Beyond Conventional Boundaries

New Technologies, Methodologies, and Procedures for the Benefit of Aerial Archaeological Data Acquisition and Analysis

Geert Verhoeven
Beyond Conventional Boundaries

New Technologies, Methodologies, and Procedures for the Benefit of Aerial Archaeological Data Acquisition and Analysis

Geert Verhoeven

Thesis Presented for the Degree of
Doctor of Archaeology

2009
Notice of Rights
The author and the promotor give the authorisation to consult and to copy parts of this work for personal use only. Every other use is subject to the copyright laws. Permission to reproduce any material contained in this work in any form by any means, electronic, mechanical, photocopying, recording, or otherwise, should be obtained from the author.

Notice of Liability
The information in this thesis is distributed on an “As Is” basis, without warranty. While every precaution has been taken in the preparation, neither the author nor the promotor or the publisher shall have any liability to any person or entity with respect to any loss or damage caused or alleged to be caused directly or indirectly by the instructions contained in this thesis or by the software and hardware products described in it.

Trademarks
All product names and services identified throughout this thesis are used in editorial fashion only and for the benefit of such companies with no intention of infringement of the trademark. No such use, nor the use of any trade name, is intended to convey endorsement or other affiliation with this thesis.

Cover
The cover shows the map of the Adriatic Roman town of Potentia (central Italy, Regione Marche) as it was deduced from aerial photographs acquired in the visible, near-infrared, and red edge spectral region. Examples of all three are shown in the upper right corner of the cover.

The research written down in this thesis was conducted with the permission and financial support of the Fund for Scientific Research – Flanders (FWO). Unless otherwise indicated, all photographs and illustrations are created by the author.

NUR 682
« I have no special talents. I am only passionately curious. »
Albert Einstein (1879-1955), physicist
Foreword

When I started my doctoral research as a Ph.D. fellowship of the Research Foundation – Flanders (FWO) in October 2004, I was confident that it should be possible to meet all the goals outlined a few months before. The research would be carried out in the framework of the Potenza Valley Survey (PVS) project, and comprise essentially the processing and analysis of all aerial imagery acquired since the beginning of the survey. However, the final outcome turned out to be slightly different from the initial purpose. The next paragraphs will sketch the underpinning reasoning for this.

In January 2000, Ghent University (Belgium) initiated the geoarchaeological PVS project in the central Adriatic Regione Marche (Italy). This interdisciplinary project was set up by the Departments of Archaeology and Geography and, under the general direction of Prof Dr Frank Vermeulen, mainly aims at reconstructing the changing physical and human landscapes along the Potenza river, one of Marche’s major rivers. Therefore, the project got the subtitle ‘From Acculturation to Social Complexity in Antiquity: A Regional Geo-Archaeological and Historical Approach’. The Potenza, which was called Flosis in Roman times, flows over its eighty kilometres course from the Apennines to the generally flat Adriatic coastline zone. To fulfil the initially outlined goals, the research focused on three sample zones systematically spaced along the river: the first was situated in the Apennine foothill landscape, the second in the undulating agricultural land of the middle valley, while the third zone encompassed the flat alluvial zone near the Potenza estuary. The interdisciplinary research initially consisted of four cornerstones: aerial survey, field survey, geomorphological survey (directed by Prof Dr Morgan De Dapper), and historical survey. In 2004, the main research techniques were expanded with a large segment of geophysical and topographical survey.

As all individual research pillars abundantly generate a wide variety of geographically linked data, it was intended from the beginning of the project to store and manage all data by means of a Geographical Information System (GIS): a completely integrated hard- and software system that enables the generation, storage, managing, analysis, and display of spatially linked data (e.g. roads, soil types, archaeological sites, field boundaries). The key of GIS is its ability to...
Foreword

connect these data (e.g. the course of a river) with non-spatial attribute data (e.g. the name of the river). This way, interactive map queries, complicated data analyses, and interesting information visualisations are possible.

In September 2002, I graduated as a Master of Archaeology with a dissertation on the archaeological use of GIS. More specifically, a GIS analysis of the field survey data generated in the first PVS season (i.e. 2000) was performed. As I was the first to do so and had already shown a considerable interest in computer applications, Frank Vermeulen – for obvious reasons the supervisor of my dissertation – proposed me to start as a scientific employee at Ghent University in January 2003. This way, I became the ‘GIS man’ of the research project. In the next fifteen months, I mainly managed the large amounts of data, built and rebuilt the PVS database, created the first PVS website, and generated all illustrative material: from detailed site maps to small-scale overviews of settlement patterns, most of them meant for publications and presentations. Moreover, these activities brought me into contact with the aerial reconnaissance data. After a few months, I was already trying to rectify and map some of the most intriguing oblique images that had been actively acquired since the start of the project. Partly out of necessity, partly because I had a serious interest in conventional ground-based photography, and partly because Ghent University had a long tradition in archaeological aerial reconnaissance, Prof Vermeulen proposed me at the end of 2003 to apply for a bursary with the Research Foundation – Flanders [Fonds voor Wetenschappelijk Onderzoek (FWO) – Vlaanderen] on the topic of archaeological aerial photography in the PVS. Even though I was initially rather reluctant (“I would never do a doctoral research”), he managed to change my mind and by the end of January, I finished my proposition which concentrated on three main research topics: rectification, interpretation, and confrontation.

First of all, I wanted to examine all possible rectification and restitution methods to geometrically correct aerial photographs for tilt and relief distortions. It was also the aim to execute a kind of cost-benefit analysis between the very expensive photogrammetric software packages and the low-cost programmes such as Aerial or AirPhoto. This knowledge would allow me to choose the right package and the most suited transformation for a particular set of images, taking also specific time and accuracy thresholds into consideration. In the ideal situation, the research and rectification of all oblique PVS photographs should take me about one year, with another six months for all historical and vertical photographs (which still had to be collected).

In a second part, these rectified frames should be interpreted and the archaeological anomalies mapped. In relation to this, several small-scale test-excavations, corings, and artefact pickups were planned in an attempt to attribute a specific function and period to particular features. Moreover, this step encompassed the testing of various data processing algorithms (e.g. filtering), which would yield a fixed set of processing routines that could maximize the extraction of relevant information embedded in the aerial frames. The thus obtained information would finally lead to an all-encompassing analysis of field survey data, oblique aerial photographs, vertical historical aerial frames, and literary sources, answering
questions such as: “Which type of features can be recognised by one or another method?”; “Do differences in detection rate occur, and if so, are these geographically, crop, soil or historically-archeologically bound or rather due to the data acquisition technique itself?”; “Is it possible to quantify the contribution of aerial photography, field survey, satellite imagery?”. In the end, this complete process was planned to quantitatively and qualitatively assess the value of oblique aerial photography in a GIS analysis of the settlement history along the Potenza River. By integrating both technical information on aerial photography and case studies from the PVS project, the end product should be a kind of archaeological manual that could be used by archaeological remote sensing specialists and novices alike.

After I was told that the FWO granted me a bursary, I presented these huge plans in September 2004 on the annual conference of the Aerial Archaeological Research Group (AARG), which took place in Münich. Although I got many positive comments from the real aerial photography specialists, there was one participant (Dr Michael Doneus) who warned me against this huge undertaking. More specifically, he told me to drop all but a few topics, as fulfilling my proposal would be “enough to get a lifetime achievement award”. Nevertheless, nothing could stop my enthusiasm and still trying to prove him wrong, I started reading as much as possible: from the principles of rectification and photogrammetry (as I proposed), to the physical behaviour of light, the psychophysical concepts of colour perception, and the various analyses of air- and spaceborne remotely gathered imagery. About one and a half year later, things started to look different. Although I had already mapped (and published) the layout of two Roman cities (Potentia and Ricina), co-organised the geoarchaeological Broadening Horizons congress, and was still producing nearly all PVS illustrations, I became aware that it was best to discard all my initial master plans: executing all proposed analyses – let alone writing a manual in English – would be impossible in a time span of four years. On the other hand, the large amount of (technical) reading on a wide variety of remote sensing topics provided me with various new insights. I became more and more familiar with the common problems and techniques of aerial archaeology. Strikingly, it seemed that archaeological aerial reconnaissance was to a very large extent still performed as about hundred years ago: flying around in a small aeroplane, using small or medium format still cameras to fastly and easily monitor archaeological anomalies in the visible domain.

In a first stage, I wanted to assess the possibilities of pure Near-InfraRed (NIR) photography on the city of Potentia. By blocking all visible light and recording only this invisible radiation, it was hoped to reveal some building structures that stayed hidden for the conventional aerial sensing methods. Moreover, this would also be a good test case for a method that was only scarcely used in archaeological research. Because the NIR-sensitised film emulsions were only moderately sensitive to pure NIR radiation, their convenient use was prohibited in all but sunny circumstances. As a consequence, a plan was conceived to build a very stable and remotely controllable aerial platform for site-based photography. Besides a means to allow NIR photography in sub-optimal circumstances, this tool would also prove useful to photograph the different stages of the planned excavations at Potentia. After more than eleven months of designing, building, and looking for funding, an NIR-enabled analogue camera could in
September 2006 finally be taken aloft using this newly built device. As it was conceived around a Helikite (i.e. a helium balloon with kite wings), the application was inaugurated HAP or Helikite Aerial Photography.

However, this first HAP test with NIR film would never be repeated. Not only because few labs could still properly develop NIR film at that time (something I learned the hard way as only the last three frames of the film were not completely ruined during its travel throughout Europe), but rather due to the fact that I became aware of the inherent spectral properties of Digital Still Cameras (DSCs or digital photographic cameras). Since nobody really seemed to take advantage of their sensitivity to both invisible Near-UltraViolet (NUV) and NIR radiation, and the potential of straightforward digital NIR (let alone NUV) photography was never thoroughly explored in archaeological reconnaissance, it seemed scientifically very interesting to focus my research more on these topics.

The last two years of my Ph.D., I mainly tried to develop low-cost methods that could allow (aerial) archaeologists to capture and analyse invisible NUV and NIR radiation. To explain why certain archaeological anomalies show up better in particular spectral bands, I browsed through large amounts of biological and (to a lesser extent) remote sensing literature to familiarise myself with the spectral reflectance characteristics of growing and senescing vegetation. Although one could obviously assume these facts to be known by the majority of aerial archaeologists, the contrary is true. As a matter of fact, it seemed that even most remote sensing manuals present a very broad, too generalised picture of the subtle reflectance characteristics of green vegetation. By exploiting these newly mastered vegetation properties, new approaches for multispectral and narrowband crop mark photography were developed, mathematical operations between spectral bands proposed, and various data acquisition methods (e.g. the use of the Helikite or flying with simultaneously operated digital cameras) tested. Since all these technologies, methodologies, and procedures have already been published (or submitted for publication) in major peer-reviewed journals, it seemed appropriate to group them. This way, the thesis under consideration comprises a collection of international papers, the first time this happens at Ghent University’s Department of Archaeology.

This approach does, however, have some consequences. First of all, the work presented here is only a portion of what has been achieved in the last four to five years. Although it might be obvious that it is impossible to include all the hours spent on database and data maintenance, map making and digitalisation, it is important to stress that more articles and chapters have been written than those included in this compilation (see Publications p. 295). Moreover, recently conducted research – which showed that it is possible to extract archaeologically relevant information from very common colour frames in far better ways than hitherto performed – could not be included. Since time was lacking to mould these results into an understandable article, its incorporation would have contrasted too much with the main body of published and submitted research articles. Additionally, the ongoing mapping of the Roman cities of Treia and Ricina was also not included for exactly the same reason.
Although the latter two mapping projects were part of the initially outlined research goals (which are, obviously, still only partly fulfilled), I currently dare to state that the thesis under consideration justifies my approach as it seems that – for the first time – economic multispectral tools are available for common archaeological research. More strikingly, these devices are not of the exotic type; they are easy to get and uncomplicated to operate, because every aerial archaeologist uses them almost on a daily basis: digital photographic small-format cameras. Together with Helikite-based photography, these small technological gems make sure that archaeological aerial imaging has the potential to never be the same again. With respect to both the spectral bands acquired as well as the altitude of image acquisition, the new approaches presented in this thesis allow to go beyond the conventional boundaries of active archaeological aerial imaging.
Preface

Take a solid manual camera, load it with some Black-and-White (B&W) film which allows a good tonal reproduction and exhibits the right amount (or absence) of graininess and – very important – choose a lens with an appropriate focal length. Afterwards, add to these three ingredients a perfect composition, a balanced proportion between out-of-focus and pin-sharp areas as well as a good dose of luck. Finally, sprinkle this mixture with some spot-on exposure plus a solid knowledge of chemistry and a stunning picture can be served. This could be a photographic recipe written by Ansel Adams (1902-1984) himself. Known from his Yosemite National Park pictures, this American photographer is – together with the recently deceased Henri Cartier-Bresson (1908-2004) – maybe the most famous and best(-known) photographer ever. Without his development of the zone system, a technique he created around 1940 in order to allow his students at the Art Center School in Los Angeles to produce better exposed photographs, probably even more photographers than now would still be struggling with determining the exact photographic exposure.

Besides pure photographical technique, chemistry played a very important photographic role in those days (and it does even now in the darkrooms of fundamental film-adepts). No matter how beautiful a latent image on a negative (or positive in the case of slide film) was, photographers could say goodbye to their ideal photograph if the film was developed poorly. However, even a well developed negative could result in positive prints with different qualities. Dedicated (aerial) photographers could create pictures with more details in the shadows and the highlights, better corrected colour casts, and a higher perceived sharpness than the local photo lab could. To a certain extent, darkroom chemistry also allowed to ‘rescue’ pictures which lacked visual impact due to bad exposure or other shortcomings. By applying the right combination of photo paper, photographic developer, and filters, the contrast of hazy aerial frames could be improved to a very large extent, making certain features (e.g. archaeological anomalies) much more distinct.

Since the advent of digital photography in the 1990s, a completely new world opened for a lot of people. Just look around: there has never been a moment in history that so many people
actually had and used a photo camera, whether it is embedded in a mobile phone, a Personal Digital Assistant (PDA) or a device designed ‘only’ with the purpose of taking photographs. The last decade, an increasing number of aerial archaeologists have also been seeing the light and converted to digital photography. Though this process was not often without striking a blow, it is safe to state that the majority of (aerial) photographers already knows more about enhancing a digital image in photo editing software than they have ever known about darkroom techniques. The direct approach (there is no ‘preview’ button in the darkroom), the ability to work in daylight with ‘clean’ computers instead of juggling with toxic products in darkrooms, and the relative easiness as well as low cost are only some of the advantages that digital image acquisition and (post-) processing benefit from. However, new technologies often entail misconceptions, a certain amount of laziness, and a decreasing understanding of fundamental operating principles. To all this, digital photography has not been an exception.

Because digital photo cameras (from now on denoted DSCs or Digital Still Cameras) are packed with several high-tech algorithms and processing components, the large majority of people completely relies upon their DSC to calculate the correct exposure and White Balance (WB), while expecting nothing less than a vibrant, pin-sharp image perfectly suited for output to a monitor, projector or printer. How the DSC calculates the exposure and the WB is not their concern, neither is the image format used to store the acquired photographic data. While the occasional photographer can surely not be blamed for this, this ‘intellectual indifference’ is less appropriate for scientists whose job it is to record archaeological anomalies by means of these devices.

To defend themselves, aerial archaeologists often use the argument that “the quality of the photograph will not change the archaeological appraisal of the features recorded”. Although this might be true to a very large extent (e.g. a digital frame with a wrong WB does not hamper the detection and interpretation of crop marks in most situations), the next chapters hopefully prove that a basic understanding of a DSC’s operating principles (i.e. the way they acquire and process the incoming radiant energy to pleasing pictures) can allow scientists to exploit their inherent (spectral) properties to a far greater extent than is conventionally done, certainly when these properties are related to the reflectance characteristics of the surfaces commonly imaged in archaeological aerial reconnaissance (i.e. plant and soils).
Roadplan. This thesis is divided into five parts, each consisting of two or three chapters. The content of every individual chapter (except one) has been published in a Journal or Proceeding that is included in the Web of Science or received an A code in the ERIH list. To indicate this, the following acronyms are mentioned at the beginning of each chapter:

ERIH (A) = publication in an A-journal as determined in the European Reference Index for the Humanities Initial List for Archaeology;
SCIE = publication part of the Web of Science, mentioned in the Science Citation Index Expanded;
SSCI = publication part of the Web of Science, mentioned in the Social Sciences Citation Index;
AHCI = publication part of the Web of Science, mentioned in the Arts & Humanities Citation Index;
CPCI-S = publication part of the Web of Science, mentioned in the Conference Proceedings Citation Index – Science;
IF = Impact Factor of this journal as mentioned on the Web of Science.

In case the article has not yet been published, it is indicated whether the manuscript was already accepted for publication or only submitted. Every chapter always bears a heading that is to a large extent similar to the article’s original title, but was reformulated for purposes of comprehensibility, clarity, and layout. However, together with the codes and the article’s status (i.e. published, accepted, or submitted), the full and correct reference is given on the chapter’s title page. After this main page, every chapter starts with the abstract of the article, followed by the main body of text. In most cases – certainly when dealing with more methodological papers – an extensive discussion and/or conclusion finalises the chapter.

Although an attempt was made to uniform the layout, terminology, and typography of every chapter, the content remained unaltered (except for a few language corrections and cross-references). As already indicated, one chapter was ‘only’ published in an ISI proceedings volume (Chapter 1), while Chapter 6 is published in a journal not appearing on the ISI record nor the ERIH list. However, both articles were included in this thesis, since it is the author’s conviction they both present a useful introduction to different but important concepts of aerial archaeology: image acquisition and rectification in the PVS project on the one hand, and the concepts of focal length, sensor size, and field of view on the other.

For the reader’s convenience and to improve the connection between the individual parts, the chapters are arranged more or less chronologically, even though their publication data may sometimes seem to indicate otherwise. The latter, however, results from two main factors. On the one hand, it takes time to develop initial ideas and write an article, so certain pieces of text were finished earlier than others. On the other hand – and more importantly – the reviewing, editing, and printing of a manuscript can suffer from serious delays once it has been submitted. In some cases, this process took almost two years. Nevertheless, presenting the chapters in their current order should facilitate an understanding of the author’s mental process in the last four to five years, while a linear reading of this thesis will also allow the reader to grasp in the best possible way the necessary concepts needed to fully understand
and qualify the information in subsequent chapters. This thesis has been written – and should therefore be read – with this in mind.

The obvious disadvantage of such a grouping of articles is the overlap in content, in ideas, in illustrative material, in references. Obviously, all articles merit an introduction, and a fluent reading of the article often necessitates a short reintroduction of concepts and methodologies that were already outlined previously. Therefore, the reader will sometimes witness the appearance of (nearly) identical illustrations, and explanations of concepts. However, this kind of structure can also be hugely beneficial in case information on a specific topic is wanted. As it was the aim of each individual research article to be as complete as possible (this often caused serious problems with the tight word limits of most journals), every chapter contains the elementary information (e.g. historical and technical backgrounds) to comprehend the subject under consideration, while the individual discussions and conclusions are specifically tailored to the particular technique or methodology being outlined. At the end of this thesis, however, a general discussion will try to integrate all the various intermediate conclusions and reflect on everything that has been written.

**Part 1 – The Potenza Valley Survey: Archaeological Research Framework.** This first part can be considered an introductory section that should allow the reader to familiarise him- or herself with the basic principles of archaeological aerial photography, and the implementation of aerial reconnaissance in the PVS project (Chapter 1). The subsequent chapters will build upon this and present the first extensive mapping projects executed in the framework of the PVS. More specifically, Chapter 2 deals with the mapping (and analysis) of *Potentia, Ricina,* and *Trea,* three Roman cities situated along the Potenza river. In a third chapter, the mapping of *Potentia* is taken to a higher level by including new aerial imagery and geophysical data. The information presented in these three chapters is very important to understand the common methodologies used in this project, as well as the landscape and the sites that are dealt with. As all aerial photographs shown in the subsequent chapters were acquired in this particular geographical area, these Roman cities served as test cases to compare the conventional approaches with the ‘revealing power’ of the newly proposed techniques. Before presenting the latter, some new terminology and file formats related to digital photography must be explained (Part 3), while the working principles of an original aerial platform – initially developed to allow these new image acquisition methodologies to be applied – shall be clarified and illustrated in Part 2.

**Part 2 – Taking Still Cameras Aloft: The Sky Must Not Always Be the Limit.** In the large majority of cases, aerial archaeologists actively acquire their own data from a low-flying aeroplane (or other manned aircraft such as helicopter, paramotor or powered parachute) or use data captured by a particular air- or even spaceborne device. In this context, one can truly state that the sky is the limit. However, remote sensing data captured from these platforms might in some occasions be totally unsuitable to meet the aims of the research (e.g. the data is acquired from too high above the surface or the cost and image characteristics are not suited for archaeological projects). In such occasions, archaeologists need a device that allows them
to capture aerial data from a relatively low altitude at all moments of the day. This part will introduce such a platform, based on the advantageous use of a Helikite. Initially, this device was constructed to apply and research the technique of analogue pure Near-InfraRed (NIR) photography on the city of Potentia, as it was hoped that it could reveal more information on the building features intra muros. Besides its initial purpose as research instrument, this platform has also been used extensively to map all the excavation phases at Potentia. As a kind of introduction to such unmanned aerial platforms and to understand why the newly developed device is in many aspects superior to existing constructions, Chapter 4 gives a detailed overview of most devices ever used in archaeological research to acquire this so-called close range imagery. Afterwards, a thorough explanation of the Helikite-based construction itself is provided in Chapter 5.

**Part 3 – Digital Still Cameras: Concepts and Image Formats.** One of these often heard and read delusions is the fact that “the focal length of lenses changes when used on a digital camera”. The purpose of this part is to prove the above statement incorrect by presenting some essential terminology (e.g. focal length, sensor size, and field of view) of the (digital) photography world (Chapter 6). Besides new concepts, a completely new image format (i.e. RAW) was introduced with the advent of Digital Still Cameras (DSCs). Chapter 7 will explain how this image format is created inside the DSC and uses this understanding to prove its elementary application in scientific aerial photography. Together with the geographical-methodological-archaeological introduction of Part 1, and the information on the new aerial platform explained in Part 2, the concepts outlined in this Part should provide the reader with the necessary understanding to fully grasp the new approaches presented in Part 4 and Part 5, as well as the general situation and conditions which made their development possible.

**Part 4 – Pushing Theoretical Limits: Near-InfraRed Aerial Archaeology.** From this point on, the inherent spectral properties of DSCs will be explored and their potential in archaeological aerial reconnaissance indicated. In an introductory text (Chapter 8), the differences between digital and film-based Near-InfraRed (NIR) photography are outlined. Furthermore, the general application of NIR in archaeological research is illustrated. The second article (Chapter 9) largely builds upon the principles given in the previous chapter to explore the different responses plants exhibit in both the visible and NIR spectrum. By presenting real-world examples, the cases in which aerial NIR crop mark imaging can be advantageous are being unravelled. Furthermore, a theoretical explanation is given as to why certain plant properties (e.g. water stress) are visualised better in the visible and others in the NIR spectral band, something that has been ignored too often in archaeological aerial reconnaissance. Chapter 10 further explores the possibilities of an NIR-converted DSC in extracting even more meaningful information from a generated NIR frame. Because this approach necessarily implies the exact spectral responses of the DSC to be known, Chapter 10 starts with a detailed description of the methodology used to acquire such channel specific spectral curves. Afterwards, this information is used to generate a vegetation index based on a mathematical operation of all channels acquired by the same camera. This way, modified DSCs are proven to be cheap, compact, robust, and easy-to-handle tools that allow a ‘spectroscopic’ aerial image acquisition.
Part 5 – Is Smaller also Better? Red Edge and Near-UltraViolet Sensing. Using the theoretical facts on plant reflectance on the one hand and camera sensitivity on the other, this part explores imaging in a very small wavelength range (i.e. the Red edge), while it also covers the archaeological potential of image generation by smaller-than-visible wavelengths: Near-UltraViolet (NUV) radiation. More specifically, Chapter 11 further refines the method explained in Chapter 10 by taking three simultaneously operated DSCs inside an aeroplane to capture information in the visible, Red edge, and NIR spectral region. This Red edge region, which forms the transition zone between the visible and NIR spectral band, is of major importance because biologists have proven several times that plants exhibit the most consistent and largest stress response in this small region centred on 700 nm. NUV photography has to be situated on the other side of the visible spectrum. This highly-energetic radiation (i.e. short wavelengths) is extremely hard to capture from conventional air- and spaceborne platforms. Consequently, it has never been used in archaeological aerial reconnaissance. Using the Helikite and a special astronomical interference filter, Chapter 12 shows how NUV aerial photographs can be generated and proves the usefulness of these smaller-than-visible wavelengths for site-based soil mark archaeology. This way, both pioneering approaches can answer positively to the question asked in the heading of this part.

Conclusion. In this last piece of text, a round-up is provided of all technologies, methodologies and procedures tackled in this thesis, while guidelines for improvements and future research, both in the PVS project and aerial archaeology in general, will be given.
Acknowledgements

The work described in this document has been carried out at the Department of Archaeology and Ancient History of Europe (Ghent University) since September 2004. This endeavour would, however, not have been completed without the support of many people. Running the risk of forgetting someone (memories move in mysterious ways), I will mention ‘some’ persons in particular. Even though most of them were not directly involved in writing my thesis, they certainly enriched my life the last five years.

First of all, I would like to thank my promotor, Prof Dr Frank Vermeulen, for taking me as a student to the Potenza valley and offering me my first real job as a scientific employee at Ghent University. Two years later, he gently pushed me towards a doctoral research, but it is a choice I have never regretted, and always fulfilled with great passion. Frank, many thanks for giving me almost complete freedom of action and the necessary support when needed. This way, I could pursue my very own research interests and find my own path to get this work done.

Secondly, I would also like to express my sincere gratitude to Prof Dr Rudi Goossens. I had the pleasure to attend some of his enjoyable remote sensing courses which helped me a big deal in my understanding of aerial photography. Rudi, thank you very much for sharing your knowledge with the archaeological community on the one hand, and your personal advice on the other.

I am particularly grateful for having Dr Michael Doneus as a reviewer for my thesis, because he is one of the most outstanding experts in the field of archaeological remote sensing. Moreover, I want to thank him for the many insightful conversations we had during various congresses and the support he gave as my foreign supervisor. His hospitality also made my short stay in Vienna a real pleasure.

The efforts of the other members of the jury (Prof Dr Jean Bourgeois and Prof Dr Morgan De Dapper) are also sincerely appreciated. Both scholars have an excellent expertise in their
particular research fields, have dealt a lot with aerial photography during their academic
career, and their review of (and opinion on) my thesis will undoubtedly be of great value.

I was also very fortunate to have the opportunity to work together with scientists from other
departments and institutes. Even though we never physically met, Dr Klaus Schmitt must be
thanked for his elementary information on Near-UltraViolet (NUV) photography. He supplied
the necessary optical hardware needed for my aerial imaging in this invisible domain, while his
contributions also considerably raised the scientific level of my NUV article. The enthusiasm of
former Master student Ir. Jo Loenders was of the utmost importance in the development of my
Helikite-based system for remote controlled aerial photography. Finally, Dr Philippe Smet and
Dr Peter Bruggeman were essential in testing spectral responses and optical transmissions of
photographic equipment. All four are greatly acknowledged for this.

The members of AARG (Aerial Archaeological Research Group) also deserve a word of
gratitude: I never encountered such a friendly and hospitable group of scientists. They gave
me the opportunity to disseminate my research results to fellow aerial photographers. Besides
the necessary discussions I had with most AARG members and the well-deserved drinks we
shared afterwards, I must particularly thank Dave Cowley. He has always been very enthusiastic
(but not less critical) about my work and was willing to proofread several of my articles.

A big thank goes to Roald Docter in helping me to get funding for my Helikite. Gitte Callaert, Kris Hemeryck, and Frederic Lamsens must be thanked for their IT support, while I also want to
express my gratitude to many (former, current, and newbie) UGent and other (scientific)
colleagues, because they provided a source of knowledge, a ‘listening ear’ or a nice drink and
chat: Joris Angenon, Peter Attema, Machteld Bats, Arjen Bosman, Ignace Bourgeois, Bart
Cherreté, Philippe Crombé, Aurelie Daems, Wim De Clercq, Kaatje De Langhe, Giuliano De
Marinis, Johnny De Meulemeester (†), Guy De Mulder, Veronica De Laet, Hilde De Paepe,
Jeroen De Reu, Tijs De Schacht, Beata-Maria De Vliegher, An De Waele, Johnny Devreker, Sarah
Deprez, Parsival Delrue, Dennis Devriendt, Eline Deweerdt, Joke Dewulf, Evelyne Dhaenens,
Lieve Donnellan, Anton Ervynck, Wouter Gheyle, Marco Giardino, Cornelia Glaesser, Tanja
Goethals, Darja Grosman, Ernie Haerinck, Davy Herremans, Karen Jeneson, Pieter Laloo, Kerstin
Leterme, Marc Meganck (†), Patrick Monsieur, Branko Mušić, Gunther Noens, Bart Ooghe,
Edvige Percossi, Dorothy Pikhaus, Ariane Raman, Joris Sergant, Božidar Slapšak, Eric Smekens,
Anke Stoker, Birger Stichelbaut, Peter Vandenabeele, Debby Van den Bergh, Martijn van
Leusen, Koenraad Verboven, Philip Verhagen, Angelo Verlinde, Jan Walstra, and Andries
Zuiderhoek. More in particular, I was very fortunate to have Catharina Boullart, Karen
Ryckbosch, Devi Taelman, Lieven Verdonck, and Hélène Verreyke as ‘office-companions’. Their
friendship and professional collaboration meant a great deal to me, and I definitely enjoyed all
afternoon breaks and countless (hilarious) discussions. An additional word of thanks also to
Devi, Karen, and Lieven for proofreading several pieces of text included in this thesis.

A big “thank you” also for Eamonn Baldwin, Carlo Alberto Bozzi, Francesca Carboni, Sarah de
Seranno, Alain De Wulf, Sophie Dralans, Gabriele Fangi, Rose Ferraby, Sophie Hay, Daniel Van
Acknowledgements

Damme, Dimitri Van Limbergen, Maarten Weyler, and all the archaeology, geography, and topography students that made the field campaigns to Italy very fruitful and certainly memorable.

My good and long standing friends Jan Broekaert, Bart De Wilde, Wouter Dhaeze and Joanna, Nele Eggermont, Tim Sautois and Veerle, Wouter Van Hecke, Julie Van Kerckhove and Ivo: you must know I really enjoyed all our trips, the conversations and drinks in the bars, the parties, the sporting. I would like you to know that you mean a lot to me. Wouter and Tim: thanks for reading through several of my articles and correcting my awkward phrases.

I would also like to thank the friends of the soccer teams ‘Campus’ and ‘FC Endelarm’, Nicolas Bracke, Olmo Cornelis, Henk Dellaert, Thierry De Coninck, Bart Sautois and Sofie, Vanessa Schelstraete, Tim Spiesschaert, Katrien Van Hecke, Liesbet Van der Heyden and Wim, Steven Verpaele, and many others for helping me to forget the hard Ph.D.-related labour. I also owe much gratitude to Iris Kolkman. Iris, thank you for the support and the many moments of joy during more than four years. You certainly coloured up my research-minded lifestyle. I wish you all the best.

Not at least, I should express my gratitude to my family: uncles, aunts and cousins. However, I am most of all grateful to my parents Gilberte De Paepe and Ronny Verhoeven. Besides offering me the opportunity to study at the university, their good advice, patience, and forbearance were essential in those moments that I was struggling with all kinds of problems. More specifically, thanks ‘moema’ for all the food you brought and the cooking you did in the last weeks. Thanks ‘poepa’ for solving all technical problems in my house and proofreading pieces of text. To my younger brothers Koen ‘den dikken’ and Bart ‘den smallen’ and their consorts Evelyne and Katrien – whom I all wish the best with their lives and careers – I want to say: go forth and multiply! (and teach your children English so they can read uncle’s little book when they are not able to get to sleep).

Ghent, March 2009

Geert Verhoeven
# Table of Contents

Foreword vii  
Preface xiii  
Acknowledgements xix  
List of Figures xxvii  
List of Tables xxxvii  
Abbreviations and Acronyms xxxix

## Part I. The Potenza Valley Survey: Archaeological Research Framework 1

### Chapter 1  From Photograph to Plan – Mapping Aerial Imagery in a GIS for Geoarchaeological Research

Abstract 4  
1.1 The Potenza Valley Survey 5  
1.2 History of Aerial Survey 7  
1.3 Principles of Aerial Archaeology 7  
1.4 Aerial Photography in the PVS 9  
1.5 Field Campaign 2004 13  
1.6 Prospects 14  
1.7 Conclusion 14

### Chapter 2  Roman Cities along a Valley – Aerial Photography and Field Survey to the Study of Urbanization

Abstract 18  
2.1 The Potenza Valley Survey 19  
2.2 Centralized Settlement before the Romans 21  
2.3 Roman Urbanization in the Coastal Area: Potentia 24  
2.4 The Inland Municipia of Ricina and Trea 32  
2.5 Conclusions 45
Table of Contents

Chapter 3  Many a Little Makes a Mickle – Combining Information Layers to Unravel the Organisation of a Roman City 49
Abstract ................................................................................................................................................. 50
3.1 Introduction .................................................................................................................................... 51
3.2 The Roman City of Potentia ........................................................................................................ 52
3.3 Survey Methodology ................................................................................................................... 56
3.4 The Survey Results and their Integration .................................................................................. 59
3.5 Conclusion .................................................................................................................................... 69

Part II . Taking Still Cameras Aloft: The Sky Must Not Always Be the Limit __________ 73

Chapter 4  Archaeological Bird’s-Eye Views – Various Ground-Based Means to Execute Low Altitude Aerial Photography 75
Abstract ................................................................................................................................................. 76
4.1 Introduction .................................................................................................................................... 77
4.2 Low Altitude Aerial Photography ............................................................................................... 79
4.3 Mast, Pole or Boom .................................................................................................................... 79
4.4 UAV ............................................................................................................................................... 80
4.5 Kite ............................................................................................................................................... 81
4.6 Balloon or Blimp ........................................................................................................................ 83
4.7 Helikite ......................................................................................................................................... 85
4.8 Applications .................................................................................................................................. 86
4.9 Conclusion .................................................................................................................................... 87

Chapter 5  Helikite Aerial Photography – A New Means of Unmanned, Radio-Controlled Close Range Aerial Archaeology 89
Abstract ................................................................................................................................................. 90
5.1 Introduction .................................................................................................................................... 91
5.2 LAAP Platforms .......................................................................................................................... 91
5.3 HAP – The System ...................................................................................................................... 94
5.4 HAP – Cost, Maintenance, and Operation ................................................................................. 100
5.5 HAP – Possible Improvements and Drawbacks ...................................................................... 101
5.6 HAP – Archaeological Applications ......................................................................................... 102
5.7 Conclusion .................................................................................................................................... 106


Chapter 6  Cropping the Aerial View – On the Interplay between Focal Length and Sensor Size 111
6.1 Introduction ................................................................................................................................... 113
6.2 Film Format Does Matter ............................................................................................................ 113
6.3 Small Format Image Sensors ..................................................................................................... 116
6.4 Some General Lens Properties ................................................................................................. 118
6.5 Tracing Rays to Reveal Focal Length Principles ....................................................................... 121
6.6 Field of View – Putting It All Together .................................................................................... 123
6.7 Crop Factor .................................................................................................................................. 128
6.8 Wide-Angle, Normal, and Long Focus Lenses ......................................................................... 130
6.9 Ground Coverage ....................................................................................................................... 132
<table>
<thead>
<tr>
<th>Chapter 7</th>
<th>It's All about the Format – Unleashing the Power of RAW Aerial Photography</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>136</td>
</tr>
<tr>
<td>7.1</td>
<td>Introduction .............................................................................................. 137</td>
</tr>
<tr>
<td>7.2</td>
<td>RAW – A Definition .................................................................................. 137</td>
</tr>
<tr>
<td>7.3</td>
<td>RAW – The Creation .................................................................................. 138</td>
</tr>
<tr>
<td>7.4</td>
<td>RAW – File Details .................................................................................. 142</td>
</tr>
<tr>
<td>7.5</td>
<td>RAW – Processing .................................................................................... 143</td>
</tr>
<tr>
<td>7.6</td>
<td>RAW – Saving ........................................................................................ 156</td>
</tr>
<tr>
<td>7.7</td>
<td>RAW – Workflow and Software .................................................................. 161</td>
</tr>
<tr>
<td>7.8</td>
<td>RAW – Disadvantages ............................................................................. 163</td>
</tr>
<tr>
<td>7.9</td>
<td>Conclusion ............................................................................................ 166</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Part IV</th>
<th>Pushing Theoretical Limits: Near-InfraRed Aerial Archaeology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 8</td>
<td>Imaging the Invisible – Modified Digital Still Cameras for Low-Cost and Straightforward Archaeological Near-InfraRed Photography</td>
</tr>
<tr>
<td>Abstract</td>
<td>169</td>
</tr>
<tr>
<td>8.1</td>
<td>Introduction ................................................................................ 170</td>
</tr>
<tr>
<td>8.2</td>
<td>NIR Imaging in Archaeology .................................................... 174</td>
</tr>
<tr>
<td>8.3</td>
<td>NIR Imaging in the Film Era ..................................................... 177</td>
</tr>
<tr>
<td>8.4</td>
<td>Digital-Based NIR Imaging ...................................................... 181</td>
</tr>
<tr>
<td>8.5</td>
<td>Real-World Examples ..................................................................... 186</td>
</tr>
<tr>
<td>8.6</td>
<td>Conclusions .................................................................................. 192</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 9</th>
<th>Near-InfraRed Sensing of Vegetation Marks – From Historical Use to Digital Approaches</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>193</td>
</tr>
<tr>
<td>9.1</td>
<td>Introduction .............................................................................................. 194</td>
</tr>
<tr>
<td>9.2</td>
<td>Crop Marks and Related Plant Reflectance .................................................................................................................. 195</td>
</tr>
<tr>
<td>9.3</td>
<td>Conventional NIR Aerial Archaeology ...................................................................................................................... 200</td>
</tr>
<tr>
<td>9.4</td>
<td>Digital NIR Archaeological Reconnaissance ............................................................................................................ 202</td>
</tr>
<tr>
<td>9.5</td>
<td>NIR Aerial Crop Mark Archaeology – Applications and Examples .......................................................... 203</td>
</tr>
<tr>
<td>9.6</td>
<td>Conclusion ............................................................................................ 215</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 10</th>
<th>Spectral Characterisation to Enhance Observation – Exploiting the Response Curves of an NIR-Modified Camera to Enable Biomass Visualisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>217</td>
</tr>
<tr>
<td>10.1</td>
<td>Introduction .......................................................................................... 218</td>
</tr>
<tr>
<td>10.2</td>
<td>DSC Characterisation – Material and Method ........................................... 219</td>
</tr>
<tr>
<td>10.3</td>
<td>DSC Characterisation – Results and Processing ....................................... 223</td>
</tr>
<tr>
<td>10.4</td>
<td>Archaeological Results ........................................................................... 227</td>
</tr>
<tr>
<td>10.5</td>
<td>Discussion ............................................................................................ 231</td>
</tr>
<tr>
<td>10.6</td>
<td>Conclusions .......................................................................................... 236</td>
</tr>
</tbody>
</table>
Part V. Is Smaller also Better? Red Edge and Near-UltraViolet Sensing 241

Chapter 11 Crop Marks Should Be Red – Imaging the Blue Shift of the Red Edge to Detect Vegetation Stress 243

Abstract ......................................................................................................................... 244
11.1 Introduction ............................................................................................................. 245
11.2 Plants and their Spectral Characteristics ............................................................... 246
11.3 Vegetation Indices ................................................................................................. 249
11.4 Material and Method ........................................................................................... 252
11.5 Archaeological Results ......................................................................................... 256
11.6 Drawbacks, Improvements, and Considerations .................................................... 260
11.7 Discussion ............................................................................................................. 261
11.8 Conclusion ........................................................................................................... 262

Chapter 12 The Attraction of the Unknown – Digital Near-UltraViolet Aerial Archaeology 265

Abstract ......................................................................................................................... 266
12.1 UltraViolet Radiation – An Introduction .................................................................. 267
12.2 NUV Imaging – Archaeological Applications ......................................................... 270
12.3 NUV Image Acquisition ....................................................................................... 272
12.4 Archaeological Results and Explanations ............................................................... 278
12.5 Final Thoughts ..................................................................................................... 286

Conclusion 289

Publications 295

References 299
List of Figures

1.1 The Marches: the southern half belongs to Augustus’ region V, the northern half to region VI............................................................................................................................... 5

1.2 Location of the three sample zones. ........................................................................................................ 6

1.3 Four different kinds of marks in the Potenza valley: 1) ruin; 2) shadow marks; 3) crop marks; 4) soil marks......................................................................................................................... 8

1.4 The projective transformation.................................................................................................................. 11

1.5 The Fishler-Bolles image rectification and an example of some image processing. ......................... 12

1.6 The mapping of the Roman town Potentia and its suburban area....................................................... 13

2.1 Localization of Roman towns and areas of systematic field survey by the PVS team. ..................... 19

2.2 Oblique view of the protohistoric hill site of Montarice. The photograph shows clear crop marks of a circuit wall/ditch around the plateau and of different phases of houses, pits, and other settlement traces (photograph by F. Vermeulen – Ghent University). .......... 23

2.3 Results of a systematic grid survey in September 2002 on the site of Montarice........................... 24

2.4 New plan of Roman Potentia and its immediate surroundings, based on aerial photography, systematic survey, and earlier excavations. .......................................................... 26

2.5 Plan of Potentia with the best known building sectors: 1) parts of two insulae excavated in the 1970s; 2) central temple sector currently under excavation; 3) buildings seen in crop marks during aerial reconnaissance in 2003. ................................................................. 28

2.6 Aerial reconnaissance in 2003 presented a clear view of the streets and circuit wall of the northern part of Potentia. Remark differences in width and in nature of the street traces (photograph by F. Vermeulen – Ghent University). ................................................................. 29

2.7 Results of a systematic grid survey (September 2002) on and near the town of Potentia............. 31
2.8 First tentative plan of Ricina, based on information from aerial photography, field checks, earlier excavations, and discoveries. Major structures observed during aerial survey: 1) theatre; 2) temple; 3-6) domus; 7-8) horrea/tabernae (?); 9) monumental building with cistern; 10) street and adjoining structures.................................................................35

2.9 Aerial photograph of crop marks in the central and eastern part of Ricina, showing traces of the main temple (2), a domus with mosaic floors (4), and a building with preserved cistern (9) (photograph by F. Vermeulen – Ghent University)..................................................................................................................37

2.10 Aerial view of crop marks in the western part of Ricina, showing traces of houses along the main street (5-6), a possible commercial complex (7) and the city wall (x). Note the clear traces of post-Roman flooding at lower right (photograph by F. Vermeulen–Ghent University).................................................................................................................................................38

2.11 Image of building features in the western part of Ricina. On the left, a broad linear trace indicates the presence of the city wall (photograph by F. Vermeulen – Ghent University)........................................................................................................................................39

2.12 General view from the north of the former urban area of Trea. Note the location of the SS. Crocifisso and of the preserved configuration of the northern part of the Roman circuit wall in the landscape (photograph by F. Vermeulen – Ghent University)..................................................................40

2.13 A tentative reconstruction of the urban topography of Trea with position of the town walls (as seen and as presumed), the urban grid (with two distinct phases), the forum (A) and the sanctuary of SS. Crocifisso (B). This sketch, based entirely on the results of oblique aerial photography, is not fully restituted.........................................................................................................42

2.14 Oblique aerial view of the central zone of Trea. The transcript of the crop marks (not yet fully restituted) shows the presence of: 1) decumanus maximus; 2) temple; 3) forum square; 4) tabernae; 5) macellum (?); 6) basilica (photograph on the left by F. Vermeulen – Ghent University)..................................................................................................................................44

3.1 The location of Potentia in the Marche region of Italy.................................................................52

3.2 New simplified plan of Roman Potentia and its immediate surroundings, based on aerial photography, systematic field survey, geophysical survey, and earlier excavations. .................................................................................................................................55

3.3 Oblique aerial view of the urban site of Potentia from the west (spring 2003). Crop marks and the excavated temple zone (centre right) can be seen (photograph by F. Vermeulen – Ghent University)..................................................................................................................................56

3.4 Results of the 2004-2005 magnetometer survey, showing subsurface streets and walls (illustration adapted from [793]: Fig. 6). ........................................................................................................................................57

3.5 Aerial reconnaissance and old vertical aerial photos show subsurface elements of the urban topography: 1) details of streets and city wall visible on RAF images; 2) rectified oblique photograph of a street pattern; 3) rectified and filtered oblique photograph with streets and building structures (photograph 1 by RAF; 2-3 by F. Vermeulen – Ghent University)..................................................................................................................................61
List of Figures

3.6 Urban structures at Potentia based on the integration of survey and excavation data. ..........62
3.7 Results of a systematic grid survey in 2002 and 2003 on the city of Potentia. .................68
4.1 (A) German Army photo rocket squad with Maul’s equipment and pigeons with cameras (B) [573]. .................................................................................................................................................78
4.2 (A) An extensible ladder ([357]: Plate IV); (B) the Holmes Boom ([634]: Fig. 6); (C) the Photo-Turret ([581]: Fig. 1). ..................................................................................................................................................80
4.3 (A) Example of an R/C helicopter to acquire digital aerial imagery ([270]: Fig. 1); (B) Microdrone MD4-200 [535]; (C) R/C paraglider [456]. ..........................................................................................................................................................81
4.4 (A) Kite used at the site of Jebel Moya ([14]: Plate XVI.1); (B) multi-flare kite with attached camera ([756]: Fig. 27d); (C) a delta kite taking off [570]; (D) a flying rokkaku kite ([512]: Fig. 2). ...........................................................................................................................................82
4.5 (A) Pioneering archaeological BAP at Megiddo ([357]: Plate III); (B) hydrogen photo balloon ([825]: Fig. 1); (C) helium photo balloon ([19]: Fig. 2); (D) photo blimp [745]. ................83
4.6 Inflating a 7 m³ Skyhook Helikite. .................................................................................................86
5.1 Inflating the Helikite. ..................................................................................................................93
5.2 Wind speed versus photographic characteristics for competing alternatives (adapted from [207])............................................................................................................................................................................93
5.3 The complete HAP system. .........................................................................................................94
5.4 The Picavet suspension. ............................................................................................................96
5.5 The cradle (A; photograph by W. Gheyle – Ghent University) which allows the camera to tilt, tip, and swing (B). ...........................................................................................................................................97
5.6 Steering the DSC by assistance of the video live link (photograph by W. Gheyle – Ghent University). ..............................................................................................................................................................99
5.7 Fishing rod and reel attached to the navigator’s body (photographs by W. Gheyle and D. Van Limbergen respectively). ..................................................................................................................100
5.8 Decumanus maximus leaving the western gate of Potentia (HAP with Nikon D70s + Nikkor 20 mm f/3.5 AI-S). ...........................................................................................................................................103
5.9 NIR photograph of Potentia’s temple area (HAP with Nikon D50Nir + Nikkor 20 mm f/3.5 AI-S). ....................................................................................................................................................104
5.10 Near vertical photograph of an excavation area with GCPs indicated (HAP with Nikon D70s + Nikkor 20 mm f/3.5 AI-S). ............................................................................................................................................105
5.11 Monitoring excavation phases (HAP with Nikon D70s + AF Nikkor 50 mm f/1.8D). .......106
5.12 Visible (A; photograph by F. Vermeulen – Ghent University) and pure NIR record (B – HAP with Nikon D50NIR + Nikkor 20 mm f/3.5 AI-S) of a Roman town wall and connected tower. .................................................................................................................................................107
6.1 The dimensions of a 35 mm frame..................................................................................................114
List of Figures

6.2 The path of a light ray in an SLR body with mounted photographic lens. .................................................................115
6.3 Different digital sensors with their characteristics. ........................................................................................................117
6.4 The Pythagorean Theorem .............................................................................................................................................118
6.5 A single thin lens system in air .......................................................................................................................................119
6.6 The graphical sign convention. ......................................................................................................................................120
6.7 A thick lens system in air. ...............................................................................................................................................120
6.8 Ray tracing for a thin lens in air ......................................................................................................................................121
6.9 Using ray tracing to show the influence of object distance on image distance and magnification. .................................................................122
6.10 The effect of increased focal length on subject magnification ..........................................................................................123
6.11 Digitally cropping the image circle. .............................................................................................................................124
6.12 Diagonal, horizontal, and vertical FOV .........................................................................................................................125
6.13 A fisheye image (photograph by D. Slater [722]). ............................................................................................................126
6.14 Deriving the formula for FOV (only for rectilinear lenses) ............................................................................................127
6.15 Classification of lenses which apply a rectilinear projection (data from [649, 651, 812]). .................................................131
6.16 Using FOV in the calculation of field coverage ..............................................................................................................132
7.1 The layout and working of a Bayer photodiode array .......................................................................................................138
7.2 Different CFAs. ...................................................................................................................................................................139
7.3 The creation of a RAW file (adapted from [93]). ................................................................................................................140
7.4 Four different tonal ranges and the occurrence of posterisation .........................................................................................141
7.5 A RAW file. ......................................................................................................................................................................143
7.6 The main processing steps of a RAW file by the camera’s firmware or on the computer. ..............................................144
7.7 The channel specific DNs and calculated normalised multipliers retrieved from a WhiBal White Balance Reference Card photographed under different illumination conditions .................................................................................................................................145
7.8 (A) A linearly processed NIR NEF file with custom channel multipliers; (B) the same image without any WB. The lower part shows the Red and Blue channel of A and B respectively. .................................................................................................................................147
7.9 Difference in diode-specific spectral information acquired by the Foveon sensor and the CFA-solution (adapted from [297]). ..............................................................................................................................................148
7.10 Comparison between demosaicing of an in-camera generated JPEG and a desktop-generated JPEG of a Nikon D70s RAW frame. ..............................................................................................................................................149
List of Figures

7.11 The effect of different demosaicing algorithms performed on an NIR image: 1) original NIR-photograph; 2) portion of 1 showing the CFA pattern; 3) bilinear interpolation; 4) median interpolation; 5) gradient Interpolation; 6) Variable Number of Gradients (VNG) algorithm; 7) Patterned Pixel Grouping (PPG) interpolation; 8) Adaptive Homogeneity-Directed (AHD) demosaicing ......................................................................................................................... 151

7.12 Stevens’ Power Law illustrated for brightness ................................................................................................................................. 152

7.13 Gamma correction of RAW data ......................................................................................................................................................... 153

7.14 Comparison between normal gamma curve and generally applied tonal curve (adapted from [93]) ............................................................................................................................................................................ 154

7.15 Relative (%) and absolute (DN) reflectance increase in stressed vegetation canopy collected from a single, but differently processed RAW frame ........................................................................................................ 155

7.16 Dynamic versus tonal range ......................................................................................................................................................... 158

7.17 Comparison between highlight details of an in-camera generated JPEG (A) and a JPEG resulting from a RAW workflow (B) .................................................................................................................................................... 160

8.1 The ElectroMagnetic spectrum (adapted from [307]: Fig. 5-7) ............................................................................................................. 172

8.2 (A) LWIR image of a man’s face and (B) MWIR image of a four-fingered hand ([659]: Figs. 6.2 and 2.19) ............................................................................................................................................................................ 173

8.3 (A) Panchromatic and (B) reflected NIR photograph of the same artefact (adapted from [196]: Figs. 13a and 13b) ............................................................................................................................................................................ 175

8.4 Kodak EIR image of a dense archaeological landscape containing Neolithic and Roman features ([109]: Fig. 8) ............................................................................................................................................................................ 176

8.5 Colour reproduction with a CIR transparency film (adapted from [255]: Fig. 2) ............................................................................................................................................................................ 178

8.6 Spectral sensitivity characteristics of Sony’s ICX205AL CCD (adapted from [731]) and a Kodak CMOS Image Sensor (adapted from [258]: Fig. 1) ............................................................................................................................................................................ 182

8.7 The layout and working of a photodiode array with Bayer CFA ............................................................................................................. 183

8.8 Wavelength versus Quantum Efficiency for (A) the Kodak KAF-8300 and (B) the Foveon X3 sensor (adapted from [261]: Fig. 5, and [502]: Fig. 6) ............................................................................................................................................................................ 184

8.9 Spectral sensitivities of (A) the Nikon D80 and (B) the Foveon X3 sensor with CM500 NIR-cut filter (adapted from [407], and [502]: Fig. 7) ............................................................................................................................................................................ 184

8.10 Modification of the digital imaging array by replacing the hot mirror ............................................................................................................. 185

8.11 Contrast enhancement (A2 and B2) on a visible (A1) and an NIR frame (B1) clearly shows the latter to reveal better Roman features as crop marks ............................................................................................................................................................................ 187

8.12 A conventional (A1) and FCIR record (B1) of the central part of the Roman city of Trea. The enhanced versions (A2 and B2) point to the fact that FCIR can be advantageous in spotting crop marks, clearly illustrated by the walls of the Roman temple in the insets ............................................................................................................................................................................ 188

8.13 The haze penetrating capabilities of the longer NIR wavelengths (B) compared to the visible record (A) of a mountainous area ............................................................................................................................................................................ 189
8.14 Mutual comparison between visible (A) and NIR image (B) of a North African Late Punic/Early Roman bowl. (1) and (2) indicate zones of increased reduction (sherd by courtesy of K. Ryckbosch – Ghent University). .......................................................... 190

8.15 Comparison between the visible (A) and NIR record (B) of a coated rim fragment belonging to a Roman handmade pot (sherd by courtesy of W. De Clercq – Ghent University). .......................................................... 191

8.16 Difference between the visible (A) and NIR rendering (B) of the inventory number on a decorated wall fragment of a late Iron Age handmade pot (sherd by courtesy of W. De Clercq – Ghent University). .......................................................... 191

8.17 Normal (A) and pure NIR (B) version of the partly obliterated word “original”. ...................... 192

9.1 Structure of a plant leaf and its interaction with incident visible and NIR radiation (adapted from [727]: Fig. 32-3, and [748]: Fig. 2). .......................................................... 196

9.2 Spectral reflectance characteristics of healthy green vegetation ([425]: Fig. 11-21a). .............. 197

9.3 Reflectance spectra of photosynthetic (green) vegetation, non-photosynthetic (dry) vegetation, and a soil ([175]: Fig. 1.18). .......................................................... 199

9.4 (A) Relative response versus wavelength of the Nikon D50\textsubscript{NIR} with Nikkor 20 mm f/3.5 AI-S ([782]: Fig. 5); (B) the layout and principal of a Bayer CFA .......................................................... 203

9.5 General comparison between the appearance of common ground features in the visible domain (A) and NIR spectrum (B): (1) healthy green vegetation, (2) soil, (3) low density canopy, (4) water, (5) trees, and (6) distant objects. Images acquired in the middle Potenza valley (N 43° 18' 02", E 13° 16' 12" – WGS84) with a Nikon D200 (A) and Nikon D50\textsubscript{NIR} (B) on May 15, 2008 at 11.40 h. .......................................................... 205

9.6 (A) Visible image of the central part of the Roman town of Potentia (N 43° 24' 53", E 13° 40' 14" – WGS84); (B) NIR image of the same scene. To obtain versions (C) and (D), very extreme local contrast enhancement was applied to (A) and (B) respectively. Images acquired with a Nikon D200 (A) and a Nikon D50\textsubscript{NIR} (B) on May 15, 2008 at 11.14 h. ............................. 206

9.7 (A) Visible image of the southern part of Potentia, while (B) displays the same view in the NIR. Imagery was acquired with a Nikon D200 (A) and a Nikon D80\textsubscript{FS} (B) on July 03, 2008 at 11.02 h. .......................................................... 207

9.8 Hydrographical features revealed by negative crop marks in the visible (A) and NIR (B) part of the spectrum. Imagery was acquired in the central part of the Potenza Valley (N 43° 19' 42", E 13° 25' 26" – WGS84) with a Nikon D200 (A) and a Nikon D80\textsubscript{FS} (B) on July 03, 2008 at 11.23 h. .......................................................... 208

9.9 (A) Visible image of the central part of the Roman town of Ricina (N 43° 19' 41", E 13° 25' 26" – WGS84); (B) NIR image of the same scene. To obtain versions (C) and (D), very extreme local contrast enhancement was applied to (A) and (B) respectively. Images acquired with a Nikon D200 (A) and a Nikon D50\textsubscript{NIR} (B) on May 15, 2008 at 11.27 h. .......................................................... 209
9.10 A visible (A) and pure NIR image (B) of Roman Tre/a’s central part (N 43° 18’ 40”, E 13° 18’ 42” – WGS84). The insets are high contrast enlargements of the same area around the left pylon. Images acquired with a Canon EOS 300D (A; photograph by F. Vermeulen – Ghent University), and a Nikon D50_nir (B) on April 23, 2007 at 11.01 h.......................................... 210

9.11 A visible (A) and pure NIR image (B) of some agricultural plots. Image (C), which shows dark, chlorophyll-rich areas and white regions without biomass, was yielded by the SR described in the text. The imagery was acquired in the central Potenza valley (N 43° 19’ 06”, E 13° 17’ 16” – WGS84) with a Nikon D200 (A) and a Nikon D80_nir (B) on May 15, 2008 at 11.39 h. .................................................................................................................................................. 211

9.12 (A) The output of a SR operation, using only the image channels of a Nikon D50_nir (B) this SR output was generated by utilizing the D50_nir’s Blue channel and the visible Red band acquired by a simultaneously operated Nikon D200. The image shows the street pattern in the northern part of Potentia and was taken on May 15, 2008 at 11.16 h. .............. 212

9.13 (A) Spectral sensitivity curves of the Kodak Ektachrome Professional Infrared EIR film (adapted from [259]); (B) spectral response of the Kodak DCS-420 CIR (adapted from [348]: Fig. 5.7); (C) the Green and Red channel response of the Nikon D200 (adapted from [712]: Fig. 6b) with the spectral sensitivity curve of the Nikon D50_nir’s Blue channel. The grey zone indicates the NIR domain. ............................................................................... 213

9.14 Visible record (A) and CIR output (B) of the street pattern in the northern part of Potentia. The images were taken on May 15, 2008 at 11.16 h and generated by a Nikon D200 and a simultaneously operated Nikon D50_nir.............................................................................. 214

10.1 Positive (A) and negative (B) crop marks (adapted from [42]: Fig. 13). ......................... 220

10.2 Kodak Ektachrome Professional Infrared image of a dense archaeological landscape containing Neolithic and Roman features ([109]: Fig. 8)..................................................................... 221

10.3 (A) The layout and principal of a Bayer CFA; (B) wavelength versus absolute Quantum Efficiency (QE) for the Kodak KAF-8300 (adapted from [261]: Fig. 5)......................................................... 222

10.4 (A) Dark current (+ bias signal) generated DNs versus exposure time (s) for very short exposures; (B) DNs generated by dark current (+ bias signal) without subsequent median filtering................................................................................................................................. 223

10.5 Relative response versus wavelength of the Nikon D50_nir with a Nikkor 20 mm f/3.5 Al-S (data and illustration by P. Smet – Ghent University).............................................................................. 228

10.6 The Red channel minus the Green channel (R-G), the Blue channel subtracted from the Green channel (G-B), and the Blue channel added to the Green one (G+B); the peak response of each band is normalized to unity (data and illustration by P. Smet – Ghent University)................................................................................................................................................... 230

10.7 Processed images from the same aerial picture taken with the D50_nir: (A) the RAW file developed by Capture NX; (B) the same RAW file linearly developed in dcr aw; (C) a contrast enhanced version of (A); (D) the output after the simple mathematical operation (explained in the text) on version (B).............................................................................................. 232
10.8 Comparison between a conventional photograph (A; photograph by F. Vermeulen –
Ghent University) and three versions of an NIR photograph: (B) is the complete NIR
frame, (C) displays the Blue NIR channel, while (D) visualizes the result of the SR...............233

10.9 Visible image of the central part of the Roman town of Ricina; (B) NIR image of
the same scene with some contrast enhancement; (C) the output when applying the SR
with the channels form (B). Images acquired with a Nikon D200 (A) and a Nikon D50NIR
(B & C). ....................................................................................................................................................................235

11.1 (A) Typical reflectance responses of a radiata pine leaf. The discontinuous (green) curve
of A represents the mean reflectance for healthy, chlorophyll-rich leaves, while the
upper (orange) curve indicates the mean reflectance for a stress situation (N or
nitrogen deficiency). The grey curve represents the STandard Deviation (STD) of
reflectance calculated from reflectance spectra of maize leaves with varying
chlorophyll content (adapted from [145]: Fig. 2B, and [379]: Fig. 2 respectively); (B)
structure of a plant leaf and its interaction with incident visible and NIR radiation
(adapted from [727]: Fig. 32-3 and [748]: Fig. 2). ...........................................................................................................247

11.2 (A) NDVI map of the world, generated from data acquired in April 2002 by the MODIS
sensor [462]; (B) reflectance in two visible bands, the Red edge, and the NIR range
plotted versus varying total chlorophyll content in maize leaves (adapted from [379]:
Fig. 3). .....................................................................................................................................................................250

11.3 (A) Spectral transmission curve of the Carl Zeiss 700 nm interference filter (data by
courtesy of P. Smet – Ghent University); (B) response curves of the same filter with
varying angle of the incident radiation (data by courtesy of P. Smet – Ghent University).....253

11.4 (A) Stereo bar with three DSCs mounted; (B) inside a two-seater aeroplane holding an
older two-DSC rig (photograph by F. Vermeulen – Ghent University). ...........................................254

11.5 Creation of the reflectance ratio image (tile I and H) and comparison with a
simultaneously generated conventional colour photograph (G). By revealing
previously invisible building structures in the central part of the Roman city of Trea (N
43° 19' 05", E 13° 17' 34" – WGS84), tile H indicates the revealing power of this $R_{700}/R_{800}$
index. ......................................................................................................................................................................257

11.6 Visible (A) and $R_{700}/R_{800}$ image (B) of a southwestern part of the Roman town of Ricina
(N 43° 19' 33", E 13° 25' 15" – WGS84). ...........................................................................................................258

11.7 (A) The reflectance image generated by $R_{700}/R_{800}$; (B) the result from $(R_{700})^{-1}$; (C) the
image generated by $[(R_{700})^{-1} – R_{800}]^{-1} * R_{800}$; (D) the same scene imaged by an
unmodified DSC. The imagery depicts the northwestern part of the Roman town of
Ricina (N 43° 19' 37", E 13° 25' 07" – WGS84). ...........................................................................................................259

12.1 An ElectroMagnetic wave (adapted from [806]: Fig. 25-3). .................................................................267

12.2 The ElectroMagnetic spectrum .................................................................................................................268

12.3 The complete UV spectrum and its principal physical and biological sub-divisions. ..................268
List of Figures

12.4 The MUV, NUV, and visible transmittance for a one-centimetre ozone layer at 0 °C and 1 atmosphere (adapted from [45]: Fig. 11a). ............................................................................................................ 269

12.5 (A) Visible (1) and NUV image (2) of Rudbeckia fulgida (photographs by K. Schmitt – macrolenses.de); (B) a lime-encrusted pot with covering slip, depicted by visible wavelengths (1), using NUV fluorescence (2) as well as direct NUV (3) (adapted from [241]: Plate 74-76) ............................................................................................... 271

12.6 The spectral, unfiltered response curves of a Kodak CCD and CMOS imaging sensor, both normalised to their largest response. The discrepancy between the curves in the grey NUV zone indicates the much higher responsivity of the CCD sensor technology in this waveband (adapted from [264]: Fig. 5, and [265]: Fig. 5).................................................................................................................................................. 273

12.7 (A) Wavelength versus Quantum Efficiency for the Kodak KAF-8300 (adapted from [261]: Fig. 5); (B) spectral response of the unmodified Nikon D80 (adapted from [407]). .................................................................................................................................................. 274

12.8 Spectral transmission of several mentioned optical elements. Compare the huge NIR-leakage of the Hoya U-340 filter with the (almost) pure NUV transmission of the Baader Venus Filter. The NIR transmission of the Novoflex Noflexar f/3.5 35 mm can be judged against the transmission of the Zeiss Topar f/2 57 mm, one of the best aerial lenses on the market and exemplary for the NUV transmission of conventional multi-coated lenses. The outcome of both lenses combined with the Venus Filter is also plotted, showing the Noflexar system to be far superior in terms of NUV transmittance (spectral measurements by K. Schmitt – macrolenses.de). ........................................................................................................................................... 276

12.9 Helikite Aerial Photography to acquire archaeological digital NUV imagery (photograph by F. Vermeulen – Ghent University). ............................................................................................................................................. 277

12.10 Comparison between a visible record of a crop mark (left) at the site of Potentia and its rendering by NUV wavelengths (right) .................................................................................................................................................. 279

12.11 (A) Spectral reflectance of several leaf surfaces ([351]: Fig. 2a); (B) absorbance of a vegetation sample in dried and fresh conditions (adapted from [96]: Fig. 2a). .................................................................................................................................................. 280

12.12 Comparison between a visible frame (A), unprocessed RAW NUV image (B), and enhanced NUV photograph (C) of an excavation area at the Roman town of Potentia. The large dark blob in the upper right corner of (B) and (C) is the shadow cast by the Helikite .......................................................................................................................................................................................... 280

12.13 (A) An image of an Airy diffraction pattern (adapted from [846]: Fig. 36.25), while (B) shows the intensity distribution of the Airy pattern, and (C) displays the Rayleigh criterion. ..................................................................................................................................................... 281

12.14 (A) NUV, visible, and NIR reflectance spectra for six different soil samples (adapted from [412]: Fig. 4); (B) absorbance of a soil sample in wet and dried conditions (adapted from [96]: Fig. 2b). ..................................................................................................................................................... 282

12.15 (A) A visual record of two soil types and vegetation, with (C) the NUV reflectance record of the same scene. While (B) represents the contrast enhanced version of (A), (D) displays the processed and contrast enhanced version of (C). ........................................................................................................................................... 284
12.16 Hayward Fault (California, USA) imaged in the NUV and Blue waveband from 600 m altitude ([396]: Fig. 56)......................................................................................................................................285
List of Tables

4.1 An overview of different BAP and/or KAP systems used since the 1960s to acquire low altitude archaeological aerial imagery. The abbreviations in the “Use” column denote Archaeological Prospection (AP), Site Documentation (SD), and Site Documentation and Mapping (SDM). ................................................................. 84

5.1 Helikite models and performance (adapted from [31]) ........................................................................... 95

6.1 Diagonal FOV for three types of photographs.............................................................................................. 128

7.1 Overview of various RAW software.............................................................................................................. 162

7.2 The extensions of some RAW formats. ........................................................................................................... 165

10.1 The appropriate symbols and units of all mentioned DSC quantities......................................................... 225

10.2 All workable spectral bands generated by the D50_{nIR} ........................................................................... 230

12.1 This enumeration lists most expert optics and some glass lenses suitable for direct NUV recording with DSCs (adapted from [690]). ............................................................................................................. 275
Abbreviations and Acronyms

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>Anti-Aliasing</td>
</tr>
<tr>
<td>ADC</td>
<td>Analogue-to-Digital Converter</td>
</tr>
<tr>
<td>ADU</td>
<td>Analogue-to-Digital Unit</td>
</tr>
<tr>
<td>AF</td>
<td>AutoFocus</td>
</tr>
<tr>
<td>AHCI</td>
<td>Arts &amp; Humanities Citation Index</td>
</tr>
<tr>
<td>AHD</td>
<td>Adaptive Homogeneity-Directed</td>
</tr>
<tr>
<td>ANSI</td>
<td>American National Standards Institute</td>
</tr>
<tr>
<td>APS</td>
<td>Advanced Photo System</td>
</tr>
<tr>
<td>ASTER</td>
<td>Advanced Spaceborne Thermal Emission and Reflection Radiometer</td>
</tr>
<tr>
<td>AVHRR</td>
<td>Advanced Very High Resolution Radiometer</td>
</tr>
<tr>
<td>B&amp;W</td>
<td>Black-and-White</td>
</tr>
<tr>
<td>BAP</td>
<td>Balloon/Blimp Aerial Photography</td>
</tr>
<tr>
<td>CAD</td>
<td>Computer-Aided Design</td>
</tr>
<tr>
<td>CCD</td>
<td>Charge-Coupled Device</td>
</tr>
<tr>
<td>CCT</td>
<td>Correlated Colour Temperature</td>
</tr>
<tr>
<td>CFA</td>
<td>Colour Filter Array</td>
</tr>
<tr>
<td>Chl $a$</td>
<td>Chlorophyll $a$</td>
</tr>
<tr>
<td>Chl $b$</td>
<td>Chlorophyll $b$</td>
</tr>
<tr>
<td>CIE</td>
<td>Commission Internationale de l’Éclairage</td>
</tr>
<tr>
<td>CIELAB</td>
<td>CIE 1976 $L^<em>, a^</em>, b^*$</td>
</tr>
<tr>
<td>CIR</td>
<td>Colour InfraRed</td>
</tr>
<tr>
<td>CMOS</td>
<td>Complementary Metal Oxide Semiconductor</td>
</tr>
<tr>
<td>CMYK</td>
<td>Cyan-Magenta-Yellow-Black</td>
</tr>
<tr>
<td>CPCI-S</td>
<td>Conference Proceedings Citation Index - Science</td>
</tr>
<tr>
<td>CT</td>
<td>Colour Temperature</td>
</tr>
<tr>
<td>CWL</td>
<td>Central WaveLength</td>
</tr>
<tr>
<td>DCF</td>
<td>Design rule for Camera File system</td>
</tr>
<tr>
<td>DIN</td>
<td>Deutsches Institut für Normung</td>
</tr>
<tr>
<td>DIP</td>
<td>Digital Image Processing</td>
</tr>
<tr>
<td>DN</td>
<td>Digital/Data Number</td>
</tr>
<tr>
<td>DNA</td>
<td>DeoxyriboNucleic Acid</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>DNG</td>
<td>Digital NeGative</td>
</tr>
<tr>
<td>DOF</td>
<td>Depth Of Field</td>
</tr>
<tr>
<td>DoF</td>
<td>Degrees of Freedom</td>
</tr>
<tr>
<td>DR</td>
<td>Dynamic Range</td>
</tr>
<tr>
<td>DSC</td>
<td>Digital Still Camera</td>
</tr>
<tr>
<td>D-SLR</td>
<td>Digital Single-Lens Reflex</td>
</tr>
<tr>
<td>EGNOS</td>
<td>European Geostationary Navigation Overlay Service</td>
</tr>
<tr>
<td>ELF</td>
<td>Extreme Long Focus</td>
</tr>
<tr>
<td>EM</td>
<td>ElectroMagnetic</td>
</tr>
<tr>
<td>ERIH</td>
<td>European Reference Index for the Humanities</td>
</tr>
<tr>
<td>ETM</td>
<td>Enhanced Thematic Mapper</td>
</tr>
<tr>
<td>ETM+</td>
<td>Enhanced Thematic Mapper Plus</td>
</tr>
<tr>
<td>EWA</td>
<td>Extreme Wide-Angle</td>
</tr>
<tr>
<td>Exif</td>
<td>Exchangeable image file format</td>
</tr>
<tr>
<td>FCIR</td>
<td>False-Colour InfraRed</td>
</tr>
<tr>
<td>FF</td>
<td>Fill Factor</td>
</tr>
<tr>
<td>FIR</td>
<td>Far-InfraRed</td>
</tr>
<tr>
<td>FOV</td>
<td>Field Of View</td>
</tr>
<tr>
<td>FPS</td>
<td>Frames Per Second</td>
</tr>
<tr>
<td>FS</td>
<td>Full Spectrum</td>
</tr>
<tr>
<td>FUV</td>
<td>Far-UltraViolet</td>
</tr>
<tr>
<td>FWHM</td>
<td>Full Width at Half Maximum</td>
</tr>
<tr>
<td>GCP</td>
<td>Ground Control Point</td>
</tr>
<tr>
<td>GIMP</td>
<td>GNU Image Manipulation Program</td>
</tr>
<tr>
<td>GPS</td>
<td>Global Positioning System</td>
</tr>
<tr>
<td>GPS/INS</td>
<td>Global Positioning System/Inertial Navigation System</td>
</tr>
<tr>
<td>GSD</td>
<td>Ground Sampling Distance</td>
</tr>
<tr>
<td>HAP</td>
<td>Helikite Aerial Photography</td>
</tr>
<tr>
<td>HVS</td>
<td>Human Visual System</td>
</tr>
<tr>
<td>Hz</td>
<td>Hertz</td>
</tr>
<tr>
<td>IF</td>
<td>Impact Factor</td>
</tr>
<tr>
<td>IR</td>
<td>InfraRed</td>
</tr>
<tr>
<td>ISO</td>
<td>International Organization for Standardization</td>
</tr>
<tr>
<td>JEIDA</td>
<td>Japan Electronic Industry Development Association</td>
</tr>
<tr>
<td>JEITA</td>
<td>Japan Electronics and Information Technology industries Association</td>
</tr>
<tr>
<td>JFET</td>
<td>Junction Field Effect Transistor</td>
</tr>
<tr>
<td>JFIF</td>
<td>JPEG File Interchange Format</td>
</tr>
<tr>
<td>JPEG/JPG</td>
<td>Joint Photographic Experts Group</td>
</tr>
<tr>
<td>K</td>
<td>Kelvin</td>
</tr>
<tr>
<td>KAP</td>
<td>Kite Aerial Photography</td>
</tr>
<tr>
<td>LAAP</td>
<td>Low Altitude Aerial Photography</td>
</tr>
<tr>
<td>LAB</td>
<td>CIE 1976 L*, a*, b*</td>
</tr>
<tr>
<td>LAI</td>
<td>Leaf Area Index</td>
</tr>
<tr>
<td>LF</td>
<td>Long Focus</td>
</tr>
<tr>
<td>LiDAR</td>
<td>Light Detection And Ranging</td>
</tr>
<tr>
<td>LWIR</td>
<td>Long Wavelength InfraRed</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------------------------</td>
</tr>
<tr>
<td>LWUV</td>
<td>Long Wave UltraViolet</td>
</tr>
<tr>
<td>MLF</td>
<td>Medium Long Focus</td>
</tr>
<tr>
<td>MODIS</td>
<td>MOderate Resolution Imaging Spectroradiometer</td>
</tr>
<tr>
<td>MSS</td>
<td>MultiSpectral Scanner</td>
</tr>
<tr>
<td>MUV</td>
<td>Middle-UltraViolet</td>
</tr>
<tr>
<td>MWIR</td>
<td>Mid Wavelength InfraRed</td>
</tr>
<tr>
<td>N</td>
<td>Nitrogen</td>
</tr>
<tr>
<td>NASA</td>
<td>National Aeronautics and Space Administration</td>
</tr>
<tr>
<td>NDVI</td>
<td>Normalized Difference Vegetation Index</td>
</tr>
<tr>
<td>NEF</td>
<td>Nikon Electronic Format</td>
</tr>
<tr>
<td>NIR</td>
<td>Near-InfraRed</td>
</tr>
<tr>
<td>NMOS</td>
<td>N-channel Metal Oxide Semiconductor</td>
</tr>
<tr>
<td>NUV</td>
<td>Near-UltraViolet</td>
</tr>
<tr>
<td>P</td>
<td>Phosphorus</td>
</tr>
<tr>
<td>PAN</td>
<td>PANchromatic</td>
</tr>
<tr>
<td>PAR</td>
<td>Photosynthetically Active Radiation</td>
</tr>
<tr>
<td>PNU</td>
<td>Photodiode Non-Uniformity</td>
</tr>
<tr>
<td>PRNU</td>
<td>Photo-Response Non-Uniformity</td>
</tr>
<tr>
<td>PVS</td>
<td>Potenza Valley Survey</td>
</tr>
<tr>
<td>QE</td>
<td>Quantum Efficiency</td>
</tr>
<tr>
<td>R/C</td>
<td>Radio Controlled</td>
</tr>
<tr>
<td>RADAR</td>
<td>Radio Detection And Ranging</td>
</tr>
<tr>
<td>RGB</td>
<td>Red-Green-Blue</td>
</tr>
<tr>
<td>RMSE</td>
<td>Root Mean Square Error</td>
</tr>
<tr>
<td>RVI</td>
<td>Ratio Vegetation Index</td>
</tr>
<tr>
<td>S/N</td>
<td>Signal-to-Noise ratio</td>
</tr>
<tr>
<td>SCIE</td>
<td>Science Citation Index Expanded</td>
</tr>
<tr>
<td>Si</td>
<td>Sillicium</td>
</tr>
<tr>
<td>SLR</td>
<td>Single-Lens Reflex</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal-to-Noise Ratio</td>
</tr>
<tr>
<td>SPIFF</td>
<td>Still Picture Interchange File Format</td>
</tr>
<tr>
<td>SPOT</td>
<td>Satellite Pour l'Observation de la Terre</td>
</tr>
<tr>
<td>SR</td>
<td>Simple Ratio</td>
</tr>
<tr>
<td>SSCI</td>
<td>Social Sciences Citation Index</td>
</tr>
<tr>
<td>STD</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>SWIR</td>
<td>Short Wavelength InfraRed</td>
</tr>
<tr>
<td>TFT</td>
<td>Thin Film Transistor</td>
</tr>
<tr>
<td>TIFF</td>
<td>Tag(ged) Image File Format</td>
</tr>
<tr>
<td>TIFF/EP</td>
<td>Tag(ged) Image File Format/Electronic Photography</td>
</tr>
<tr>
<td>UAV</td>
<td>Unmanned Aerial Vehicle</td>
</tr>
<tr>
<td>ULM</td>
<td>Ultra Léger Motorisé</td>
</tr>
<tr>
<td>UV</td>
<td>UltraViolet</td>
</tr>
<tr>
<td>VF</td>
<td>Vegetation Fraction</td>
</tr>
<tr>
<td>VI</td>
<td>Vegetation/Vegetative Index</td>
</tr>
<tr>
<td>VIN</td>
<td>Vegetation Index Number</td>
</tr>
<tr>
<td>VLF</td>
<td>Very Long Focus</td>
</tr>
</tbody>
</table>
Abbreviations and Acronyms

VUV  Vacuum-UltraViolet
VWA  Very Wide-Angle
WA   Wide-Angle
WAAS Wide Area Augmentation System
WB   White Balance
WGS84 World Geodetic System 1984
The Potenza Valley Survey
Archaeological Research Framework

« If we knew what it was we were doing, it would not be called research, would it? »
Albert Einstein (1879-1955), physicist
CHAPTER 1

From Photograph to Plan

Mapping Aerial Imagery in a GIS for Geoarchaeological Research

The content of this chapter is published in Remote Sensing in Transition [CPCI-S]

Abstract

In January 2000, Ghent University initiated the geoarchaeological Potenza Valley Survey (PVS) project in the central Adriatic Regione Marche (Italy). One of the methods employed in this interdisciplinary research is oblique aerial photography from a small, low-flying aeroplane. This method has already proven countless times its enormous value in archaeological survey projects. A major drawback of the generated imagery, however, is the complex geometric distortions incorporated in the photographs. In the PVS, AirPhoto is used as a low-cost software package to rectify the images that are made with hand-held uncalibrated cameras. Due to the varied terrain along the Potenza, different transformation procedures have to be followed. Afterwards, the input of the corrected imagery in a GIS enables on-screen digitizing and an interpretation of all archaeological and geomorphological features.
1.1 The Potenza Valley Survey

During his reign, Rome’s first emperor Augustus (27 B.C.-A.D. 14) divided Italy into eleven regions. More than half a century and eight emperors later – in A.D. 77 – Gaius Plinius Secundus (also called Plinius Maior or Pliny the Elder) wrote his Naturalis Historia. In addition to the knowledge concerning ethnography, anthropology, human physiology, zoology, botany, and mineralogy, Pliny the Elder’s Natural History also contains information about the geography of that time. Glancing through the third volume – in which Pliny deals among other things with the eleven Italian regions – one might read “Quinta regio Piceni est, quondam uberrimae multitudinis…” (Plinius Secundus, Naturalis Historia III.13): “The fifth region is that of Picenum, which formerly was very densely populated…” [643]. It is the first sentence used by this author to describe the fifth region, denominated as Picenum. Nowadays, the northern half of the former Picenum forms the southern part of the Marches: one of the contemporary Italian regions. The northern half of the Marches (Regione Marche) was part of Augustus’ sixth region: Umbria (Figure 1.1).

![Figure 1.1. The Marches: the southern half belongs to Augustus’ region V, the northern half to region VI.](image)

In January 2000, Ghent University (Belgium) initiated a geoarchaeological research project called the Potenza Valley Survey (PVS) in what was once the fifth and sixth region of Roman Italy. This interdisciplinary project was set up by the Departments of Archaeology and Geography under the direction of Prof Dr Frank Vermeulen. The major aim of the PVS was – and still is – to measure the evolution of social complexity within the valley of one of the Marches’ major rivers.

Although Pliny describes a densely populated land, this project wants to discover how populous the region was, what the different types of settlements were, and how the settlement patterns changed, in time as well as in space. Therefore, the project got the subtitle ‘From Acculturation to Social Complexity in Antiquity: A Regional Geo-Archaeological and
Chapter 1

Historical Approach’. To fulfil the initially outlined goals (see [786]), three sample zones, systematically spaced around the river, are focused on (Figure 1.2).

The Potenza – called *Flosis* in Roman times – flows over its ca. 80 km long course from the Apennines through a wide and fertile Apennine foothill landscape before it ultimately runs into the generally flat Adriatic coastline zone. Although it already had some importance in Prehistory, the Potenza river valley only became a major commercial route during the Italic Protohistory. In later periods, the valley remained an important corridor for political, economical, and cultural contacts between the Tyrrhenian and the Adriatic coast. During Roman times, several cities developed in (or near) the valley floor (*Potentia*, *Ricina*, *Trea*, *Septempeda*, and *Prolaqueum* – Figure 1.2) and even a southern branch of the Roman *Via Flaminia* passed through it. In the subsequent period, the area remained of importance as it formed the contact zone between *Longobards* and *Byzantines*. These are only a few of the reasons why a geoarchaeological survey was launched in this particular area.

This research project consists of four important cornerstones: field survey, geomorphological survey, historical survey, and aerial survey. Systematic prospection of the ploughed fields, always carried out in the month of September, is the method used in the field survey. In search for archaeological evidence, **intensive linewalking** was chosen as a prospection method. Therefore, an interval of 10 to 15 meters between different walkers is aimed for. Scholars of the Department of Geography conduct the **geomorphological research** that, amongst many other techniques, consists of **augering** to reconstruct the evolution of the coastline and tackle the problem of alluviation and colluviation. To study known archaeological sites, an intensive historical survey is executed as well. Moreover, some research of toponymical and historical
written information is ongoing. Finally, there is the aerial survey, on which this chapter will focus.

1.2 History of Aerial Survey

Since Joseph Nicéphore Nièpce (1765-1833) invented ‘drawing with light’ in the 1820s, photography can almost celebrate its second centennial. Aerial photography in archaeology covers approximately one half of that time span. Although there was a certain tradition of spying balloons in the nineteenth century – especially during the American Civil War (1861-1865) – it was not until the end of World War I that archaeologically interesting pictures were made from an aeroplane [202]. In this first phase of archaeological aerial reconnaissance, much credit must be given to O.G.S. Crawford. This Englishman is considered to be the inventor of scientific aerial archaeology and his work in the 1920s and beyond (e.g. [199, 200, 201, 202, 203, 204]) was the basis for the future development of the subject.

The major advances achieved and the different archaeological projects initiated in the Interbellum were abruptly interrupted by the second World War. Nevertheless, this period also witnessed an outbreak of significant technical developments. As an additional result of the war, many archaeologists were trained as air photo interpreters.

It was, however, not until the late 1950s that academic research took place and thousands of new archaeological sites were discovered. Accordingly, a growing number of governments began to understand the need to protect the historic environment. After the Iron Curtain was lifted, many countries were finally able to enjoy the benefits of archaeological aerial reconnaissance [80, 834].

1.3 Principles of Aerial Archaeology

In aerial reconnaissance, two methods can be employed: vertical and oblique photography. Vertical aerial photography is achieved by taking a series of overlapping images at regular intervals with a calibrated camera mounted onto an aeroplane and pointing directly down to the earth. The generated imagery can also be viewed stereoscopically. However, archaeologists mostly use hand-held, uncalibrated cameras and simply take oblique photographs at an angle from a low-flying aeroplane. Although this approach is extremely flexible and cost-efficient, the generated imagery suffers from both tilt and terrain distortions. The latter two must be corrected to a certain extent, before any archaeologically interesting feature can be mapped.

Such archaeological remains can be seen from the air in a number of ways. Besides large material remains (e.g. castles, churches, bridges, etc.) and (partly) eroded remains of structures
(earthen banks, ditches, low walls, mounds, etc.), most of the features that can be viewed from above are the remains of buried archaeological sites. While the first type of features is directly visible (Figure 1.3-1), the second type – referred to as earthworks – is often noticed through shadow marks. In the right conditions (i.e. with the sun low in the sky) these archaeological features can reveal themselves by the pattern of sunlight and shadow (Figure 1.3-2). To obtain the best results, the linear earthworks should be at right angels towards the sunrays. The more parallel they are, the weaker the shadows will be.

Soil and crop marks might disclose the buried or levelled remains. Crop marks are patterns of differential growth in vegetation, caused by subsoil variations. Trenches or pits will often be filled with organic material or new soil, having a greater moisture retention and more nutrients than the surrounding matrix. In periods of drought, these humous soils hold the available water for an extended period, allowing the plants to grow longer and fuller (for a more detailed overview of the physical principles and related spectral properties of vegetation marks, consult Chapters 9, 10, and 11). The adjacent plants will be less tall or ripen quicker. These differences in height and/or colour result in patterns that can be seen from above (i.e. positive marks – Figure 1.3-3). The reverse is caused by archaeological features like stone walls or floors. They are characterised by less water retention and nutrients than the undisturbed subsoil, which obviously corresponds to weaker and shorter plants (i.e. negative marks). Although many crops have a high potential to display such marks, the best crop marks are commonly seen in cereals (wheat and barley in particular). It is also noteworthy that the possible height differences in the crops might yield shadow marks as well.

Soil marks are mostly caused by ploughing because this might bring the archaeological subsoil to the surface. Since these deposits often differ from the non-archaeological soil, ploughing

Figure 1.3. Four different kinds of marks in the Potenza valley: 1) ruin; 2) shadow marks; 3) crop marks; 4) soil marks.
often results in patterned colour differences (Figure 1.3-4). The latter can be twofold: darker or lighter than the undisturbed soil. As the refilling of a deepening sometimes contains more humous material with a better water storage capacity, it has a good chance to appear darker than the surrounding soil matrix. If the archaeological feature is not buried too deep, this colour difference might even be visible without the terrain being ploughed.

The opposite happens when dealing with walls, which might generate distinct bright traces. However, soil marks are not only restricted to the soil itself. One might certainly not forget that artefacts are able to generate various soil marks as well, because concentrations of dark pottery or red tiles can easily be seen from the air against a natural background.

Two infrequent kinds of marks are snow and water marks. The former are created by dissimilarities in the temperature of the subsoil. Walls and trenches are always colder or warmer compared to the adjacent ground, causing the snow to melt accordingly faster or slower. Earthworks can create snow marks too, because snow on the sunlit areas will melt the quickest. These marks are typically detected in only a few hours’ time span. By filling the lower parts of the landscape, water marks can be encountered in rainy periods and areas that contain plenty of water.

1.4 Aerial Photography in the PVS

1.4.1 Photography

“Be at the right place at the right time” is valid for most kinds of archaeological marks. Therefore, archaeological aerial reconnaissance should be very intensive. In the PVS, several flights are performed during different seasons each year in order to obtain a representative picture. Additionally, the aerial survey is also very extensive with flights over the whole valley, although a specific interest for the three sample zones remains (for more information consider [787, 790, 794]). Using a small Cessna, the pilot flies at an observation altitude of approximately 300 m. Once a feature is detected, the photographer (in most cases Frank Vermeulen) takes an oblique photograph at an altitude of approximately 100 m to 150 m with a Canon 35 mm Single-Lens Reflex (SLR) camera. The sensitive medium applied is generally Fujichrome slide film (ISO 100). (Note: since 2004, digital cameras have been used – see 1.5).

1.4.2 Inventory

From the beginning of the project, it was the aim to incorporate all gathered survey data into a Geographic Information System (GIS). Until now, a GIS still is the easiest way to manage and analyze large quantities of geographically linked data. However, a whole process needs to be applied before any of the acquired aerial images can enter the PVS GIS.
First of all, every slide gets a unique number which is written onto the slide mount. Afterwards, all slides are inventoried in the PVS database: one major Access database that contains all attribute date of the project.

In addition to the slide number, a whole series of photographic metadata fields have to be completed in the database (e.g. the time and the place of image acquisition, the kind of anomaly photographed, etc.). Because also ground-based photographs are taken, additional attribute information is used to identify whether or not the photograph was obtained from the air and which film, camera, and lens were used for its acquisition.

Thirdly, the best and most unveiling images are digitized and stored on CD-ROMs as well as on the PVS’ central computer. This way, a direct link is enabled between the database and the imagery, while a back-up copy is safely stored elsewhere. Besides their illustrative function, the main purpose of these aerial images is – obviously – to reveal new archaeological sites or confront the already known or discovered features with the archaeological aerial evidence. Secondly, geomorphological marks of interest (e.g. paleofluvial gullies) are looked for as well. Therefore, one needs to map the marks that can be distinguished on the digitized slides as accurate as possible.

1.4.3 Rectification

Because the oblique aerial photographs are made with a hand-held SLR equipped with uncalibrated lenses, only a semi-quantitative result is attainable. Therefore, there is no use in obtaining a highly accurate photogrammetry package. Presently, at least two low-cost programmes exist that are specifically designed to correct oblique images for tilt distortion (i.e. rectification) and incorporate them into an existing coordinate system (i.e. georeferencing): Irwin Scollar’s AirPhoto and John Haigh’s AERIAL. Because AirPhoto has more functionality, the PVS team purchased Scollar’s software.

AirPhoto has the ability to create (pseudo-)orthophotos from extreme oblique images in an economic way [698, 699]. The package offers several algorithms to geometrically align the image with a topographic map. Besides the well-know polynomial transformations (which are, however, less useful for frame images acquired by photographic cameras), the multipoint, the projective, and the Fischler-Bolles algorithms are offered. Although it is not a photogrammetric plotting programme sensu stricto, the latter algorithm can generate true orthophotos if a Digital Elevation Model (DEM) is provided. Even though altitude information can be incorporated, the other methods do not really require them. Furthermore, the package offers three interpolation methods, calculates the mismatch between the source and the target image (i.e. a control of the goodness-of-fit of the rectification), incorporates a lot of image processing techniques, and supports output to different coordinate systems and GIS formats.

The ability to work with different transformations is very useful when one is working in different topographic regions, as is the case in the PVS. Because flat terrain does only introduce
tilt distortion into the photograph, a simple **projective transformation** can be applied. This transformation only requires four corresponding points (i.e. **Ground Control Points** or GCPs) that are easy to define in both the source (i.e. photograph) and the target (i.e. topographic map or orthophoto) (see Figure 1.4). The accuracy of the transformation can be improved by adding more control points.

![Figure 1.4. The projective transformation.](image)

For undulating terrain, the **Fischler-Bolles image rectification** is used, a method generally unknown by photogrammetrists as it was developed in the scientific field of computer vision. This option only requires three corresponding GCPs. Furthermore, some camera parameters need to be filled in: the **focal length** of the lens (see Chapter 6), the width and height of the film frame (or digital sensor), and the **principal point** (i.e. the intersection of the optical axis of the lens system with the focal plane). The latter is accurately determined in the case of a calibrated lens, but working with a normal reflex camera necessitates some estimation. For convenience, the geometrical centre of the image is chosen. Finally, at least five heights not too far from the GCPs are needed to apply the transformation. As all topographic maps (scale 1:10 000) are digitally stored as vector files in the PVS GIS, it is easy to calculate the necessary DEMs (Figure 1.5), a process described in [775]. In the end, this procedure corrects the aerial imagery for both terrain and tilt distortions, hence yielding an orthophoto (i.e. a photograph adjusted for several distortions such that it can be considered a map with a uniform scale).
1.4.4 Mapping and Interpretation in a GIS context

Once the most interesting images are rectified, they are opened in ESRI's ArcView 3.2 GIS software. In this digital environment, all the photographs are displayed in their right geographical location because they were georeferenced during the rectification process (more in particular, they all were tagged with Italian Gauss-Boaga coordinates). This enables a rigorous comparison of all archaeological features visible in different photographs. Moreover, the georeferenced imagery can serve as a background base layer for the various vector layers which hold the on-screen interpretations. To make the features become more distinct, some image processing might be advisable (the most common routines are also available in AirPhoto – Figure 1.5). The possibility GIS offers in adding other digital information layers such as hydrological, geological, and field survey data seriously helps in the interpretation of the imagery, hence increasing this process’ reliability. Since the topographical map is on a scale of
From Photograph to Plan

1:10,000, this mapping method is adequate for the production of 1:2500 (and smaller) archaeological maps (Figure 1.6).

In search of Potentia and the suburban zone

<table>
<thead>
<tr>
<th>Known archaeological structures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temple of Jupiter</td>
</tr>
<tr>
<td>Roman cemetery</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Features discovered by aerial survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roman urban road</td>
</tr>
<tr>
<td>Supposed Roman urban road</td>
</tr>
<tr>
<td>Roman suburban road</td>
</tr>
<tr>
<td>Supposed Roman suburban road</td>
</tr>
<tr>
<td>City wall</td>
</tr>
<tr>
<td>Supposed city wall</td>
</tr>
<tr>
<td>Prospected field</td>
</tr>
<tr>
<td>Recent road</td>
</tr>
<tr>
<td>Railway</td>
</tr>
</tbody>
</table>

Figure 1.6. The mapping of the Roman town Potentia and its suburban area.

1.5 Field Campaign 2004

During the fifth PVS September campaign, total station height measurements were obtained. This additional information allows the geometric correction of even the most difficult
photographs acquired over very undulating terrain. Moreover, the precision of these data allows the generation of interpretation maps on a larger scale (e.g. 1: 250).

Since 2004, the aerial reconnaissance data have been recorded by means of a digital photographic camera (Canon EOS 300D and various Nikon still cameras – see next Chapters). As one can imagine, this new hardware has significant consequences: no film and processing expenses, no extra costs in digitizing images, the freedom to take as many pictures as wanted without needing to worry about the number of pictures that is affordable, an abundant amount of accurate metadata information that is simultaneously stored with the photograph (such as focal length, shutter speed, and aperture – see Chapter 7), instant display of the image, the possibility to shoot on different types of ‘film’ just by adjusting an internal setting, etc. Even though archaeological aerial reconnaissance may be one of the oldest prospection methods, these new digital products certainly have their consequences.

1.6 Prospects

As the flights are ongoing, new features (archaeological as well as geomorphological) are regularly discovered. Therefore, rectification and interpretation is a continuous process. At a certain stage, however, these mapping results have to be confronted with the information yielded by the various field surveys. On the one hand, the images might give a better idea of the ‘real’ proportions of sites (like the Roman city of Potentia in Figure 1.6), while the sherds and pieces of marble and glass can on the other hand tell something about the site’s internal organisation and give a temporal delineation of the features. In Chapter 2 and Chapter 3, this approach is clarified by several examples.

Besides, this highly significant data can (and will) be combined with the geomorphological and historical information obtained in all three survey zones, just to make the whole settlement process easier to understand and the interpretations more reliable. This way, the capabilities and advantages that GIS and oblique aerial photography offers will be completely expressed in the PVS project.

1.7 Conclusion

Oblique aerial photography is already an old and established archaeological prospection technique. In the Potenza Valley Survey, this method has once more proven to give much additional information that otherwise might be overlooked. As the photographs are made without calibrated lenses and metric cameras, it is possible to utilize specifically designed low-cost software which allows archaeologists (and others) to rectify the data with ease, without being a specialist in photogrammetry. One of these software packages is called AirPhoto.
Because this software offers many useful transformation and interpolation functions, it enables working in a topographically varied research area, as long as a DEM is available for the undulating terrains. Once transformed, the incorporation in a GIS allows the imagery to be combined with other spatial data sets, making the on-screen digitizing and interpretation of the georeferenced imagery more reliable and the analysis of all data layers a more fruitful process.
CHAPTER 2

Roman Cities along a Valley

Aerial Photography and Field Survey to the Study of Urbanization

The content of this chapter is published in Journal of Roman Archaeology [ERIH (A)]

Abstract

This chapter is a contribution to the study of Roman urbanization in a part of the central Italian region of Marche. It presents results of ongoing systematic survey by a team from Ghent University in one of the valleys of ancient Picenum, the valley of the river Potenza. An intensive, non-destructive survey approach combining active remote sensing and regular gridwalking is particularly highlighted here. The possibilities of archaeological aerial reconnaissance on so-called ‘unsuccessful cities’ in Adriatic Italy are assessed. We present current knowledge concerning the main Roman towns in this area and synthesize new views on their topographic development. This new information is integrated in a view on processes of urbanization in a still understudied region of the peninsula.
2.1 The Potenza Valley Survey

Systematic survey work has been conducted since 2000 in the central Italian region of Marche. This multi-disciplinary, geoarchaeological project, organised by the University of Ghent, aims to study changes in the landscape and occupation patterns in the period 1000 B.C.-A.D. 1000 throughout the Potenza river valley (Figure 2.1). The intensive survey of the valley embraces full aerial photographic coverage of 400 km² between the Apennines and the Adriatic, and systematic fieldwalking in three carefully selected sample areas in which we also assess the differentiation of landscape types and their influence on human settlement systems [791]. These landscapes comprise, amongst others, the narrow Apennine valleys, the woodlands and higher grasslands of the mountain areas, small intermediate basins, the undulating rich agricultural land of the middle valley and the lower hill slopes and widening coastal plain near the river mouth. Here the Potenza stream lies only some 15 km south of the dominating and attractive coastal promontory of the Monte Conero, near Ancona. This interesting location partly explains why the Potenza valley was, during the whole period concerned, an important corridor for political, economical, and cultural contacts between both the Adriatic and the Tyrrhenian sides of the peninsula.

The objectives of the project and preliminary field results have been published in BABesch [790, 792, 794]. Here we will present some of the more striking results relating to the evolution of
concentrated human settlement towards an urban society with greater social complexity, and to the subsequent breakdown of this urban pattern.

The field methodology was designed with the focus on urbanization particularly in mind. The three sample zones chosen for intensive and systematic fieldwalking (each between 10 km\(^2\) and 25 km\(^2\)) were located in light of supposed or known major settlements along the Potenza corridor (Figure 2.1). It is clear that these centres could in Protohistory (with a series of hilltop sites) or in Roman times (with at least four real valley cities) have generated more complex settlement systems. These areas were most likely to demonstrate the evolution of social complexity and the effects of political and cultural change on the concentration of human settlement.

Beside systematic **linewalking** in order to map and study all possible forms of (especially rural) settlement in the three zones concerned, research by the Ghent team in the field has started to focus also on the existing and formerly located Roman town sites in the valley. Four sites in the Potenza corridor had an urban character during most part of their Roman history; they are from coast to mountains: *Potentia*, *Ricina*, *Trea*, and *Septempeda*. All four of them were during late antiquity unsuccessful in some way and were then, like almost half of all Roman towns in *Marche*, completely abandoned [24].

As no or almost no present-day habitation covers these towns and as agricultural practice still dominates the areas where their ruins lie buried, they can still be studied with the use of common non-destructive archaeological survey methods. Since it is no longer feasible to carry out excavations of such a magnitude [251], research on ancient cities increasingly relies on non-destructive research techniques [602], such as aerial photography, intensive surface surveys, and geophysical prospections, or a combination of these techniques (e.g. [84, 434]).

Since 2000 we have conducted regular aerial photography in the valley from a low-flying aeroplane. Especially intense flying during the dry spring of 2003 has produced remarkable images, visualizing many aspects of the urban topography of three of these towns, namely *Potentia*, *Trea*, and *Ricina*. Although the mapping and especially the interpretation of the structures on these oblique images is still in progress, we can already present here some preliminary results of this work.

The results from this type of remote sensing have also induced us to focus part of the regular field survey on the urban sites concerned. Therefore, a specific programme of **gridwalking** was planned on the three cities mentioned. This is not only necessary to control the remotely sensed images, but it also allows a more detailed study of chronology, functional zoning, and spatial development at these centres.
2.2 Centralized Settlement before the Romans

The middle and upper river valleys of Marche saw a different history of settlement and urbanization to that found in the more prosperous and politically active regions of Latium and Campania, where urbanization was already fully developed during the 2nd century B.C. (cf. [604]) thanks to those regions’ fertility, the efficient exploitation of farmland through new methods, and the wealth acquired by the local élites through their involvement in Rome’s expansion. It resulted there in the construction of urban public buildings financed from municipal resources or individual benefactions. By the 1st century B.C. many of those towns had already embarked on the process of rivalry between cities, which was to prove an important feature of municipal life under the Empire and a real motor for the further process of urbanization.

In large parts of Picenum, with the exception of parts of the coast where Rome’s colonial involvement may be seen during the 3rd and 2nd centuries B.C., the situation was very different. Picenum is a land of rugged mountains – the destination of transhumant flocks and their shepherds – and hilly country cut by small but fertile plains, which could best be exploited by small-scale nucleated settlements which visually control the areas. These settlements are known as vici or pagi. They were the centres of administration in these naturally separated districts, the home for members of the free peasantry (also living in small dispersed sites around these nuclei), and the focus of the ambition and generosity of the local élites. Also important in Piceni culture were rural sanctuaries, which acted as regional cult centres, although not many are yet known.

As only few settlement structures have been discovered till now, most knowledge about Piceni and eastern Umbrian cultures in Marche derives from funeral sites. Although some of the richer cemeteries give good indications about elite concentration in certain areas favoured by natural dispositions (good agricultural land, dominating hilltops, river crossings, a wide view on the sea, etc.), the location and features of elite and/or concentrated settlement of these peoples are still ill-known [105, 786].

The ongoing systematic field surveys of our team and a re-assessment of old finds is beginning, however, to produce some insight into the protohistoric settlement system of the area before the start of slow Romanization during the 3rd and 2nd century B.C. It seems that the evolution of society towards social complexity and some weak form of proto-urbanization had its roots in the later Bronze Age. Several hilltop sites dispersed along the valley became the focus of settlement concentration at least from the 12th century B.C. onwards. In the course of the Iron Age, these sites, now more and more dominated by elite groups within Piceni society, long remain crucial for the control over the valley. A certain topographic continuity of most (if not all) of these hilltop sites until the later Iron Age can no longer be excluded.
The fieldwork campaigns demonstrated clearly the importance of these relatively small hilltop sites. Inland sites such as the Monte Primo near Camerino, the Monte Pitino near San Severino Marche and especially the Monte Franco at Pollenza seem to be the best exponents of this phenomenon. These sites, although different in altitude, extent and general appearance, all have a search for a defensible environment and for an ideal location with respect to the control over movement in the Potenza plain in common. Available archaeological data seem to support their long life as centres for elite power, as well as their use for hierarchical control over a partly concentrated population.

The topographic position of the upper valley site of Monte Primo and the character of the finds lead some scholars to the conclusion that we deal here essentially with a cult place, originally situated on the very top of the Monte Primo [101]. The selection of this place for a sanctuary-like destination is not arbitrary. The Monte Primo dominates the crucial and narrow passage of the Potenza river valley through an Apennine gorge and is an excellent spot to observe bird migration, a very important religious activity in protohistoric Italic culture. It seems, however, very likely that it was also the place of at least seasonal concentrated settlement activity in the transition phases of the Bronze and Iron Ages.

The attraction of the summer grazing grounds for pastoral activities, as well as its control function for an emerging elite of society, are both particular assets in this respect. This role as a settlement centre is certainly sustained by the discovery of an intricate system of earthworks surrounding the c. 4 ha large site [101, 790]. Further fieldwork and particularly excavations are needed to narrow down the chronology and character of this important hilltop occupation site.

The sites of Monte Pitino and Monte Franco, both in the middle Potenza valley, are since long known by their rich early Iron Age (9th-7th centuries B.C.) burial grounds (e.g. [489]). Our surveys in the area of the latter have, however, produced new data on the long life and particular richness of the inhabited zones in the later Iron Age (6th-3rd centuries B.C.), immediately preceding the first Roman influences in the Treia area [794]. The discovery of an intensively occupied area near the eastern flanks of the promontory, displaying fine regional wares as well as a good selection of imports from southern Italy and the Aegean world, underscores this. It seems that these hill sites have continued to play an important role in Piceni society, especially as motors in the newly established exchange patterns with Greek merchants in the coastal area.

From the 6th century onwards, the role of the Greek emporion at Numana and the progressive cultural contacts via the Adriatic Sea, seem to speed up a process, which will ultimately transform Piceni society. Although this did not lead to a fast development into an urbanized development until the coming of the Romans, it probably was a factor of some importance in the development of socially more complex structures.

This should be best visible in the coastal area, near the mouth of the Potenza. Here our 2002 and 2003 surveys showed an interesting pattern of occupation [792]. It was established that
both hills, bordering the valley and almost touching the coastline, attracted settlement in protohistoric times. In the south, at Monte dei Priori near Potenza Picena, possibly only a smaller Bronze Age occupation occurred. In the north, on the high plateau of Montarice (Porto Recanati), our fieldwalking and aerial photography surveys revealed the existence of an important Late Bronze and Iron Age centre, just north of the actual Potenza river mouth.

This site of Montarice, situated on a plateau of some 4.2 ha, was first studied from the air, revealing the soil marks of the extension of settlement all across the plateau. Crop marks photographed in different seasons indicate a huge number of settlement structures, especially the traces of enclosure walls and ditches, accentuating the natural defence of the site during several phases of occupation. The aerial imagery acquired during the extreme drought of April-May 2003 even suggests an almost organised aspect of this imposing site, showing clearly several lined and/or grouped houses and other structures like pits and possible cisterns (Figure 2.2).

During the September 2002 field campaign the Ghent team undertook a detailed intra-site surface survey of Montarice (Figure 2.3). This revealed the presence in the ploughed field of great numbers of protohistoric coarse wares, Piceni buccheroide wares, imported south Adriatic ceramics, and even Greek pottery. Although this material is still being processed in view of dating and intra-site dispersion analysis, we can already state that the high density of the pottery finds and a first evaluation of the chronology supports the idea of a long and
almost continued occupation of the hill from the middle Bronze Age into the late Iron Age and
even the Early Roman period. The good quality of the wares, especially of the Greek (black
glazed and black- and red-figured pottery) and southern Adriatic imports (e.g. Daunian and
Messapian), points no doubt to the presence of Piceni elites, probably controlling maritime
transports and contacts with Numana and other commercial centres along the coast. In fact,
the site fits between old river mouths and neighbouring protohistoric altitude settlements
which are observed along the middle Adriatic coastline and can be connected with the
maritime commercial routes of the Greek merchants [48, 498]. The pre-urban hill site certainly
had a role to play in the control over the entrance to the river corridor and over the flow of
goods to inland sites (and ultimately to the Tyrrhenian area). There are, however, no clear
indications yet that the relatively small site was eventually turned into a centre with real urban
allure, as should be deduced from the presence of a central square or important public
building infrastructure.

Figure 2.3. Results of a systematic grid survey in September 2002 on the site of Montarice.

2.3 Roman Urbanization in the Coastal Area: Potentia

Rome’s military campaigns of 269-268 B.C. in the Adriatic region profoundly changed the fate
of the Piceni and partly Celtic and Umbrian populations living in what is now Marche. The slow
but profound Romanization process that followed brought major changes in the social and
economic tissue of the society, including the real introduction of urbanization and at least a
partial re-organization of the countryside. The immediate consequences of the Roman victory
over the region were quite drastic: parts of the territory were confiscated and groups of Piceni
were deported to southern Italy. Yet this did not provoke a complete breach with the pre-
Roman Iron Age. The majority of the population was soon incorporated in the Roman state,
first with an incomplete citizenship, from 241 or 233/2 B.C. onwards with full rights. The installation of the Latin colony at *Firmum* in 264 B.C. was a major impulse for the Romanization of a region that, apart maybe from the centre at *Asculum*, knew no real urbanized society [136, 226, 551]. This progression was fiercely interrupted in the last decades of the 3rd century B.C. by the incursions of the Carthaginian armies, who used *Picenum* as a base for their attacks on the Roman State. The foundation in central *Marche* of the colonies of *Potentia* and *Auxinum* during the first half of the 2nd century B.C. meant, however, a new impulse for the Romanization process. In the lower Potenza valley the Roman impact became thus very visible from 184 B.C. onwards. With the foundation of the colony for Roman citizens at *Potentia*, a whole series of foundations of maritime colonies on the Adriatic coast, which started shortly after the battle of *Sentinum*, was finalised.

The urban site of *Potentia* was first located by Nereo Alfieri on a beach ridge a few hundred meters south of the river Potenza (ancient *Flosis*), immediately south of the present-day coastal town of Porto Recanati [23, 25]. Since this discovery in the 1940s important archaeological work was done on the site. Rescue excavations in the 1960s and 1970s have revealed parts of its northern cemetery and elements of a housing sector in the northeastern corner of the town [526, 528]. A study by Moscatelli of good vertical aerial photographic data (e.g. RAF-pictures of WW II) revealed many indications of the town’s regular street grid [552], while Paci produced a bibliographical synthesis and analysis of all known monuments and inscriptions [591].

Since the mid-eighties small scale excavation campaigns, under the direction of Edvige Percossi, have been organized on a regular basis in the monumental centre of the city. The *Soprintendenza Archeologica per le Marche* discovered here a republican temple, surrounded by a portico and other buildings of republican and imperial age [610, 611, 612]. The excavations show a great vitality of the town during the late Republic and early Empire, not really interrupted by an earthquake that hit the area in 56 B.C. According to the date of the portico surrounding the original temple and of a macellum north and a luxurious building east of this probable capitolium, the city centre was further monumentalized in the time of Augustus. A flourishing period under Trajan and the Antonines was, however, followed by the 3rd century crisis. During the economic revival of the late 3rd and 4th century, the central area was reorganized and most finds (e.g. the coin series) suggest a positive atmosphere until the beginning of the 5th century, followed by a clear decline. The latest archaeological finds belong to the 7th century, but the exact character of the occupation in the town at this period is unclear.

The results of our remote sensing operations represent no doubt a further contribution to the knowledge of this late republican colony. Although the site is currently crossed by the wide coast road and adjoining railway, two modern houses cover part of the ancient town, and grass vegetation of a segment of the archaeological area protected by law hinders aerial observations, conditions today are still fairly good for aerial photography. Since 2000 regular flying over the city site, especially during spring and early autumn, has produced a good set of images, most of which contain information that was not available before. Soil marks in the
ploughed soil on oblique photographs made in late September/early October helped to better understand the geomorphological situation, such as the location of the whole Roman city on a narrow ancient beach ridge, parallel with and at some distance inland of the current coastline. This beach ridge is cut to the south by what can now be determined as the main ancient river bed in Roman times. Detailed mapping of the ancient coast plain by means of augering and earth resistance measurements, in process under the direction of M. De Dapper, should further elucidate the exact position and immediate surroundings of Potentia during its long history [792, 798]. It should also clarify the precise role of river floods, alluvial activity, and sea level change in the whole process of gradual abandonment of the town during late antiquity.

Figure 2.4. New plan of Roman Potentia and its immediate surroundings, based on aerial photography, systematic survey, and earlier excavations.
Soil marks and many good crop marks observed in fields of grain during spring have helped to locate the street network and town walls. Although the older vertical photographs already revealed much of the regular town plan [552, 613], many new elements can be added, and older ones corrected or observed in more detail. Most of the streets are distinguished by very straight and clear pale linear features which probably mark stone covered surfaces with an average width of 5 m. In the eastern part of the city a street feature is somewhat wider and exceptionally bordered on one side by a thin darker trace, probably a shallow ditch or sewer. We also see here the crop mark of a street with darker colouring, lined on both sides by a thin pale trace: possibly an un-metalled street lined by rows of stones. The town wall is suggested by a less clear cut, wider pale crop mark of some 8 m wide, probably the trace of a destroyed city enclosure.

Centrally placed in the northern and southern short side of the rectangular town's enclosure we could distinguish wider crop marks, probably indicating the positions of the northern and southern gates. Thanks to the computer rectification of these obliquely photographed traces of streets and city wall with AirPhoto (see Chapter 1) as well as GIS analysis incorporating vertical images and existing map information, it is now possible to produce a fairly consistent map of the city grid (Figure 2.4 and Figure 2.5).
Figure 2.5. Plan of *Potentia* with the best known building sectors: 1) parts of two *insulae* excavated in the 1970s; 2) central temple sector currently under excavation; 3) buildings seen in crop marks during aerial reconnaissance in 2003.

Current data suggest a quite regular, rectangular town grid measuring 15 *actus* by 10 *actus*, which is some 525 m by 350 m or almost 18.4 ha, measured *intra muros*. If we include the walls in the town’s surface the figure is 19.1 ha (note that this size is almost identical to that – 19.3 ha – presumed for the Adriatic colony of *Pisaurum*, founded in the same year as *Potentia* [728]. The town is oriented along a longitudinal axis with a north-northwest–south-southeast orientation, no doubt chosen parallel with the former and current coastline. The streets (Figure 2.6), which according to Livy were laid out in 174 B.C. together with the walls and the temple of
Jupiter, subdivide the urban space into *insulae* of different size. The width of these city blocks, not incorporating the streets, is fairly consistent and measures some 35 m or exactly 1 *actus*. This system is close to that proposed for the republican colony of *Sena Gallica* further north on the Adriatic coast [215].

The length of the *insulae* is less constant. The row of *insulae* immediately west of the *cardo maximus* seems quite strictly modulated: these blocks have an inner length of 2.5 *actus* (c. 87 m). The other *insulae* have more varying lengths, with an average around 1.75 *actus* (c. 61 m). It is still difficult to calculate the exact number of *insulae* of *Potentia* as it is not sure that the peripheral areas, bordering the city wall *intra muros*, were (initially) all subdivided in blocks. If we do not count this peripheral space, which was possibly only built up in a later phase of the life of the city, then a maximum of 33 *insulae* were designed here.

Figure 2.6. Aerial reconnaissance in 2003 presented a clear view of the streets and circuit wall of the northern part of *Potentia*. Remark differences in width and in nature of the street traces (photograph by F. Vermeulen – Ghent University).

The *cardo maximus* lies more or less central and is connected with the two gates, north and south out of the city. This main *cardo* was probably almost centrally crossed by the *decumanus maximus*, an east-west oriented main street whose trace is still difficult to detect fully on aerial photographs. It was probably also connected with two gates, west and east, although the existence of a *porta marina* is still uncertain. Obscure is also the precise location and configuration of the *forum*. One would expect it to be situated immediately southeast of the crossing of the two main city streets, but this area does not show well from the air as it now lies under the modern coast road and railway. Furthermore, excavations of the temple area have
demonstrated that immediately north of the temple, in its axis, a *macellum* was positioned, which makes the location of the *forum* square in that area somewhat unlikely [613]. The *forum* was perhaps not centrally placed in the town (e.g. [152, 182, 675, 728]) and a position south of the southwards oriented *capitolium* would then be more appropriate. In that case a position central in the southern half of town, in accordance with the incoming coast road (which determined the longitudinal axis of the city plan), would be most acceptable. In this respect, the position and layout of this town reflects the situation of the typical maritime colonies known along the Tyrrhenian coast, built as bridgeheads for land and sea routes and spatially organized along the coast road (e.g. *Via Aurelia Vetus* built in 241 B.C. [182]). A similar system must have functioned along the Adriatic during the early colonisation phase of this side of the peninsula.

Our aerial survey has not (yet) succeeded in producing much evidence about the town buildings and houses of *Potentia*. Fragments of buildings are visible on several images, often showing parts of structures neatly lined according to the street grid. This is especially the case near the eastern and southern fringes of the city. In the latter, near the southern gate, we can distinguish parts of a major building complex (*thermae*) with rectangular and rounded rooms and probably with some well-preserved floors, indicated by clear cut rectangular colorations in the crops (Figure 2.5, 3). In different areas, pits, and other irregular structures can be recognized, showing the evident intense use of most city quarters. For the further detection and identification of housing and public buildings a campaign of geophysical survey is planned. This must also produce more information about the density of occupation of the colony. The literary sources do not mention the exact numbers of colonists that settled here in the 2nd century B.C. It is only known that each new family received a piece of land of 6 *iugera* to cultivate [593]. The colony was certainly more populated than most small *coloniae maritimae*, where normally only some 300 colonists lived. A figure between 2000 and 3000 republican inhabitants, as presumed for the republican colonies of *Parma, Mutina*, and *Cosa*, is more realistic [593].

In 2002 the PVS team started with intensive and systematic fieldwalking in and around the town of *Potentia*. That part of the wide urban area, *intra* as well as *extra muros*, was subdivided in regular units. Large samples of datable ceramics, building materials, and other artefacts were collected in a systematic way (see [792]; Figure 2.7). The first available distribution maps, showing artefact density in and around the presumed habitation centre, already reveal the differences in occupation density in several sectors of town. Their chronological and functional significance will be further analysed once the pottery studies are finalised. In general, however, the chronological span of the town’s life, from the early 2nd century B.C to the early 7th century A.D., was confirmed by the pottery finds [549]. Several tracts of the street grid, known from our aerial photographs, were identified on the ground, mostly as clear concentrations of river pebbles. The survey results were also indicative for the precise localisation of the northern and southern gates in the circuit wall, both built mostly with regular blocks of limestone.
The surveys on arable fields surrounding the urban site of Potentia were most relevant for our knowledge of the suburban and rural hinterland of this town (Figure 2.4). Immediately to the north of the colony, a possible extra muros settlement area could be distinguished. It borders the Roman coast road identified on our aerial photographs. Further north, this road was – according to the remote sensing – lined with at least two funerary monuments of which we found clear surface indications in the shape of many fragments of worked limestone blocks. An archaeological dig started in the beginning of 2004, at the same moment this text was written. The Soprintendenza per le Marche, who is responsible for this work, has indeed already confirmed the existence of at least six imperial funerary monuments lining the coast road to the north. Excavations in the 1960s and 1970s had already indicated that this area was further extended to the east with a large cemetery, used between the 2nd century B.C. and the 4th century [528].

A Roman road, which we discovered from the air in 2000, leaves the presumed southern gate in a southwestern direction to the ruins of a Roman bridge at Casa dell’Arco [483] and was recognized in the field. Our images show that it is bordered by at least five funerary monuments of which we found clear surface indications: fragments of architectural blocks of marble and limestone, and of some fine and common early imperial pottery. Immediately north of this road, which possibly linked Potentia with the nearby towns of Pausulae and probably Urbs Salvia, we located several suburban extramural settlement zones. Finally, also along the Roman coast road leading from Potentia to the south (which we discovered from the air in May 2002), we now traced a few suburban settlements. Their exact chronology still needs to be established.

All data assembled in and around Potentia so far, from excavations as well as from our recent surveys, indicate that from its early 2nd century B.C. start this colony had a fully developed
urban pattern. Its important political and economic impact on the area of the valley mouth must have radically drawn the pre-existing Piceni into the Romanization process, with a transfer of the local elites and power from the nearby hill of Montarice to the Potenza river plain. The discovery during our recent field surveys of several late republican villae in the hinterland of the colony concords well with this image of shifting powers and economic transition [773]. But although the local elites and populations must have had a role to play in this early process of full urbanization, the Latin and Roman colonies erected on the Adriatic coast between 289 B.C. (Hadria, Sena) and 184 B.C. (Pisaurum, Potentia) were no more than rather brusque and artificial urban creations of Rome’s politico-military machine. They did not have an immediate impact on the deeper inland territories of Picenum.

2.4 The Inland Municipia of Ricina and Trea

Apart from these ‘Roman islands’ and their respective territories, essentially located in the coast area, there was probably not much change visible in the rest of the conservative inlands of Picenum until the 1st century B.C. Yet as a result of the Lex Flaminia de agro Gallico et Piceno viritim dividundo of 232 B.C. a major reorganization of this part of the peninsula was initiated, probably with the creation of new administrative, political, and social centres, not necessarily of urban character. Together with the existing colonies on the coast, a network of newly founded inland praefecturae was established to fulfil the Roman desire of organisation. These establishments (archaeologically still difficult to trace) would become the core of further urbanization of the region. As elsewhere in Italy, the end of the Social War in 90 B.C. – which brought full citizenship to many Italic people – was a strong impulse for administrative reorganisation and Romanization of the region. The widespread urbanization of major parts of Picenum and the ager Gallicus only occurred during the second part of the 1st century B.C., when the ‘imperialistic’ system of praefecturae was abolished and a whole series of municipia were developed, that is real towns with their own territory and administration. Roads were further developed, such as several branches connected with the Via Flaminia in the northern and central area of Marche and with the Via Salaria in the south. Other important developments in that period are the assignation of land to veterans of the armies of Caesar, Marc Anthony, and later Augustus within the territories of municipia (such as Tolentinum, Truentum, and Urbs Salvia) and the foundation of new colonies at Ancona (losing its state of independent city), Firmum, Asculum, and Falerio [226, 590, 592].

The inland part of the Potenza valley was no exception to this general development. The archaeological data, which start to emerge from our recent surveys and from a re-evaluation of older finds, indicate that until the beginning of the 1st century B.C., towns were not of major importance. When the Italian allies of Rome acquired Roman citizenship as a result of the Social War, the elites in many parts of Italy began to transfer their attention to the city of Rome and to political competition there. During the 1st half of the 1st century B.C, this must have
resulted in Picenum in a gradual decline of the vicus-based system and to some urban development in many parts of the interior. It is precisely in that period that we see the gradual rise of the three other cities in the Potenza valley, the municipia of Septempeda and Trea in the middle valley and of Ricina in the lower plain. One or two of these, like Septempeda and maybe Trea, were possibly already vici within the Piceni system, while the settlement of Ricina, located on the immediate left bank of the Potenza near an important river crossing, was perhaps a relocation in a better suited area where full advantage could be taken of the economic possibilities of the valley road and fluvial transport system. During the period of the second triumvirate (43-33/32 B.C.) new territories for Romans were assigned in Picenum. Among those, which were mentioned in the so-called Liber Coloniarum, are the four towns in the Potenza valley: Potentia, Ricina, Trea, and Septempeda together with the nearby inland centres of Pausulae and Tolentinum (river Chienti) and Urbs Salvia (river Fiastra). This no doubt gave a new impulse towards more urbanization in this region.

This gradual process was further strengthened under the Principate of Augustus. He sought to create a systematic organisation for rural Italy: the division of the whole country into municipal territories and pagus districts was but one aspect of this scheme. One effect of this was that it distracted the attention of the municipal elites from the former vici towards the municipia, where they would be able to compete for office and prestige, eventually aiming at a career (as senator) in Rome itself. The elites accordingly began to build in the newly created municipia and neglected the old pagi and vici. By the early 1st century A.D., the Piceni municipia were beginning to acquire the facilities which their counterparts in Latium or Campania had gained a century earlier. As elsewhere in Marche, there are enough indications from surviving monuments and inscriptions that the growing wealth of the local elite and the gradual growth of the Potenza valley towns as central places are responsible for an almost continuous prospering of municipal life during the 1st century of our era (see [551, 590] for a good synthesis on the situation in the Marche). The towns are provided with public infrastructures (e.g. theatres, temples, aqueducts, and public baths). Created originally as artificial administrative centres, they became flourishing centres of population, and hence of marketing and exchange.

Our archaeological survey of these three urban sites is still in its initial phase. Nevertheless, especially systematic aerial photography combined with field checks since 2000 has already met with great success. Results in two of these inland towns, Trea and Ricina, can now supplement and even completely alter all earlier attempts to reconstruct the general topography of the urban tissue, while also in Septempeda new topographic information is available.

2.4.1 Ricina

The Roman city of Ricina, of which only a rather well-preserved theatre building is fully visible above ground level today, lies in the lower Potenza valley, some 15 km inland from the river
mouth. Although there have been a series of investigations to understand the character and extent of this city, almost nothing was known about its general layout and organization until the start of the PVS project (see [136, 609]). Already since the 15th century local searchers have studied the many, still standing remains of this Roman town. These early and often erudite studies have been synthesised by Nereo Alfieri in 1937 [22]. According to most archaeological finds and inscriptions, the city has to be located on the immediate left bank of the river Potenza in an area that is today partly occupied by the small roadside agglomeration of Villa Potenza, partly still used as arable land. This location was not random: it marks the junction of the crossing of the river (which was in antiquity probably still navigable) with an important crossroads of the Salaria Gallica (connecting Urbs Salvia with Aesis) and an offshoot of the Via Flaminia along the Potenza corridor to Potentia (where it joins the coast road Ancona-Aternum (Pescara). According to scarce information from small scale rescue digs in several parts of the town, the site knew an already quite extensive occupation since the later 2nd century B.C. [551].

A segment of a southwest-northeast oriented street lined with shops, found near the modern day Septempedana road that runs parallel with the river, was probably arranged in that time [529]. Most data about the urbane phase of the site are, however, to be placed between the 1st century B.C. and the 4th century. Ricina became a Roman municipium from the mid-1st century onwards, when also the first colons, veterans of the civil wars, were settled here. The city flourished in the time of Augustus-Tiberius, according to a series of funerary monuments and inscriptions probably originating from a cemetery located southwest of the settlement [136, 527, 609], to the construction of an aqueduct [154] and to the building of the largest theatre in Picenum [609]. During the 2nd century much public building work was accomplished and squares and streets were re-metalled [136]. Most discoveries of scarce elements of a thermal complex (near the theatre) and of small parts of houses with mosaic floors [525, 529, 609] are dated in this century.

Already in the first half of the 2nd century, city finances seemed to dwindle, as a curator rei publicae Riciniensium was installed here. Under Septimius Severus (A.D. 205) the city became a colony under the name Helvia Recina Pertinax (in honour of his predecessor) and plans were made to restructure the town, but unfinished sculptures seem to indicate that most of these plans were never executed [136]. After this, written sources remain silent and only some building structures (mosaic pavements) found in the southwestern part of town date certainly from the 4th century. The area of the street with shops was also occupied until the 4th century, but two graves found above the settlement structures indicate a later reduction of the city sometime in late antiquity. Not much more is known of the late antique evolution of Ricina, which suffered possibly from barbaric incursions during the 5th and 6th centuries. The remaining population no doubt sought new living areas in the hills east and west of this valley site. Until late medieval times the ruins were well-preserved [22], but now only the theatre and some minor structural remains are visible above ground.
Figure 2.8. First tentative plan of *Ricina*, based on information from aerial photography, field checks, earlier excavations, and discoveries. Major structures observed during aerial survey: 1) theatre; 2) temple; 3-6) *domus*; 7-8) *horrea/tabernae* (?); 9) monumental building with cistern; 10) street and adjoining structures.

The first aerial photography campaigns by the PVS team on the site of *Ricina* gave no results at all. The spring 2003 campaign, followed by surface sampling of artefact scatters in September of that year, has, however, fundamentally altered this state of affairs [798]. While still in the process of mapping and interpreting some smaller features, we can now put forward many new elements regarding the town’s topographic situation, its overall layout, and its probable extension and wall circuit, besides proposing functions for several buildings of public and
private signature newly discovered from the air (Figure 2.8). The confrontation of our obliquely photographed information with results of ground survey in arable land shows indeed that the city was situated on the left bank of the river. Pale crop marks, some 5 m wide and noticed in several fields, give us a fair idea about the presence of a town wall. The position of the circuit wall is still partly hypothetical in some zones and further checking with geophysical or other methods will be necessary. Nevertheless, we can already state that it seems to delimit a fairly regular and quite flat, almost rectangular area of about 22 ha, positioned between the Potenza valley floor and the trace of a now disappeared subsidiary brook, which once flowed more or less parallel with the river. The wall was traceable in the field by a slight difference in surface level and by the occurrence of concentrations of gravel and fragments of limestone building material. Its position could be mapped near the southern and eastern corners of the town area, while its trace in the northern and western area remains somewhat uncertain. Part of the southern longitudinal side of the town wall is most probably erased by the action of river erosion and sub-recent human interference, most likely gravel exploitation.

The Roman town was more or less centrally crossed, from southwest to northeast, by the valley road between *Trea* and *Potentia*, acting here as *decumanus maximus* of the street network. This main street, which today is almost completely covered by modern housing and roads, was locally excavated in the 1960s, together with a row of late republican shops lining its northern fringes [529]. Clear cut *insulae* cannot yet be distinguished on the aerial images, but some short linear crop marks of possible urban roads in the eastern part of town suggest the existence of a system of several fairly regular streets parallel with or perpendicular to this *decumanus*.

One probable main northwest-southeast axis, now covered by the agglomeration of Villa Potenza and therefore not seen from the air, could have been connected with the former Roman bridge over the Potenza, which earlier archaeological observations have located a few meters upstream from the current bridge. This street, if confirmed by further fieldwork, could have linked the *decumanus maximus* with the Roman bridge [154], passing directly in front of (and parallel with) the *scenae* building of the theatre along its way (as was already suggested by Moscatelli [551]).

Near the intersection of this northsouth axis (the *cardo maximus*?) and the main *decumanus* we can now propose the location of the *forum* of this town. Although today a major part of this ancient city centre is built over by the houses and streets of Villa Potenza, we were able to distinguish several large Roman buildings in the crops of the arable fields north of this central area. One of them is clearly a temple precinct (Figure 2.9). This can be deduced from its typical plan, some above ground *in situ* remains of Roman concrete walls (*opus testaceum*) and from a series of diagnostic surface finds, such as marble fragments with fine architectural decorations (e.g. cornice fragments), all pointing to an early imperial date.

The plan suggests the presence of a rectangular building (dimensions min. 18 m by 33 m), with a northwest-southeast orientation and almost centrally placed in a precinct of some 55 m wide
and at least 90 m long. The imposing building, probably the main temple of *Ricina*, was oriented towards and perpendicular to the presumed *decumanus maximus* and possibly bordered a *forum* square located south of the sanctuary (cf. the situation in the *municipium* of *Suasa* [215]). It is not known to which god it was dedicated, but epigraphic dedications to Augustus, Jupiter, and Mercury are known from this town [154] and an identification as a *capitoline* temple for Jupiter or Augustus seems therefore reasonable.

Figure 2.9. Aerial photograph of crop marks in the central and eastern part of *Ricina*, showing traces of the main temple (2), a *domus* with mosaic floors (4), and a building with preserved cistern (9) (photograph by F. Vermeulen – Ghent University).

Traces of other building structures observed in the fields near the temple concern complexes situated near the probable *forum* and along or near the main EW street of the city. Among the structures of several buildings seen immediately northeast of the temple, we distinguish a large building with a complex array of rooms, possibly surrounding an *atrium* and a *peristyle*. This entity could represent an important *domus*. Except for its typical plan this interpretation is indicated by the presence of several rooms with tessellated or tiled floors, as can be concluded from surface scatters of such building materials in corresponding locations of pale rectangular crop marks in the field.

The presence of much fine pottery of imperial date (*terra sigillata*, ARS, etc.) and of painted stucco sustain this hypothesis. To the southeast of the temple at least two separate buildings are indicated by several still standing wall structures in *opus testaceum*, as well as by corresponding crop marks of walls. Because one of the standing features is identified in the field as a large cistern, measuring 12 m by 5.35 m, one could suppose that a thermal complex
or other public facility for the intensive use or storage of water was located in this sector. Most chronological indications from the intense surface scatters of building materials and pottery in this whole central sector of the city point to activities during the Principate, although some republican and late imperial sherds were also found.

Figure 2.10. Aerial view of crop marks in the western part of Ricina, showing traces of houses along the main street (5-6), a possible commercial complex (7) and the city wall (x). Note the clear traces of post-Roman flooding at lower right (photograph by F. Vermeulen–Ghent University).

Good crop marks of Roman buildings were also discovered in the southwestern zone of the intramural city area. Most of these structures are connected with the decumanus maximus. Two concentrations of buildings with a complex organisation of rooms (Figure 2.9 and Figure 2.10) could again belong to large city houses. Their intricate wall structures, as well as their dense surface scatters of artefacts, indicate a long life of these houses, with possibly repeated restructuring of the domestic architecture. In one of them a hypocaust floor and indications for an aqueduct suggest the presence of baths (Figure 2.11).
Apart from late republican and early imperial finds, especially the late Roman phase (4th to 5th century) with several imports of ARS-wares seems well represented. Interesting is also the presence (in this area near the southern city wall) of linear crop marks that could belong to a commercial building complex (Figure 2.10-7). The regular array in a row of similar rectangular rooms, flanked by a corridor or portico suggests a possible identification of horrea or a set of tabernae. The presence in the field of many dolia and some amphora sherds, as well as the economically suited location of the complex at the city edge and near the (ancient) river bed of the Potenza support this idea.

Finally, we must mention a few linear traces visible in two extramural zones, probably indicating an extension of habitation outside the located city wall. To the immediate southwest of town, outside the probable southwestern city gate (already presumed by Moscatelli [553]) and alongside the road to Trea, such a small settlement area lies between the city wall and the Roman cemetery which was located here by earlier discoveries [609]. Also outside the northeast exit of the main road around which this city was developed, some crop marks (confirmed by a wider scatter of Roman surface material) indicate such an extramural settlement. We cannot exclude that more extramural activity zones existed around this town (e.g. near the bridge and even on the other side of the Potenza), but the modern village inhibits a good evaluation of these phenomena.

A first and fairly general evaluation of the small sample of datable surface finds seems to confirm the general date obtained from earlier excavations in Ricina. As can be expected from
surface material, the late republican finds are somewhat underrepresented, while the late Roman finds suggest an important occupation of this town at least into the 5th century.

2.4.2  Trea

The Roman town of Trea lies in the middle valley of the Potenza, some 30 km from the Adriatic. The hilly area, situated generally between 250 m and 350 m above sea level, is characterized by a narrowing of the valley formed by two axial hill spurs, now respectively occupied by the medieval town centres of Treia and Pollenza. On the south side the river is dominated by a very conspicuous promontory, the Monte Franco. The presence of important protohistoric settlements in this zone, mostly known from their adjoining cemeteries [488, 489] and now also from our surveys (see supra), does not surprise at all when we take into account the strategic value of this particular area. On a dominant plateau, one km northwest of present day Treia, lays the site of the Roman municipium Trea, in an agrarian area around the convento of SS. Crocifisso (Figure 2.12). The only remaining visible ruins are a small section of the city walls connected to the western gate. They are partly incorporated in a now abandoned farm house. According to the Itinerarium Antonini this Roman city was located on the Via Flaminia per Picenum Anconam, a diverticulum from the main Rome-Rimini road, leading via Septempeda, Trea, and Auximum towards Ancona.

Figure 2.12. General view from the north of the former urban area of Trea. Note the location of the SS. Crocifisso and of the preserved configuration of the northern part of the Roman circuit wall in the landscape (photograph by F. Vermeulen – Ghent University).

Since the 16th century many isolated finds and epigraphic monuments concerning Trea were discovered in this general area [69, 284, 553] (for a summary consider [509]). The first major
excavations by Fortunato Benigni in the late 18th century determined the exact location of the town and revealed parts of its city wall, a **basilica** (not exactly located by him somewhere in the western part of town) and a sanctuary with possible thermal building under the cloister of SS. Crocifisso [612]. Since the 1970s the University of Macerata intensified research in this area, with surveys and topographical studies by Moscatelli based on vertical aerial photographs [551, 553] on the one hand, and excavations by Fabrini in the convento compound (in the eastern part of the city) between 1985 and 1988 [284] on the other hand. These studies produced a first hypothesis about urban organisation, especially the location of the city wall, and evidence that the site of the later monastery and church was in the 2nd century organised as an Egyptian sanctuary.

As a result of this research the main traits of Trea's development are now gradually becoming clear, although the precise origin of the site remains unknown. Its location on an elevated plateau could indicate that it was already a pre-Roman centre, possibly later chosen by the Romans to establish one of their controlling **praefecturae**. It became a Roman **municipium** shortly after 49 B.C. [592] and it is conceivable that its concrete circuit wall, built in a **quasi-reticulatum** technique with blocks of whitish local limestone, was erected around that time [551, 612].

According to the **Liber Coloniarum** the territory of Trea was centuriated during the second triumvirate, an intervention which left its trace in the nearby Potenza plain, southwest of the town centre [551]. Many funerary monuments, statuary, and epigraphic evidence that are now displayed in Treia's Museo Civico, indicates that the Roman town flourished particularly between the reigns of Augustus and Antoninus Pius [509, 612]. Like many towns in Italy, later phases are less well documented. The last epigraphic evidence dates from the 4th century [509], but some archaeological finds (ARS pottery, African lamps, and coins) from excavations and surveys prove later (5th and 6th century) habitation in Trea, with a coin of the Byzantine emperor Foca (A.D. 602-610) as the most recent piece [284].

The Lombard 7th century remains obscure, although an ornamental bronze object and a possible grave of that period suggest some continuity at the site. It is conceivable that during the early Middle Ages the remaining habitation was restructured in connection with an old ‘pieve’, a simple early Christian sanctuary for the plebs, here to be located at the site of the SS. Crocifisso. Although this sanctuary is only found in documents from the mid-12th century onwards, many early medieval **spolia** used in the later church of SS. Crocifisso indicate the presence of a much older phase. During the main period of **incastellamento**, the population probably moved towards the easily defendable hill site of Montecchio (later called Treia) around A.D. 1000, and the original city site remained almost deserted.
Figure 2.13. A tentative reconstruction of the urban topography of Treia with position of the town walls (as seen and as presumed), the urban grid (with two distinct phases), the forum (A) and the sanctuary of SS. Crocifisso (B). This sketch, based entirely on the results of oblique aerial photography, is not fully restituted.

Notwithstanding the fact that Treia received full scientific attention these last thirty years, information about the precise location, extent, and urban organization of the Roman city remained very limited and partly hypothetical. This did not change much during our 2001 PVS campaign of systematic fieldwork in the territory of Treia, organized in the eastern part of the territory of the city, nor during the first campaigns of aerial photography. Several of our aerial photography flights in 2002 and 2003 over the town of Treia delivered, however, spectacular results concerning the urban topography. Many traces of buildings and streets were visible in the grain and are now being studied, rectified, and mapped. They completely revolutionize the current knowledge concerning this municipium, altering some of the earlier hypotheses and complementing the current data with a whole series of new identifications.

A first and still preliminary interpretation scheme of these crop marks (Figure 2.13) indicates, among others, the location of: parts of the circuit wall, the near complete pattern of city streets delineating several regular insulae, the forum, and most of its surrounding public buildings and a whole series of other town structures. As such, almost 70% of the town’s infrastructure can be mapped in the near future [the necessary Ground Control Points (GCPs) were measured in the field during the 2008 campaign]. A succinct terrain control, with systematic sampling of some surface material on the ploughed fields of this area (performed in September 2003) produced additional information for the comprehension of some of these urban structures. A first synthesis of the urban infrastructure of Roman Treia can be presented here.
According to our observations the town wall delimitating the main urban area of Trea has an irregular oval shape which agrees with the general topographical configuration of the hilly plateau determining the location of the town. It seems that, on its long northern and short western and eastern sides, the trace of the town wall is still more or less preserved in low earthworks bordered by modern roads, while parts of the long southern city limits, lying on the slopes of the small valley of the Rio Palazzolo, have been remarked as distinct crop marks on some of our oblique photographs. The wall traces seen from the air were locally confirmed in the fields in September 2003 as 6 m wide linear zones with surface concentrations of white limestone rubble and pinkish mortar. As such, the total city area delimited by a circuit wall is probably only about 10 ha (which leaves us with a much smaller city than proposed by Moscatelli [553]). This does, however, not exclude the existence of extramural habitation areas, particularly in eastern and western directions where the less articulated topography allows it.

The aerial images produced good evidence for the decumanus maximus, cutting the city in two halves from east to west. This circa 6 m wide pale crop mark of a probably paved structure represents the main street of town around which most of the urban grid was developed. This road enters the city by the western gate, near the upstanding remains of a tower compound in opus quasi-reticulatum. After some 150 m this road bends markedly towards the east-southeast before continuing in a straight line towards the probable location of an eastern gate. This decumanus maximus could well correspond with the Via Flaminia towards Ancona, but we suspect that its prolongation in a southeastern direction, parallel with the Rio Palazzolo, brings it into the Potenza plain, where it joined the valley road towards Potentia at the site of a small vicus under present-day Villa Potenza [773].

As a result of this angle in the main road, the pattern of the town streets and buildings shows two different predominant orientations (Figure 2.13). The smaller western part of town corresponds with the highest part of the plateau. The crop marks of streets and buildings in this area were more confusing, with indications for several phases of urban development. Still, it is possible to observe the existence of a series of narrower streets parallel with and perpendicular to the main east-west axis. They seem to demarcate several regular insulae, but the picture is too precarious to conclude about their exact dimensions. Several building structures are present in this amalgam of linear structures, most of which could belong to houses, but it is too early for a definite interpretation. Surface survey showed the presence of several zones with mosaic floors and many pottery finds confirm the partial function as habitation quarters. It is interesting to note that most republican surface finds were observed in this sector of town. This could mean that this highest part of the city, which comprised the source of the Rio Palazzolo brook, was also the area of the earliest settlement. Also interesting is the presence here of more late Roman finds, such as 4th to 5th century African Red Slip pottery, perhaps an indication for more continuity of settlement in this part of town.

The larger eastern part of town gives much more precise information. The aerial views made us distinguish a whole series of buildings and public areas, as well as several streets constituting the backbone of the urban space within this sector of town (Figure 2.14). The streets have an
estimated average width of some 4 m and describe a regular grid of *insulae*, having their longitudinal axis oriented parallel with the central *decumanus maximus*. By extrapolating the visible crop marks on our oblique aerial images with information from earlier topographical observations (such as vertical aerial photographs [553] and the presence of the excavated structures under SS. Crocifisso [284]), we can propose the existence of at least eight rectangular *insulae* with regular dimensions of 3 *actus* by 1 *actus* (c. 105 m by 35 m). The shape and dimensions of a series of more irregular additional *insulae* lying in the periphery are conditioned by the presence of the circuit wall.

One full *insula*, almost central in the town and immediately north of the *decumanus maximus*, is clearly the *forum*. It is composed of an open rectangular square, bordered on three sides (north, west, and south) by porticoes. Centrally placed on its eastern side, one finds the configuration of a rectangular and axially placed building of some 20 m by 10 m. Its position, as well as the large surface concentration of marble *crustae* fragments found on this spot, facilitate a determination as temple of the *capitolium* type. The podium building is clearly subdivided in an approach with stairs, a deep *pronaos* and a *cella* with internal infrastructure for the statuary of the venerated deity.

On the south and north sides of the *forum*, the porticoes border rows of narrow rectangular buildings (shops or *tabernae*) with their short sides towards the square. A larger and more complex building in the southeast corner, however, could be a *macellum*. It is probably a
building with several rooms centred on a paved courtyard. It is oriented towards the square with a northern short side of at least some 25 m, while the exact length cannot be determined yet. Finally, the *forum* is bordered to the west by the long side of a large rectangular building, clearly planned as part of the *forum* complex. The building is probably lined on its other sides by a series of shops and its total dimensions approach 35 m by 20 m. A function as *basilica* is not only suggested by its position and typology, but also by the many surface fragments of rich marble building materials (*crustae*, *opus sectile* fragments) found on this location. The whole spatial setting of this centre displays the typical features of a planned *forum* with a dominating sanctuary of the *capitolium* type, a *basilica* on the opposite side, and rows of shops and a food market behind monumental porticoes lined with columns. Although we see an obvious resemblance with several early imperial Italian *fora* (e.g. the mid-1st century A.D. *forum* in *Brescia*), and many fragments of Roman pottery were found during surface control, we are unable to propose a date in this stage of research. It seems, however, likely that this *forum*, and with it the whole gridded central and eastern sector of *Trea*, was only constructed under the reign of Augustus when much epigraphy was produced (or under his immediate successors).

At some distance from this *forum* and especially along the main *decumanus* (which bordered the square on its southern long side), we clearly distinguish crop marks of several buildings oriented in accordance with the grid. According to the surface scatters connected with some of these buildings (e.g. mosaic *tesserae*, *tubuli*, painted stucco, and fine pottery) and their more intricate multi-room plans, some seem to display a great deal of comfort. They could be of the extensive *domus* type, although in a few cases a public function (e.g. *thermae*) and simple *tabernae* should be considered. Fragments of slag found in the northeastern sector of town are possibly indications for local artisanal activity in a part of town with less pronounced crop features. The pottery and other finds on the surface do not (yet?) indicate pre-Roman settlement in *Trea*, but late republican finds are enough frequent to suggest an already important occupation during the 1st century B.C. Most survey artefacts confirm the expansion during the first two centuries of imperial rule, while they also clearly suggest continued human presence into the 5th or early 6th century. Further fieldwork and pottery studies are awaited before we can refine these data.

### 2.5 Conclusions

While the PVS project still has a long road ahead with many archaeological field checks, geomorphological surveys, and pottery studies, it has already contributed substantially to the subject of early urbanization in *Picenum*. The case study of the Potenza valley has for the first time in this part of Adriatic Italy been able to bring forward very tangible elements for the reconstruction of a pattern of grouped settlements, controlling large areas during most of the Piceni Iron Age (9th–early 3rd century B.C.). Although large scale excavations remain necessary, the survey approach to nucleated settlement sites such as Montarice and Monte Franco
procures a first insight in the topographic features of concentrated elite habitation. As probably existed in many other valleys of the hilly and undulating Marche region (but now hidden underneath modern hilltop towns), we observe here a pre-Roman system of relatively compact defended hilltop settlements, with a high potential for the control over movement in the plain and along the coast. Most of these defended villages or vici seem to have antecedents in the late Bronze age and at least some of them were probably almost continuously occupied until the period of their first contacts with Rome during the 4th/3rd centuries B.C.

The transition period of most of the 3rd and early 2nd centuries B.C. is still less tangible in the archaeological record. The Roman military victory over Picenum, finalised in 268 B.C., only had consequences for specific maritime areas chosen for the erection of colonies. After the incursions of the Carthaginian armies, who used Picenum as a base for their attacks on the Roman state, the foundation of the colonies of Potentia and Auxinum during the first half of the 2nd century B.C. meant a new impulse for the Romanization process.

For the area of central Marche, these foundations are a first phase of real urbanization. Although these cities and their respective territories functioned for a while only as Roman islands in a still very conservative land, they were fully developed urban centres. In Potentia (184 B.C.), at the mouth of the Potenza stream, we could study the town’s topography through a combination of survey techniques. This coast colony displays all features of a small, newly founded city, with a regular modulated plan, circuit walls, and a public centre. Its location in the plain, on the coast and along the river course exemplifies the changed political and economic reality of the new age. Its vicinity to a former elite controlled population centre on an adjoining hill (Montarice) probably facilitated the transition of the Piceni into this new reality.

In the interior of central Marche things had probably not changed much by that time. Here we have to wait for a second phase of Roman intervention and urbanization, which started after the social war in 90 B.C. In the inner Potenza corridor, real urbanization, with population centres developing near the valley bottom, dominated the settlement pattern of the 1st century B.C. It is likely that these new centres grew in or very near former occupation areas of the Piceni elite, but so far we could not establish a direct topographic continuity of (grouped) settlement on the sites that eventually developed into typical Roman towns.

Interesting though is that the three interior Potenza cities, Septempeda, Trea, and Ricina, demonstrate a similar system of urban layout. They are all to be characterized as road-towns, because of their positioning on a mainly east-west oriented road, which acts as decumanus maximus for the town grid. In Septempeda and Ricina this is the road through the valley floor ending up at the town of Potentia, in Trea it is a road diverting from this valley road. Furthermore, the three cities are all walled and their intra muros building infrastructure comprises during the early imperial period a multitude of public and commercial complexes (temple, theatre, basilica, market, baths, shops, etc.) arranged along this decumanus and its
Roman Cities along a Valley

adjoining forum. Although they are real cities according to Roman tradition, their relative compactness and their focus on the main transport corridor concords well with the centres of the former vicus/pagus system in the Piceni era. In a similar way, they are characterized as rather small habitation centres controlling the flows of goods and ideas through the valley, while at the same time acting as service towns for the surrounding rural population.

While the success of the Potenza valley cities during the first centuries of the imperial period is well visible, we cannot evaluate yet what happens with this urban network in late antiquity. It seems at first hand that invasions and economic and demographic decline, interrupted by short periods of reorganisation, lead to a slow but irreversible abandonment of many urban centres and of large tracts of the countryside and its related habitation [24]. The scarce dating evidence from surface surveys and small-scale excavations indicates that until the early 6th century some form of concentrated habitation in the towns pertained.

A very serious blow to the Roman occupation system occurred probably during the Greco-Gothic wars (A.D. 535-553), resulting in a drastic contraction of the rural and urban population and a progressive abandonment of the valley settlement. This process becomes irreversible the moment that the Lombards invade central Italy. For the valley towns this is quite disastrous, especially as the Potenza axis will be abandoned by the Lombards as a main communication road between east and west. As can be seen in other areas of Adriatic Italy some ideas could be put forward about what happened in most urban centres between the 6th century decline and the final abandonment of former Roman towns in favour of fortified centres on the hilltops. The survival of some form of permanent human presence could be characterized as small nuclei of poor structures inside the former habitation zone, especially concentrated near places of the Christian cult. These poor structures are often made of wood and clay and are generally built against remaining parts of Roman stone buildings or walls. They leave very little traces and are hard to discover if only survey techniques or small-scale digs are applied.

It is interesting to note that in the area under investigation, we can observe how temporary the Roman urbanization phase has been. We are confronted with a region in which the natural occupation pattern of the Iron Age is one of strategically placed fortified villages, controlling movements of peoples and goods. The Romans implanted their system of organized urban settlement, firmly linked to a network of permanent and well-served roads. When their world finally collapsed, we see a gradual return to a natural state of smaller scale fortified settlements on the hill crests, making full use of the natural assets of the central Marche.
CHAPTER 3

Many a Little Makes a Mickle

Combining Information Layers to Unravel the Organisation of a Roman City

The content of this chapter is published in Journal of Field Archaeology
[ERIH (A); AHCI – IF 2007: none]

Abstract

Our research contributes to the study of Roman urbanization in the Italian peninsula, both in the central Adriatic area and beyond. It focuses on the integrated use of archaeological field methods and non-destructive techniques. The study of the urban layout of the city of Potentia is an example of the use of aerial reconnaissance information combined with regular gridwalking and geophysical survey to intensively investigate abandoned classical town sites. Information can be combined in a new approach to the study of urbanization processes in the heartland of the Roman world, and this integrated methodology can be applied to any large archaeological site with regular patterning such as orthogonal street layouts and planned civic centres.
3.1 Introduction

We present here some results of integrated archaeological survey work carried out since 2000 in central Adriatic Italy (Regione Marche). Our research is part of the long term Potenza Valley Survey Project (PVS) of the Department of Archaeology at Ghent University (Belgium). The main aim of the PVS Project is the study of the urban and rural occupation patterns in the valley of the river Potenza, from Prehistory to the Middle Ages, with a special focus on the period of first urbanization and so-called Romanization of the area (ca. 300 B.C.-A.D. 500). Apart from objectives connected with wider themes such as Italian settlement history and Roman colonialism, this predominantly geoarchaeological research also pursues some methodological objectives. These include the development of interdisciplinary survey methods and the refinement of integrated historical-archaeological GIS work. The Potenza region can thus be regarded as a test-case for the development of methods for landscape research within a well-defined archaeological and chronological framework.

The intensive surveys of the Potenza valley include systematic aerial photography in the ca. 400 km² large valley area between the Apennine hills and the Adriatic coastline, as well as regular archaeological fieldwalking and detailed geoarchaeological field studies in three sample zones spread between the upper valley and the coast. Apart from these rural surveys, which were the subject of most of the first phase of the project [790, 792, 794, 795], intensive intrasite fieldwork is now also being carried out on the main protohistoric centres in the valley (Piceni culture, 9th-early 3rd centuries B.C.), as well as on the four Roman towns located in the Potenza corridor. From west to east, the latter are Septempeda, Trea, Ricina, and Potentia.

As argued in Chapter 2, at least half of the Roman town sites in the central Adriatic region of Marche have been partly or completely abandoned in the early Middle Ages and were never built upon afterwards. It is therefore not surprising that all of the former urban Roman sites in the Potenza valley have reasonable potential for surface survey. Because of the small scale of excavation work previous to our work in the area, much can be learned from an integrated survey approach. Our surveys are designed to improve general understanding of the topography of proto-urban and urban settlement in this part of Italy and to deepen our knowledge of the processes of early (mostly Roman) urbanization in Marche. They integrate the results of non-destructive survey methods, such as active aerial photography and air photo-interpretation, geophysical surveys, geomorphological surveys, detailed topographic measurements, and classic fieldwalking. The results of small-scale, previous or ongoing, excavation work on these cities and artefact studies of survey materials and excavation finds are integrated in this picture, as they progressively become available.
3.2 The Roman City of Potentia

Like almost half of all Roman towns in the Marche region, Potentia (Figure 3.1) became from late antiquity onwards ‘unsuccessful’ in some way and was completely abandoned in medieval times [24]. Almost no modern habitation is present and as agriculture now dominates the former town area, the ancient topography of the city can still be studied with our survey methodology. We concentrated on the all-important intramural part of this Roman city, where the conditions for such an integrated analysis were best.

Figure 3.1. The location of Potentia in the Marche region of Italy.

3.2.1 Historical Context

Roman military dominance over the central Adriatic Regione Marche was established by 268 B.C. The majority of the population living here, mostly belonging to the Italic Piceni peoples, was soon incorporated in the Roman state. The installation of a series of mostly maritime Latin and Roman colonies during the second part of the 3rd and the 2nd centuries B.C. was a major impulse for the Romanization of a region that knew no real urbanized society before the Romans (e.g. [26, 226, 552, 590]). In the Potenza valley, one of the most important natural corridors linking the hilly Adriatic area with the broad Tiber valley and the Tyrrhenian heartland, the Roman impact became very visible from 184 B.C. onwards. With the foundation of the coastal colony Potentia for Roman citizens in that same year (Livy, History of Rome: 39.44 [658]), the lower Potenza valley (and with it the whole area of central Marche) entered its first
phase of real urbanization. Although this city and its territory probably only functioned for a while as a ‘Roman island’ in a still very conservative Italic territory, it grew as a fully developed urban centre, with a major impact on Piceni society as a whole.

According to the written sources, which are particularly relevant for our topographic work, the colony soon received a circuit wall with three arched gates, a street network with sewers, an aqueduct, a temple for Jupiter, and a portico with shops to close the forum in 174 B.C. (Livy: 41.27 [658]). Although the ‘military’ or ‘civil’ character of this first Roman colony, and the number of its initial colonists – only some 300 or approximately 2000 persons – are still matter of scholarly debate [590], the urban pattern of the coastal settlement was fully developed, only ten years after its foundation. Official sources for the later history of the town are, however, minimal. Around the middle of the 1st century before our era, probably in 56 B.C. (Cicero, de Haruspicum Responso: 62 [617]), a major earthquake destroyed part of the town. Epigraphic evidence, however, testifies of the flourishing development of the town from the Augustan age onwards deep into the 2nd century [591, 594]. The lack of inscriptions from the 3rd century A.D. onwards could point to a downfall of the city’s prosperity. By the end of the 4th/early 5th century A.D., Potentia became a bishops’ seat. There is no reference to Potentia in the work of Procopius, who, in his History of the Wars (6, 7) discusses the devastation of the Gothic-Byzantine war (A.D. 535-553) in this region of Picenum. Under Lombard dominance (after A.D. 580) the coastal cities suffered as the result of the collapse of the maritime economy, and although Potentia is mentioned at the beginning of the 7th century A.D. in the Cosmographia of the Ravennatis Anonymi (31, 34; 36, 8 [691]: 68, 84) and in the Geographia of Guido (21, 35; 68, 54 [691]: 117, 129), the city clearly did not survive into the Middle Ages [24, 26].

3.2.2 Archaeological Investigations before 2000

The Italian archaeologist Nereo Alfieri started to work intensively in Marche shortly after the second World War and was the first to locate the urban site of Potentia. He could establish that the town was erected a few hundred meters south of the river Potenza (ancient Flosis), immediately south of the modern coastal town of Porto Recanati [23, 25]. Since this discovery, several archaeological studies were undertaken on the site. Rescue-excavations in the 1960s and 1970s have revealed parts of its northern cemetery and elements of a housing sector in the northeast corner of the town [526, 528]. A study by Moscatelli [552] of good vertical aerial photographic data (e.g. RAF-pictures of the second World War and shortly after) revealed many indications of the town’s regular street grid, while Paci [591] produced a bibliographical synthesis and analysis of all known monuments and inscriptions.

Since the mid-1980s small scale but systematic excavation campaigns under the direction of Edvige Percossi have been organized on a regular basis in the monumental centre of the city. These stratigraphical excavations by the Soprintendenza per I Beni Archeologici delle Marche revealed that the site was possibly already occupied in the early Bronze Age, but nothing so far contradicts the idea that the Roman colony was created here from scratch. The digs in an area
just south of the geographical centre of the Roman city were particularly relevant for the
discovery of a late republican temple (2nd century B.C.), surrounded by a portico and other
buildings of republican and imperial age [613]. According to the date of the portico
surrounding the original temple and of a possible market building (macellum) north and a
luxurious building east of the sanctuary, the city centre was further monumentalized during
the reign of Augustus. A flourishing period under Trajan and the Antonines was, however,
followed by the 3rd century crisis. During the economic revival of the late 3rd and 4th century
the central area was reorganized and most finds (e.g. the coin series) suggest some prosperity
until the beginning of the 5th century, followed by a clear decline. The latest archaeological
finds in this area are dated to the 7th century, but it remains unclear what exactly the character
of this late human presence in Potentia was.

Of some relevance for our topographic work in and around the city area was a 290 m long
archaeological clean up operation of a ditch which crosses the city area in its southern part
[121]. This field exercise during 2000 revealed the locations of several walls and floors of town
buildings, several elements of ancient roads, as well as the circuit wall, the western section of
which had been cut by the ditch.

3.2.3 Topographic Setting

At present the archaeological site of Potentia lies some 100 m from the Adriatic coastline, on an
almost northsouth oriented beach ridge [792]. The terrain is usually flat, lying on average 3 m
above sea level, but it is possible that the Romans slightly adapted the local topography for the
purpose of building the city. The geology of the area consists of a sandy beach ridge with local
gravel beds lying under a thin layer of alluvial clay. This beach ridge is cut to the south by what
can now be determined as the main ancient river bed in Roman times. The present-day
position of the river Potenza more than 1 km north of the Roman palaeochannel is a result of
late and post-medieval human interference with its course [25]. Detailed mapping of the
ancient coastal plain by means of augering and earth resistance measurements (under the
direction of M. De Dapper of Ghent University) should further elucidate the landscape situation
in the immediate surroundings of Potentia during its long history [792, 798]. It should also
clarify the precise role of river floods, alluvial activity, and sea level change in the process of the
gradual abandonment of the town during late antiquity.

The site is currently crossed from its northwestern edge to the southeastern limit by the wide
coast road and adjoining railway (Figure 3.2 and Figure 3.3). A few modern houses and gardens
cover small parts of the former city area, but almost three quarters of the original urban site is
in use as agricultural land today. The central part (west of the main coastal road) lies currently
under grass, as this segment near the excavation area of the temple is now an archaeological
zone protected by law. Most other fields on and near the urban site are arable land.
Figure 3.2. New simplified plan of Roman Potentia and its immediate surroundings, based on aerial photography, systematic field survey, geophysical survey, and earlier excavations.
Chapter 3

Figure 3.3. Oblique aerial view of the urban site of Potentia from the west (spring 2003). Crop marks and the excavated temple zone (centre right) can be seen (photograph by F. Vermeulen – Ghent University).

3.3 Survey Methodology

3.3.1 Aerial Photography

From 2000 onwards the PVS team has developed a programme of intensive flying, with yearly campaigns of systematic aerial photography between the months of April and October [788, 798] (see also Chapter 1). A small aeroplane piloted by members of a local Aeroclub was used to photograph with the help of standard reflex cameras (35 mm) for colour slides and with digital still cameras. The area of the town of Potentia has been continuously monitored from the air, with regular flights during different seasons and in different weather conditions. Especially intense flying during the dry spring of 2003 has produced images locating and visualizing many aspects of the urban topography including two Roman cemeteries, four roads connected to the town, and several extramural settlement areas. Some flights were also useful for monitoring the geomorphology of the changing river course and the precise location of the beach ridges and lagoons. The regular flights over the original intramural area of Potentia resulted in some excellent views of crop and soil marks. The latter are mostly the result of ploughed-up larger stone structures, such as the city streets and remains of the circuit wall, but also occupation layers, zones with locally more organic substance in the upper layers, and humidity traces caused by differential drying of the soil in some parts of the town. The detected crop marks are connected with the colony’s street grid and circuit wall, and with
many building structures. The aerial photos were stored and further manipulated (image enhancement, rectification, mapping) within a GIS environment (see Chapter 1). For this mapping and interpretation work the team also made use of existing aerial photographs, including those made in 1947 by RAF flights over the area and those made by the *Istituto Geografico Militare* in 1972 and 1974. The integrated GIS mapping of all the aerial data is facilitated by the use of AirPhoto: i.e. software developed to rectify archaeological images made with non-calibrated lenses [699] (see also Chapter 1).

### 3.3.2 Geophysical Survey

Although the aerial survey revealed much new and valuable information, locating and mapping the remains of sub-surface archaeological features in the central (protected) part of the colony, near the recent excavations of the sanctuary, remained difficult. To counter this, a large-scale geophysical survey was initiated here in 2004 [793].

![Figure 3.4. Results of the 2004-2005 magnetometer survey, showing subsurface streets and walls (illustration adapted from [793]: Fig. 6).](image-url)
An area in the south of the town was also surveyed to compare the results with data obtained by aerial photography and gridwalking. The magnetometer survey aimed to elucidate details of the layout of the Roman colony, the location of the city walls, and the position of the western gate, particularly in an area where the grass cover prevented good data acquisition by means of aerial photos or fieldwalking. It was also intended to test the capability of earth resistance survey on a targeted area of the site.

For the geophysical survey, grids of 30 m by 30 m were set out to ensure that the survey traverses crossed the line of potential archaeological features at an angle of approximately 30°. The magnetometer survey was undertaken using a Geoscan Research FM36 Fluxgate Gradiometer. Readings were taken at 0.5 m intervals along traverses every 1m. An automatic encoder trigger was used to record the readings, allowing the survey to be conducted more rapidly as the area was relatively free of obstructions. The depth to which the Geoscan Research FM36 Fluxgate Gradiometer can take readings varies with the geology but is usually up to 1m. The magnetometer survey covered in total an area of approximately 7.2 ha (Figure 3.4).

In addition, a small area covered by magnetometry (near the western gate) was chosen to test the response to electrical resistance survey techniques. This was carried out using a Geoscan Research RM15 Multiplexer set up with a double twin probe array in order to speed up the data collection process and therefore cover more ground. Readings were recorded at 1m intervals along 1m traverses.

### 3.3.3 Topographic Survey

During the 2005 field campaign, a series of topographic measurements of the former urban site were initiated. At a moderate level of precision (1 m contours) this strategy allows a secure topographic model to be produced, in accordance with the data available from the topographic map of the area on scale 1:10 000. This allowed a better understanding of the overall topography of the site and enabled us to better rectify the oblique aerial photographs and to position the results of the geophysical surveys in a GIS model.

### 3.3.4 Fieldwalking and Artefact Studies

In 2002 and 2003 the PVS team undertook two short campaigns of intensive and systematic fieldwalking in and around Potentia. That part of the wide urban area, inside as well as outside the proposed wall circuit (which is currently used as arable land) was subdivided in regular units and large samples of datable ceramics, building materials, and other artefacts were collected. Generally 40 m by 40 m blocks were walked, but during the second campaign in 2003 on a freshly ploughed plot in the western part of town, we used a system of 19 m by 19 m blocks. Every square of the 40 m by 40 m grid was walked by 3 or 4 persons within a 20-minute time limit. The smaller 19 m x 19 m grid was walked in somewhat shorter time span and with
only two persons per block to allow reasonable comparison with the other blocks. The collected material is a small but representative sample of the potential surface finds. The results are evenly distributed over the whole (intramural) area of the city as conditions of visibility were almost the same in all prospected fields and all walkers had more or less the same survey experience.

Once collected, the ceramics were sorted by class (fine wares, coarse wares, and amphorae) and production (Black gloss, African Red Slip, etc.), and the building materials by category (mosaic tesserae, bricks, etc.). After a first interpretation of the finds [549, 795], their relative densities across the site were mapped within the GIS. This particular field approach was necessary not only for the control and confirmation of remotely sensed images (such as the aerial photographs and the results of the magnetometer survey), but also for a more detailed appreciation of crucial aspects like the chronology, functional zoning, and spatial development of the urban centre. This intensive field survey of Potentia was carried out in close collaboration with the geomorphological team in order to take into account biases induced by physical processes at the site, such as erosion and riverside sedimentation.

Artefact density maps of the presumed habitation centre revealed differences in occupation density in several sectors of town. Although their chronological and functional significance needs to be further analyzed, they generally confirm the chronological span of the town’s life as deduced from the recent excavations in the monumental centre [613]: from the early 2nd century B.C. to the 6th or early 7th century of our era.

3.3.5 Re-examination and Integration of Excavated Data

The confrontation of the excavated with the survey data is essential not only for identification and dating purposes, but also for the interpretation of the survey results. Thanks to collaboration with the Italian excavators we can use the published material from the old excavations in the northeastern quarter of the town, and integrate still unpublished structural data and find materials from the ongoing excavations in and near the republican sanctuary situated at the heart of the city. When the northeastern quarter of the city was studied by L. Mercando, the site was interpreted as an elaborate farm building (villa rustica). Later, the buildings were assigned to two intramural town quarters (insulae) of Roman Potentia [613], an identification which helped us to propose a new reconstruction of the city plan.

3.4 The Survey Results and their Integration

Although Potentia was first identified in the 1940s and some small-scale excavations have revealed some of its topography and stratigraphy, the PVS field surveys made a detailed plan of the urban pattern possible. Our results identify the location of the town defences (including
the gates), the street grid, the **forum**, several other monumental complexes, many elements of city housing, three extramural funerary areas, and a large segment of the suburban and rural settlement system and roads, connecting the city to its territory. Although all newly discovered structures seem at first sight to be only ‘superficial’ evidence, their contextualization as a result of gridwalking and subsequent artefact studies and their integration with partly re-studied excavation data and with the very valuable historical information about the early years of *Potentia*, allow us to make major progress in the archaeological and historical research of the topography, urbanization, and settlement history of this Adriatic town, building further upon the excellent recent synthesis of Percossi Serenelli [613], who focused on the excavation data.

### 3.4.1 Location and Geomorphological Context

The remote sensing information, in conjunction with data from geomorphological and geophysical fieldwork, demonstrates that the ancient longitudinal beach ridge, lying parallel with the coastline and just north of the Roman mouth of the Potenza stream, was narrow and not completely flat. It was, for instance, noted that in the area of the geophysical survey the results became faint towards the western edge of the town site. There is no doubt that the ancient topography is the cause of this and that the beach ridge must slope relatively sharp down to the west, a fact also confirmed by our recent augering campaign. The depth of mostly clayey material now covering the Roman town masks the archaeological remains to such a degree that the detection of buried remains using magnetometry is very limited. This phenomenon could also account for the fact that the Roman streets and other structures immediately near the town wall were not identified here from the air. The deeply buried remains cease to affect the growth pattern of the grass seen on the surface, which is the fundamental element in distinguishing features in aerial photographs.

The same phenomenon applies to the lower southern and southwestern part of town. Here the arable fields are much eroded due to post-Roman flooding of the river Potenza in this area, an observation again attested by way of geomorphological survey using resistance measurements and augering [792]. Dominating the magnetometer results in this area is a curvilinear positive feature on a northeast-southwest orientation that cuts across the southeast corner of the survey area. To the south of this line it is evident that no trace of the town layout remains and also the clearly visible city streets (see *infra*) stop abruptly at precisely the point where they meet this feature. This phenomenon is also attested on our aerial photographs and on earlier vertical images, and is further indicated by the sharp reduction in surface artefacts towards the south. Together, all these elements will in the near future allow very detailed mapping of the Roman aspect of the full beach ridge. It is already obvious, however, that its quite limited surface, surrounded by an originally very wet landscape, must have put major constraints on the Roman town builders.
3.4.2 City Wall

As identified by Moscatelli [552] on the basis of older aerial photographs (Figure 3.5), the town defences of Potentia can be better distinguished with our new data. Re-examination of these images in GIS overlay with our excellent 2003 oblique aerial photographs and with the information from the magnetometer surveys demonstrates that the situation is quite complex. Although we must remain very careful when considering the diachronic aspects of these defences (as no regular stratigraphic excavations have touched upon them), we would like to put forward the idea of three major phases in the development of the Potentia circumvallation. We also would like to suggest that some of the detailed historical information about the early years of Potentia can be linked to this phasing (Figure 3.6). In a first phase, no doubt coinciding with the official foundation of the colony in 184 B.C. as mentioned by Livy, the chosen town area was probably surrounded by a max. 2 m wide ditch (fossa), possibly flanked by an earthen bank (agger) made up of soil from the ditch. This ditch structure was only clearly seen as a dark and sharply delineated crop mark on vertical aerial images of the northern periphery and on some of the best oblique photographs in the eastern and western parts of the city area. It seems to surround a very regular rectangular area (525 m x 300 m) with an north-northwest–south-southeast orientation, possibly representing the initial settlement of the first colonists. Shortly after in 174 B.C., according to Livy (Livy: 41.27 [658]), the site of the colony was fully urbanized. Thanks to the financial intervention of the censor Fulvius Flaccus, the new citizens
clearly expressed the will to develop from a simple defended military-looking settlement into a real town with structures providing the base for full social and economic development. These urban structures included a temple for Jupiter, a circuit wall with three arched gates, a regular street network with sewers, an aqueduct, and a portico with shops to close the forum square.

Figure 3.6. Urban structures at Potentia based on the integration of survey and excavation data.
The archaeological data support this detailed historical information, for instance in the case of the town defences: the building of a circuit wall. The presence of a circuit wall was first attested during the cleaning in 2000 of a modern ditch, crossing the beach ridge from west to east [121]. Spatially the wall structure was also attested during the geophysical surveys and showed on the vertical and oblique aerial photographs as a quite sharp and pale crop mark, only some 2 m wide. Its observed building technique, an ashlar structure of dry masonry made with large regular blocks of regional sandstone (*opus quadratum*), fits well with the chronology of the early urbanization of the city according to Livy [613]. The observed town wall seems to delimit a very regular, rectangular area of some 525 m by 343 m or almost 18 ha (measured *intra muros*). In the ancient Roman system this represents some 15 *actus* by 10 *actus*. When we include the walls in the total city surface we obtain almost 19 ha. It is certainly significant that this size is quasi identical to the 19.3 ha presumed for the central Adriatic colony of *Pisaurum* (Pesaro), founded in the same year as *Potentia* [728].

Still hypothetical is our identification of a third phase in the development of the town defences. We suggest that the circuit wall initially surrounded an area of the same size as the space confined by the first ditch of the colony (see *supra*), but that this intramural area was enlarged some 50 m to the east in a later stage of the settlement’s history. The initial eastern side of the city wall was then replaced by a street – evidenced by a larger and more pronounced trace in the aerial photographs than most other streets – and the new wall (and ditch?) was built on the edge of what is now a local coast road bordering the current beach area. This hypothesis, to be verified in the field, would certainly explain why the sector in the northeastern corner of the city (excavated by Mercando in the 1970s) was only inhabited from the Augustan era (late 1st century B.C.) onwards. It would also explain why the eastwest streets of the housing blocks in the eastern periphery of the town have a different look in the aerial photographs compared to the other city streets (see *infra*). Such a partial reorganization of the town is also suggested by the stratigraphic discovery of a destruction layer in the temple sector dated around the mid-1st century B.C., when Cicero mentions a major earthquake destroying *Potentia* [613]. It could indicate that the town might have suffered much by this catastrophic event, and that major reorganization and rebuilding was necessary in the decades following the earthquake, as can also be seen in the whole temple area.

3.4.3 Town Gates and their Suburban Links

Whatever the exact phasing of the defences may be, it is likely that the town wall had three major gates, as suggested by Livy’s information mentioned above. As can be expected in a regularly planned colonial town built from scratch, they were more or less centrally placed in their respective town sides, namely north, west, and south. Only for the eastern coastal side of the city the presence of a gate cannot be attested. Thanks to our surveys the northern and western gates are now located in the field, and only await confirmation by excavation.
For the northern gate there is evidence from several aerial photographs displaying a clear breach in the city wall and ditch traces, and the gridwalking of the ploughed field in this area revealed a dense surface concentration (covering an area of 15 m by 10 m) of building materials, consisting of many tiles, some pieces of white limestone, and several large river pebbles. Some aerial photography traces and especially geophysical evidence hinted at the location and quite complex phasing of the second entrance to town, the western gate. Combining both magnetometer and electrical resistance approaches in this area has revealed a 15 m long eastwest oriented zone full with anomalies which would seem to indicate the presence of high concentrations of building material and of a complex gate system. The third gate, in the southern part of the wall, can also be quite well-positioned, despite the fact that it is located in an area which is now much eroded by post-Roman stream movements. This southern town exit links up neatly with a road leaving the town area towards the southwest to reach the main Roman bridge over the river Potenza (see also Chapter 2).

This road, and also the major roads leaving the northern and western gates, have all been discovered and mapped thanks to the study of vertical photographs and especially with the help of our active flying since 2000. Thanks to these aerial surveys and connected ground observations we know that all three roads leaving the city gates were bordered by cemeteries with funeral monuments facing the roadway. A large cemetery north of the city was already known thanks to rescue digs in the 1970s and was recently further confirmed by still unpublished excavations in 2004 by the Soprintendenza per I Beni Archeologici delle Marche, during which the traces of several square limestone funerary monuments along the road, seen as crop marks during one of our flights, were excavated and dated.

A comparable picture emerges along the southwest oriented road leaving the city from its southern gate. Here crop marks of at least five such funerary monuments along the road trace were observable from above and (although no excavations were performed) surface concentrations of marble, limestone, and early imperial pottery (1st century A.D.) sustain this interpretation. Finally, the situation in the extra urban area of the supposed western gate seems quasi identical. Here, the main road leading out of town in a westerly direction along the still standing concrete core of a Roman funerary monument (il Toraccio), had more or less suggested another cemetery area (see Chapter 2). In 2005, the magnetometer survey revealed, immediately beyond the limits of the town wall, several possible traces of mausolea flanking the sides of the road. They can be distinguished as a number of rectilinear and curvilinear positive anomalies, further confirmed by the results of the restricted grid survey of this area in 2003. The latter not only suggested the presence of the circuit wall and its persistence from late Republic to late antiquity, but also demonstrated that the surface finds (mostly building materials, pottery, and glass) were much more abundant intra muros than outside the town area. We can now link the presence of fragments of marble and mosaic tesserae in the extramural area to this funerary area along the western road out of town. Finally, we must stress that some of the extramural surface finds and traces found during remote sensing are certainly to be linked to small areas of habitation and other activities outside the city walls.
is particularly evident in the southern sector where ongoing geoarchaeological research suggests the presence of a Roman harbour near the ancient river mouth.

3.4.4 Street Network

The gates link up well with the regular north-northsouth–south-southeast oriented street system. Several tracts of this orthogonal street pattern, mostly seen on our aerial photographs as some 6 m wide, pale and linear crop (Figure 3.5) and soil marks, were identified on the ground and were complemented by the results of the geophysical surveys in parts of town. Most tracts are characterized in the ploughed surface as clear concentrations of river pebbles mixed with some artefacts, mostly pottery and tile. According to the evidence from the excavations near the temple, most streets are probably some 5 m wide and normally consist of a surface of battered river pebbles. The digs also indicate that at least some of them had one or two sidewalks and brick sewers, the latter lying underneath the axis of the road bed [613]. According to our aerial photography interpretation, the eastwest oriented and somewhat broader (7 m to 8 m) street segments in the (younger?) peripheral eastern part of town seem to be even simpler, with a road bed of battered earth and a stone lining. The two dominant main northsouth and eastwest street axes of the planned Roman town, respectively called *cardo maximus* and *decumanus maximus*, were probably constructed with much more care. It is likely that both streets, which link up nicely with the town gates, were paved. This can not only be deduced from their different aspect in the crop marks seen from above, but also from the stronger linear positive anomalies seen in the magnetometer survey. As the nature of their traces is not continuous, we may presume that the pavement of these main city roads did not completely survive post-Roman spoliation of the building materials of the town.

We can now map the parallel city streets with some confidence and subdivide the urban space into housing blocks (*insulae*) of different size. It is difficult to calculate the exact number of blocks of *Potentia* as we still lack details about some areas. It is also not clear yet whether we should interpret some streets as being fully part of the major grid, or just as local subdivisions of *insulae*. For the moment we hypothesize the presence of a total number of 51 to 57 *insulae*. According to our present interpretation, the city grid has three main zones: northern, southern, and central.

The northern and southern zones seem to have respectively five and four eastwest oriented rows of five *insulae*. The width of these city blocks, not incorporating the streets, is fairly consistent and measures some 35 m or exactly 1 Roman *actus*. The length of the *insulae* is less constant; we propose sizes between 2.5 *actus* (c. 87 m) and 1 *actus*. This system is less regular but comparable to the proposed street pattern with more or less eastwest oriented longitudinal blocks of 70 m by 35 m (i.e. 2 *actus* by 1 *actus*) for the republican colony of *Sena Gallica* (Senigallia), up north along the Adriatic coast [215].
The central zone of Potentia, determined by the presence of a series of large public complexes, seems to have six larger and northsouth oriented insulae with a length of some 3.5 actus or a little more than 120 m and widths varying between 1 actus and 1.75 actus. If all or some of these insulae were subdivided by a central eastwest street then a total number of twelve for this central zone is also possible.

3.4.5 The Civic Centre

Basic rules of Roman regular town planning and some fifteen campaigns of excavations in the central area of the urban space had made clear, long before the start of our systematic surveys, that the forum of Potentia had to be situated in or near that area. The temple found during these digs is most probably to be identified as the late republican sanctuary of Jupiter, mentioned by Livy (41.27). Although good epigraphic support for this still lacks, its prominent location and architectural style, as well as the dating evidence from the sacred pits (favissae) found during the digs, place this important sanctuary in the very beginning of the life of the colony, around the time of the financial support given by the censor Fulvius Flaccus to build a sanctuary and a forum [613].

As the digging campaigns remained restricted to the area of the temple and parts of some surrounding buildings, and because our aerial photographic evidence is still much limited because of the grass vegetation (no deep ploughing!) on this central protected part of the urban site, the exact location of the forum square remained long unknown. It is the great merit of the 2004 geophysical surveys in the large zone adjoining the excavated temple area to have located the forum and to have proven that the central insulae, around the junction of the main cardo and decumanus, represent the civic centre of Potentia.

The magnetometer results are very explicit in situating the forum square, a long rectangular and almost northsouth oriented open area of some 120 m x 30 m, immediately southwest of this junction. It seems to have been bordered on both long sides by rows of shops fronted by a portico, some of which (according to the many tesserae found here during the grid survey) probably contained mosaic floors. The magnetic anomalies testify that several large public buildings surrounded the forum. To the east, the excavations have already identified a temple and part of a food market (macellum) directly north of it. On the southern short side geophysical research suggests a large building, such as a civic basilica, but the magnetic field of this area was too much disturbed by the excavation fence to allow further precision. On the northern side a whole complex of buildings can be distinguished (maybe a temple and/or bath complex), but also the presence of an early Christian church and/or different civic buildings cannot be excluded. The location in this northern area of many older epigraphic discoveries with a public character [593] and of great number of marble fragments found during our fieldwalking could support this suggestion. The presence of good quantities of late Roman fine wares in this general area north and northwest of the forum seems to suggest a long and complex history.
3.4.6 Houses and City Quarters

Finally, our surveys and a re-evaluation of all excavated evidence have also succeeded in producing some information about city housing and the spatial and temporal development of city quarters. Thanks to the magnetometer surveys and intensive site monitoring from the air, fragments of buildings are visible all over town, as a rule neatly lined according to the street grid. For this type of evidence (which is generally less pronounced than city walls, broad streets and large public buildings) it was necessary to constantly confront and integrate the data from these two approaches. Thanks to this effort and despite the ascertainment that the magnetic surveys did not produce very clear evidence of housing structures, a good number of partial plans of houses is now discernable. In the most prominent groupings of traces, some of which were clearly linked to the main northsouth street axis, we can distinguish remains of large buildings, probably single family residences (domus type), with several rooms grouped around what seems to be a central courtyard. At the ploughed surface level they are characterized by a high density of finds and the pieces of brightly decorated wall plaster, sherds of fine wares, marble fragments, mosaic tesserae, and masses of roof tiles and amphorae hint towards an area rich in style. In some other cases where the overall plan of the buildings remains obscure, a series of rectangular patches in the crops, visible from the air, could be interpreted as individual floors. As these particular plots often coincide with concentrations of mosaic tesserae, a certain quality of architecture is indicated. In several areas simple rectangular structures, facing the street with one of the short sides, suggest the presence of shops and modest houses of the so-called tabernae type. Still, as the overall form and plan of many groups of features is not clear, an interpretation as to their function (houses, shops, warehouses, etc.) cannot be attributed. Nevertheless, all these new data add much knowledge to the already available partial ground plans of two insulae near the northeastern corner of the city, were a full battery of house walls, floors, sewers, etc., was excavated in rescue circumstances [526].

Thanks to these rescue excavations and to the evidence from our grid surveys, some chronological information is now also available about housing and about the general evolution of the use of urban space. Although surface finds picked up in the ploughed fields – mostly building materials (tiles, building stones, mosaic tesserae, pieces of wall plaster, etc.), pottery and glass – were abundant throughout the whole intramural area, differences in date and density of the artefacts suggest some interesting phenomena. While a clear drop in the density of surface materials in the most southern and southeastern blocks of our survey grid coincides with the effects in this area of post-Roman flooding and erosion by the original Potenza stream bed [795], a few marked density peaks can be observed elsewhere.

The regular but quite low quantity dispersion over the city surface of black gloss pottery (i.e. good indicators for late republican settlement activity) seems to indicate that most intramural sectors were already (thinly?) inhabited or used in the first two centuries of the life of this town. It is likely that especially the eastern fringe of the urban area, maybe as a result of later expansion of the town in the direction of the coastline, were only fully exploited after the
earthquake of the mid-1st century B.C., as can be deduced from the mentioned dating evidence of the Mercando excavations in the northeastern corner of the city [526].

Figure 3.7. Results of a systematic grid survey in 2002 and 2003 on the city of Potentia.
During the Principate (1st-3rd centuries A.D.) probably the full intramural area was intensely built over and indications from our surveys suggest that some suburban extramural areas (e.g. southwest of the city) were also occupied now (Chapter 2). In different sectors of the city strong concentrations of terra sigillata pottery (i.e. good indicators for richer contexts of early imperial date) can be proof for the uneven spread of rich housing during the booming first two centuries of our era (Figure 3.7).

In late antiquity (4th-7th centuries A.D.), however, a serious contraction of the inhabited space occurred. The artefact maps show a remarkable and consistent concentration of abundant late Roman fine wares in the northern central part of the town. This seems to suggest a clear regrouping of the town population and activity in the area around the forum and immediately north of it. These data agree with the youngest date (4th century) of the excavated housing structures in the northeast corner of town and with the discovery in that peripheral zone of a series of late graves whose proposed date is 5th-6th century [613]. All this demonstrates that parts of the city, and at least the outer northeastern edges near the town walls, were most likely not inhabited anymore in the later life of Potentia, when some formerly inhabited space was used for funerary activities.

### 3.5 Conclusion

It is commonly accepted that the study of towns is central to any understanding of the Roman world, since they were the nodes through which the administration, economy, and dominating culture were negotiated [435]. Among Roman archaeologists involved in intensive urban research it is agreed that there is a firm need for the collection of more data about the internal organization and chronology of the full range of urban sites in different Roman provinces. As much of our knowledge about Roman urbanism in Italy and other parts of the Empire is based on an uncertain sample of well-preserved but atypical towns such as Pompeii and Ostia, there is a need for more systematically collected data, different from the large body of earlier work on Roman towns (which far too often focused upon individual buildings taken out of the context of their urban environment). Furthermore, a major part of Roman urban archaeology is constrained by medieval and modern developments in those towns that remained occupied today.

We believe that carefully defined strategic use of non-destructive field techniques and systematic aerial reconnaissance in investigating such ancient city sites can be useful. We also ask for the careful selection of Roman urban sites to study, focusing on the available range of abandoned sites with well-preserved subsoil structures that allow a fully integrated geoarchaeological approach. To sustain this choice and thus to contribute to the broader historical-archaeological debate on Roman urbanization, we provide the first synthetic results of a project that examines urban settlement patterns in the Potenza valley. Our work complements earlier town-based research in the Regione Marche, which has a long and
distinguished tradition of historical and archaeological work on the development of Roman towns [499]. This research provided important and detailed information about the development of parts of individual towns and town buildings, involving both the excavation and publication of particular structures, and the analysis of material evidence such as pottery, inscriptions, and sculptures. Our approach, by contrast, provides broader information for a series of urban settlements located in a central area of Marche. We are interested in the patterns of urban development, distribution, size, and form and believe that new information can be collected from intensive and integrated surface survey and fieldwork rather than excavation or detailed artefact studies.

If detailed excavation evidence is available it must, however, be fully used. Integrating data can advance our knowledge of the topography and internal organization of the ancient towns and contribute to a wider vision of urban development in the Roman era. Our project complements previous work by providing a broader context for the interpretation and comparison of individual sites.

It has been argued that systematic surface survey and large-scale geophysics contribute to our understanding of Roman urban topography, as was recently demonstrated in classical town sites such as Tanagra in Boeotia [84] and Falerii Novi and Portus in the Tiber valley [434, 435]. Both techniques are easy to use and allow archaeologists to cover substantial areas in a short time span. As we have demonstrated, this is also true for a detailed investigation of available aerial images and especially for an intensive monitoring of the urban site with the help of archaeological aerial photography during different seasons and years.

It is our firm belief that a simultaneous application of these non-destructive techniques with feedback from applied geomorphological research creates a method ideally suited to characterize the extent, organization, and chronology of urban sites typical of central Italy and other parts of the western Mediterranean. By contrast, the traditional strategy of attempting to excavate urban landscapes is expensive and not always of great scientific value. The Potenza Valley Survey project builds upon these experiences and attempts to refine the non-destructive exploration of Roman urban landscapes.

Despite the success of the magnetometer and earth resistance surveys, the active aerial photography, and the analyses of these data in a GIS model, it must be stressed that they revealed little with regards to the chronology or specific function of buildings detected (see also Chapter 2 or [434]). Viable hypotheses about functions and even general chronologies of Roman urban patterns remained problematic. Therefore, we conducted systematic fieldwalking in grids over large sectors of town, contextualizing the remote sensing imagery. Together with further integration of the valuable stratigraphic data from different excavations in the town area, this work has proven to be of crucial importance for our understanding of the city topography and part of its evolution. We are convinced that this integrated methodology can be widely applied beyond the geographical and chronological limits of Roman
archaeology and should be particularly useful on all types of large open archaeological sites with some form of regular or understandable patterning.
Taking Still Cameras Aloft
The Sky Must Not Always Be the Limit

« Sine pennis volare haud facile est. »
(It is not easy to fly without wings.)
Titus Maccius Plautus (254-184 B.C.), playwright
CHAPTER 4

Archaeological Bird's-Eye Views

Various Ground-Based Means to Execute Low Altitude Aerial Photography

The content of this chapter is submitted to Archaeological Prospection [ERIH (B); SCIE, AHCI – IF 2007: 0.660]

Abstract

Since the beginning of aerial photography, researchers used all kinds of devices (from pigeons, kites, poles, balloons to rockets) to take cameras aloft and remotely gather aerial data, needed for a combination of research goals. To date, many of these unmanned devices are still used, mainly to gather archaeologically relevant information from relatively low altitudes, enabling so-called Low Altitude Aerial Photography (LAAP). Besides providing a concise overview of the unmanned LAAP platforms commonly used in archaeological research, this chapter weighs the pros and cons of every device and provides an extensive reference list.
4.1 Introduction

Aerial photography was given birth in 1858 when Gaspard-Félix Tournachon, called Nadar by himself, took the first aerial image of Petit Bicêtre (a village near Paris) from a tethered hot-air balloon some 80 m above the ground [189, 243, 574]. Afterwards, balloon photography became gradually more established, largely helped by the development of the dry-plate process pioneered in 1871 by Richard Leach Maddox [243, 573]. It was not, however, until June 1899 that the first (European) archaeological balloon photograph (of the forum in Rome) was taken by Giacomo Boni [153, 157, 625].

Although the English Meteorologist E.D. Archibald [189, 377, 772] claimed to have taken aerial imagery from a kite around 1882, the lack of any surviving imagery to prove his point allowed Arthur Batut to become the pioneer in what was later called Kite Aerial Photography or KAP [573, 772]. In 1888, Batut took the first aerial photographs by means of a large kite [59], followed two years later by the French Emile Wenz. Besides different ingenious man-lifting kite systems – the most spectacular being those of the American born Samuel Franklin Cody [377, 573, 653, 670] – more unmanned solutions to acquire aerial imagery were being proposed at the end of the nineteenth and beginning of the twentieth century.

A notable year was 1903: in addition to the first flight of a manned, heavier-than-air motor-driven machine built by Orville and Wilbur Wright, the German Alfred Maul applied and patented his invention to launch a camera using a powder rocket (Figure 4.1A). This seemed to be a far more reliable and successful method compared to the patented proposal by Ludwig Rahrmann of 1891 and the first photographs taken in 1897 with a small rocket designed by the Swedish Alfred Bernhard Nobel [573, 772]. Also in 1903, the German engineer Julius Neubronner experimented with breast-mounted cameras for carrier pigeons (Figure 4.1B), using timers to take pictures along the flight track of the bird [76, 573, 772].

Those techniques were developed further – one noteworthy example is George R. Lawrence’s imagery of San Francisco’s devastated buildings, acquired six weeks after the great fire following the earthquake of 1906 by a train of kites to lift a large-format panoramic camera [47, 573] – before L.P. Bonvillian took the first true photograph (which is in fact one frame of a motion picture) from an aeroplane in 1908 as a passenger on Wilbur Wright’s aeroplane when flying near Le Mans, France [243, 573].
During World War I, aerial photography by means of an aeroplane became a standard practice [230], and by the end of the war its usefulness in civilian applications became (commercially) exploited. To date it is this form of aerial photography (and later on also spaceborne imaging) that has remained important and has seen technical improvements throughout all kinds of research fields. In (European) archaeology, aerial reconnaissance has – since the 1930s – largely been characterised by individuals acquiring their own data from the cabin of a small, relatively low-flying, conventional fixed-wing aeroplane, utilizing 135 mm format (or slightly larger or smaller) hand-held still cameras to acquire imagery that is mostly oblique in nature. The value of such conventional Small Format Aerial Photography (SFAP) is obvious and has, over the years, been subject to certain improvements (e.g. the use of digital photo cameras, multispectral approaches, camera rigs, several methods of geocoding, the combination with satellite imagery, etc.). However, in some countries, flying aircrafts for aerial image acquisition is simply forbidden by the military [35]. Even if it is allowed, this conventional way of image acquisition might be inconvenient and/or logistically difficult due to weather conditions, topographic features, flying restrictions, required qualifications or other factors making it hard to (safely) fly over the region of interest at an appropriate height.

On some occasions, such as large-scale photography (e.g. 1:500) or narrow-band imaging (e.g. from 520 nm to 570 nm), the forward movement of the aeroplane is far too fast to compensate for the situation-specific shutter speed needed, hence creating excessive blur. Renting an aeroplane might also be too expensive if only a few images from a rather limited area are needed, certainly when frames have to be generated on a daily basis (i.e. large **temporal resolution**). Even if all aforementioned issues are not applicable, the **Ground Sampling Distance (GSD)** or the distance on the ground between two adjacent samples that form an image point or pixel of conventional SFAP can be too broad (rather expressed in decimetres than centimetres) to resolve the finest details needed in research such as excavation.

---

Figure 4.1. (A) German Army photo rocket squad with Maul's equipment and pigeons with cameras (B) [573].
photography, recording standing ancient remains or capturing very narrow crop and soil mark features.

4.2 Low Altitude Aerial Photography

To deal with these issues and still be able to obtain qualitative archaeological imagery, lower and/or slower flying manned platforms like helicopters (e.g. [295, 438]), balloons (e.g. [137, 822, 825, 827]), powered parachutes (e.g. [362, 458]), gliders, paramotors (e.g. [287]), and ULMs (Ultra Légers Motorisés)/ultralight aircrafts (e.g. [18, 106, 789, 805]) or unmanned devices such as balloons, kites, model aircrafts, blimps, and poles are used to acquire imagery from the air. Although there are exceptions in both approaches, the latter systems generally allow the Radio Controlled (R/C) or action delayed (digital) still or video camera to be more or less stationary over specific spots of interest at particular altitudes, difficult or impossible to achieve through the former, manned platforms. Additionally, most unmanned aerial platforms also allow the operation height to be very low (e.g. 30 m), enabling Low Altitude Large-Scale Reconnaissance (LALSR) or Low Altitude Aerial Photography (LAAP) (also called low level, large-scale or close range aerial photography) to acquire image data that even resolves the finest details [688]. However, these systems vary a great deal in capabilities, costs, operation conditions, flexibility, and maximum working heights.

The following overview will therefore present the ground-based means commonly used in archaeology to lift photographic or videographic devices and acquire close range, large-scale aerial imagery. Besides a short historical note and general description, the particular drawbacks and advantages of these unmanned camera platforms will be mentioned and weighed against each other.

4.3 Mast, Pole or Boom

Already employed at the beginning of the twentieth century (e.g. [455]) and ranging from simple telescopic masts to ingenious, adjustable booms (or related forms using a bipod, tripod or quadrupod, scaffold, crane, turret, ladder or aerial bucket trucks – the latter five manned however), these constructions (Figure 4.2) are used to document excavations and generate large-scale plans (e.g. [14, 44, 108, 165, 192, 241, 286, 293, 349, 357, 365, 427, 444, 522, 530, 581, 634, 657, 693, 716, 737, 743, 746, 818, 817, 821, 825, 827, 828, 835, 841]). Although the potential temporal and spatial resolution is extremely high and most of these camera platforms (but some of the manned versions such as scaffolds and cranes) are cost-efficient, portable, and extremely stable, these inexpensive systems have the major drawback of an operation height limited to only 15 m - 20 m [241, 620]. Moreover, most systems generally cast a shadow on the area to be photographed [737].
4.4 UAV

Initially developed for purely military applications, Unmanned Aerial Vehicles or UAVs [also known as Unmanned Vehicle Systems (UVSs), Remotely Operated Aircrafts (ROAs), Remotely Piloted Vehicles (RPVs) or drones] are powered aerial vehicles that do not take a human operator aloft but fly and manoeuvre in the air autonomously or completely remotely controlled [100, 268]. Although in use for intelligence gathering since the 1950s, the first UAV used for civil mapping purposes was a model aeroplane, reported on by Przybilla and Wester-Ebbinghaus [639]. Together with model helicopters (Figure 4.3A), both UAV types have been equipped with cameras to obtain aerial imagery, also for archaeological purposes (e.g. [72, 133, 268, 269, 270, 375, 387, 388, 422, 540, 603, 694, 695, 718, 719, 757, 850]).

Today, many manufacturers offer R/C aeroplane and/or helicopter models that can lift several kilograms of payload, but the major drawback still is their cost and need for an experienced operator, because flying these systems – certainly helicopters – is far from easy [269, 375, 428, 582, 641, 718, 719, 755, 757, 815]. Moreover, the final cost is generally much higher than the price of initial purchase, as the ever present risk of crashing can largely destroy the camera and the aircraft itself. Thirdly, one needs to make sure that the serious vibrations, induced by both the motor and rotary wings of a helicopter [71, 171, 805, 815] and the airframe vibrations in a model aeroplane [375], are damped. Even if current suspension systems can largely eliminate these vibrations, they still remain a serious issue [268]. Consequently, high shutter speeds are still needed to combat the remaining vibrations in both aircrafts and the blurring effect of relative ground speed that occurs at lower altitudes in case of an aeroplane [374, 375, 452, 540, 755, 805, 807, 815], limiting the application of both systems for photographic situations where low amounts of reflected radiation need to be recorded (and to very low altitude photography in case of the model aeroplane). Flying the latter device can also be severely restricted when wind speed largely exceeds 30 km/h [540], and unsuited ground might be risky to land on.
Finally, problems with petrol, gas, and engines [268, 269, 620] as well as civilian licensing [688, 719] can still occur.

On the other hand, these devices allow very accurate navigation, certainly if the UAVs are equipped with a GPS/INS (Global Positioning System/Inertial Navigation System)-system [268, 269, 270, 582]. In case of the helicopter, a major advantage is the ability to operate close to the ground, while both helicopter and aeroplane are also readily available [375]. Notwithstanding, as long as the aforementioned disadvantages are not completely solved and the birotors (e.g. Workfly’s Eyesfly) and easy-to-fly quadrotors (e.g. Microdrones – Figure 4.3B, Draganflyers, Intelicopters) remain excessively expensive and their payload too little (≤ 1 kg), non-UAV platforms generally are more practical to use. However, the current technological innovations will most likely allow these UAVs to become more powerful and their price less high, which means that such devices – and more in particular quadrotors – could become the archaeological LAAP solutions of the near future.

A rather new type of UAV, the so-called R/C parachute or R/C paraglider (Figure 4.3C) – a combination of a parachute with a motor and radio gear suspended below [411, 456, 598, 638] – will most likely not alter this situation. Although the many variants of these easy-to-handle and slow-flying devices have been used in low level archaeological photography (e.g. [78, 411]) because their payload capacity easily accommodates a (still) camera, these UAVs are also wind sensitive (maximum wind speed is about 20 km/h [40, 638, 644]), while their working principle does not allow for hovering and keeping them at a constant altitude is difficult [78, 456, 638]. Due to the combination of all aforementioned UAV-related issues, kites and balloons currently are still the most widespread used platforms to acquire low-cost, close range archaeological aerial imagery.

4.5 Kite

After its initial use by Sir Henry Wellcome to stereoscopically record the Sudanese site of Jebel Moya in 1911 [14, 164, 203, 230, 852] (Figure 4.4), archaeological KAP (Kite Aerial Photography) was only practiced at very few occasions (e.g. [57]) before it became more common in the 1970s and 1980s [10, 165]. Nowadays, KAP is practiced by several individuals
and teams to obtain archaeological LAAP (Table 4.1, Figure 4.4B-D), showing that “in the hands of scientists, a toy does serious data gathering” ([614]: 186). KAP’s popularity is due to the fact that the system is highly portable and inexpensive [7, 8, 57, 162, 283, 385, 824], as the initial purchase of all necessary material is minimal and only manpower and wind are needed to get it working (which allows the operating costs to be virtually zero [35]).

Moreover, the equipment is less fragile, smaller and lighter compared to UAVs [588, 587], while still accommodating a few kilograms of payload [370], with the exact amount largely depending on the wind speed as well as the size and design of the kite [161, 183, 570, 756]. These characteristics also made KAP very popular in a wide variety of non-archaeologically related disciplines: hydrology (e.g. [696]), forestry (e.g. [6, 10, 83]), bird study (e.g. [138]), wetland mapping and analysis (e.g. [4, 8]), geomorphological and soil studies (e.g. [9, 10, 98, 512, 513, 514, 517]), humanitarian purposes (e.g. [721]) and others (e.g. [570, 673]). Although KAP is possible at very moderate heights [57], and therefore an excellent system in microstructural mapping [4], the system does not work as a tool for large area archaeological reconnaissance [35].

Besides, KAP comes with one big disadvantage: one does need a steady wind as irregular winds are not suited for KAP [19, 161, 283, 370, 447, 587, 589]. According to the speed of the wind and the payload to get aloft, the type of kite and its size have to be adapted – or kites must be flown in tandem (e.g. [183]). In general, large, rigid kites are applied to handle moderate winds (about 10 km/h – 25 km/h), while soft, smaller kites are best used in stronger winds (25 km/h – 40 km/h) to lift the equipment [1, 3, 7, 160, 161]. Some have even used extremely small kites (e.g. 1 m³) to fly in winds surpassing 40 km/h [160, 161]. In cases where the wind is insufficient, other solutions are needed as lifting platform [756, 772]. In this respect, kites can be seen as complementary to balloons.

Figure 4.4. (A) Kite used at the site of Jebel Moya ([14]: Plate XVI.1); (B) multi-flare kite with attached camera ([756]: Fig. 27d); (C) a delta kite taking off ([570]; (D) a flying rokkaku kite ([512]: Fig. 2).
4.6 Balloon or Blimp

Besides its use in atmospheric and meteorological observations (e.g. [541, 770]), balloons and blimps (Figure 4.5) are largely applied as aerial platforms to monitor crops and vegetation (e.g. [8, 122, 208, 220, 359, 360, 392, 409, 410, 486, 545, 546, 556, 568, 629, 630, 631, 682]), rock, soils and geomorphology (e.g. [46, 98, 308, 513, 515, 516, 663]), hydraulics and hydrographical networks (e.g. [66, 98, 170, 679]), and to acquire imagery for several other applications (e.g. [185, 228, 519]).

Besides, Balloon/Blimp Aerial Photography (only rarely called BAP) has since long been established in archaeology. The first attempt to record archaeology from an unmanned balloon was Guy’s work at Megiddo (Palestine) around 1930 [357] (see Figure 4.5(A)), after which archaeological BAP gained momentum in the 1960s and 1970s (e.g. [561, 819, 820, 821, 823, 824, 825, 827, 826, 828] – see Table 4.1).

![Figure 4.5. (A) Pioneering archaeological BAP at Megiddo ([357]: Plate III); (B) hydrogen photo balloon ([825]: Fig. 1); (C) helium photo balloon ([19]: Fig. 2); (D) photo blimp ([745]).](image)

Although it was said in the early days that windless conditions were a condition sine qua non [357], conventional blimps and balloons filled with hot air, helium (He) or hydrogen (H) can remain stable platforms in very light wind conditions up to about 10 km/h - 15 km/h [1, 2, 5, 28, 514, 561, 629, 696], after which they become very difficult to position and hold steady [35, 510, 515, 536, 663, 745], with spherical balloons the most unstable of these devices [563, 565].

Besides sensitivity to wind, these lighter-than-air devices can also suffer the problems of moving and safely storing the inflated device in addition to the cost and availability of helium or hydrogen [409]. After all, it might not always be possible to purchase helium or hydrogen locally and get it on site; in some situations, this can be a reason not to opt for these devices (e.g. [28, 40, 83, 98, 587, 589, 696, 827]), although some authors also point to the fact that helium allows for very silent operation and lets the balloon/blimp to be aloft for extended periods of time [510].
Table 4.1. An overview of different BAP and/or KAP systems used since the 1960s to acquire low altitude archaeological aerial imagery. The abbreviations in the “Use” column denote Archaeological Prospection (AP), Site Documentation (SD), and Site Documentation and Mapping (SDM).

<table>
<thead>
<tr>
<th>Lifting device details</th>
<th>Range (m)</th>
<th>Cameras mounted</th>
<th>Camera type</th>
<th>Live View</th>
<th>Use</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Helium balloon (8 m³)</td>
<td>50 m</td>
<td>1</td>
<td>Nikon F70</td>
<td>Yes</td>
<td>SDM</td>
<td>[381]</td>
</tr>
<tr>
<td>Hot air balloon (37 m³)</td>
<td>250 m</td>
<td>1</td>
<td>Hasselblad 500 EL/M</td>
<td>No</td>
<td>SDM</td>
<td>[493, 802]</td>
</tr>
<tr>
<td>Hydrogen balloon (15.3 m³)</td>
<td>100 m</td>
<td>1</td>
<td>Canon EOS 5D</td>
<td>Yes</td>
<td>SDM</td>
<td>[536]</td>
</tr>
<tr>
<td>Helium balloon</td>
<td>300 m</td>
<td>1</td>
<td>Yes</td>
<td>SDM</td>
<td>[432]</td>
<td></td>
</tr>
<tr>
<td>Hot air balloon</td>
<td>1</td>
<td>No SDM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hydrogen balloon</td>
<td>100 m</td>
<td>1</td>
<td>Nikon F2</td>
<td>No</td>
<td>AP</td>
<td>[664, 665]</td>
</tr>
<tr>
<td>Helium balloon (300 m³)</td>
<td>300 m</td>
<td>2</td>
<td>Hasselblad EL-500</td>
<td>No</td>
<td>SDM</td>
<td>[432]</td>
</tr>
<tr>
<td>Helium balloon (19.8 m³)</td>
<td>3</td>
<td>Hasselblad EL 500</td>
<td>No</td>
<td>SDM</td>
<td>[823, 824, 826, 827]</td>
<td></td>
</tr>
<tr>
<td>Helium balloon (19.8 m³)</td>
<td>3</td>
<td>Rollei 35</td>
<td>No</td>
<td>SDM</td>
<td>[826]</td>
<td></td>
</tr>
<tr>
<td>Helium balloon</td>
<td>2 or 3</td>
<td>Hasselblad El-500 and Canon AE-1</td>
<td>No</td>
<td>SDM</td>
<td>[828]</td>
<td></td>
</tr>
<tr>
<td>Helium balloon (2.1 m³) and multiflare kite</td>
<td>500 m</td>
<td>1</td>
<td>SLR and point-and-shoot camera</td>
<td>No</td>
<td>SDM</td>
<td>[61]</td>
</tr>
<tr>
<td>Hydrogen balloon (17 m³) and Jalbert Airfoil (3.9 m³)</td>
<td>600 m</td>
<td>1</td>
<td>Linhof Technica, Graflex XL and Nikon SLR</td>
<td>No</td>
<td>SDM</td>
<td>[819, 820, 821, 824, 825, 827]</td>
</tr>
<tr>
<td>Helium balloon and aerofoil (2.8 m³)</td>
<td>50 m</td>
<td>1</td>
<td>Olympus Stylus 800</td>
<td>No</td>
<td>SD</td>
<td>[838]</td>
</tr>
<tr>
<td>Helium blimp (7 m³)</td>
<td>150 m</td>
<td>No SDM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hydrogen or helium blimp (35 m³)</td>
<td>800 m</td>
<td>2</td>
<td>Hasselblad ELM/ 500 and Canon AE-1 or Nikon 2020</td>
<td>No</td>
<td>SDM</td>
<td>[194, 560, 561, 562, 563, 564, 565]</td>
</tr>
<tr>
<td>Helium zeppelin (11 m³)</td>
<td>850 m</td>
<td>1</td>
<td>Nikon D70</td>
<td>Yes</td>
<td>SDM</td>
<td>[674]</td>
</tr>
<tr>
<td>Helium blimp (20 m³)</td>
<td>1</td>
<td>No SDM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stratoscoop aerofoil Mark 3</td>
<td>1</td>
<td>Home-made 5 in x 4 in camera and Olympus OM-1</td>
<td>No</td>
<td>SDM</td>
<td>[35]</td>
<td></td>
</tr>
<tr>
<td>jalbert aerofoil (4.18 m³)</td>
<td>1</td>
<td>Home-made 5 in x 4 in camera and Olympus OM-1</td>
<td>No</td>
<td>SDM</td>
<td>[35]</td>
<td></td>
</tr>
<tr>
<td>Different sizes of airfoil kites</td>
<td>1</td>
<td>Home-made 5 in x 4 in camera and Olympus OM-1</td>
<td>No</td>
<td>SDM</td>
<td>[35]</td>
<td></td>
</tr>
<tr>
<td>Rokkaku</td>
<td>1</td>
<td>No SDM</td>
<td></td>
<td></td>
<td></td>
<td>[104]</td>
</tr>
<tr>
<td>Sonjo Rokkaku</td>
<td>1</td>
<td>No SDM</td>
<td></td>
<td></td>
<td></td>
<td>[112]</td>
</tr>
<tr>
<td>Rokkaku (5 m²) and carré japonais (2 m²)</td>
<td>100 m</td>
<td>1</td>
<td>Canon Prima 5 / Olympus µ1</td>
<td>Yes</td>
<td>SDM</td>
<td>[159, 161, 162, 320, 385]</td>
</tr>
<tr>
<td>Dunford Flying Machine 2000, Delta 1800 and 2500</td>
<td>&gt; 500 m</td>
<td>1</td>
<td>Canon AE1 and home-made 5 in x 4 in camera</td>
<td>No</td>
<td>SD</td>
<td>[183]</td>
</tr>
<tr>
<td>Delta Conyne and FlowForm 16</td>
<td>150 m</td>
<td>2</td>
<td>Minolta 7000 Maxxum and Yashica T4</td>
<td>No</td>
<td>SD</td>
<td>[244]</td>
</tr>
<tr>
<td>Parafoil (3m²)</td>
<td>120 m</td>
<td>1</td>
<td>Olympus OM 1</td>
<td>No</td>
<td>SD</td>
<td>[399]</td>
</tr>
<tr>
<td>Sparless aerofoil</td>
<td>1</td>
<td>No SDM</td>
<td></td>
<td></td>
<td></td>
<td>[447]</td>
</tr>
<tr>
<td>Diamond-shaped kite</td>
<td>180 m</td>
<td>1</td>
<td>No SDM</td>
<td></td>
<td></td>
<td>[521]</td>
</tr>
</tbody>
</table>
Notwithstanding these disadvantages, balloon and blimp aerial photography can be extremely flexible to set-up and operate [98, 122], is characterised by its ease of use and maintenance [515] while the platform itself is virtually vibration free [515, 663]. As with KAP, photographs from very low altitudes are possible [493, 515, 563], although with BAP there is a greater risk of photographing the tether [35].

This way, it is clear that a combination of a balloon and a variety of kites is often essential to maximise the conditions for low level aerial archaeology, as using only one of these two systems makes it more likely that attempts to acquire large-scale aerial imagery will be compromised. However, rather than using two or more separate lifting platforms (e.g. [2, 19, 61, 820, 825, 827, 838]), a Helikite can be used.

### 4.7 Helikite

The Helikite is a hybrid between a balloon and a kite patented in 1993 and currently manufactured by Allsopp Helikites Ltd. By combining a helium balloon with kite wings (Figure 4.6), this lighter-than-air device joins the best properties of both platforms without incurring too much of their disadvantages. The helium filled balloon allows the Helikite to take off in windless weather conditions, whereas the kite components become important in case there is wind: first of all, they lift the construction up in the air to altitudes higher than the pure helium lift. Moreover, the Helikite’s lift becomes stronger with increasing wind speed (with an upper lift limit depending on its size). Secondly, the wings counteract any unstable behaviour that is characteristic of traditional balloons and blimps flown in windy conditions. Moreover, the construction supports more payload for its size when compared to ordinary aerostats [31].

Due to the fact that it can be used in adverse weather (e.g. rain, fog, freezing conditions) it might also be more flexible in operation than most UAVs. Based on these properties, Helikites are employed as versatile devices to lift GPS pseudolites [671] or radio and weather equipment, they allow long-term aerial surveillance for the military [31] or the monitoring of river systems [785]. Although limited to a moderate extent, Helikites have been used in aerial archaeology as well. In Italy, they are employed by the author to map excavations and acquire beyond visible information ([776, 781] or Chapter 5, Chapter 9, Chapter 10, and Chapter 12). Besides, the Egypt Exploration Society and, more recently, the Amarna Trust have used a Helikite for general survey and for detailed imagery of excavations at Tell el Amarna in Middle Egypt (pers. comm. Gwil Owen). Further more, archaeologists of the Discovery Programme (Ireland) use a small Helikite to provide the necessary aerial images for photogrammetric recording of excavations and medieval structures [705].
However, Helikites also do have some important drawbacks. It remains challenging to accurately establish a precise camera location or take aerial imagery along an outlined track – a problem encountered with all forms of photography using kites and blimps [432, 562, 718, 756].

As the tether might also conflict with trees, houses, scrub, and power lines, not each and every single place is suited to perform HAP. Additionally, the wind direction and topographical setting can prohibit its position above the spot of interest. Consequently, the positioning capabilities of UAVs still remain superior for R/C LAAP.

4.8 Applications

Although Schlitz [688] and Walker and De Vore [807] identified some crucial points for a LAAP system to be effective (e.g. low velocity, small take-off and landing space, portability, low cost, minimal operational staff, low vibration, reliable power supply, fast to set up and employ, low risk and low impact), the choice for one system over another will largely depend on the best compromise between all device-specific limitations on the one hand, and the topographical setting, the weather conditions, the expertise, the (running) costs, and the particular application on the other. To maximally exploit the whole range of possibilities offered by LAAP devices, one could even use several platforms side-by-side.

Despite the fact that previous overview clearly indicates that these devices can be applied for a variety of archaeological tasks (see [688] for a more detailed overview of archaeological
applications), documentation of archaeological excavations and their direct environment is
generally LAAP’s main application (e.g. record every individual phase of an excavation or
generate small-scale overviews). Besides their use in folders, presentations, books, and
websites, these images largely aid the final interpretation of the excavated features by putting
things into a new perspective and revealing minute aspects about individual features that can
be disregarded in traditional plans and photographs. This is the reason why Żurawski once
stated that all ongoing excavations should have a “handy, reliable and inexpensive vehicle
capable of shooting aerial pictures at a particular moment of the fieldwork” ([852]: 244).

Moreover, low level aerial photography has the potential to break the boundaries of
conventional documentation purposes, as the acquired aerial imagery can be used to generate
Digital Elevation Models (DEM), create orthophotos with accurate metric information, and
obtain consistent maps that are distilled faster than other, low-cost mapping techniques.
Additionally, the colour and textural information provided by the orthorectified photographs
remains essential to complement laser scanning [705], allowing the generation of
photorealistic 3D site representations in past landscape settings. The fact that this approach is
not restricted to conventional sites has been illustrated by several scholars who used such
unmanned platforms to map underwater remains (e.g. [419, 562, 563, 820, 821, 824, 825, 827,
828]).

Although not suited for extensive reconnaissance, most unmanned systems can also be used
to monitor sites and their direct environment because they permit the acquisition of imagery
at very specific time intervals and enable a fast response to events. Besides, new site-based
imaging techniques can be exploited, as it is often only a matter of lifting the appropriate
device. This way, several archaeologists already tried close-range Near-InfraRed photography
(e.g. [4, 6, 7, 8, 44, 98, 122, 360, 758, 776, 781, 823, 824, 826, 827, 828] or Chapters 5 and 9), but
also less straightforward Near-UltraViolet (e.g. [777] or Chapter 12), and thermal imaging (e.g.
[426]) has been proved possible.

4.9 Conclusion

Several unmanned devices allow for archaeological ground-based low level photography, but
this does not mean that they can be applied equally successful in every possible situation.
When deciding upon which platform to choose, archaeologists should take all device-specific
drawbacks and advantages in consideration. Some situations might even ask for several
platforms. But irrespective of the platform chosen, all of them will be more cost effective, offer
a larger flexibility, and yield superior results when compared to site-based aeroplane
photography if imagery has to be generated on a frequent basis and/or has to meet certain
requirements (slow shutter speed, high resolving power, etc.). Although LAAP is already
applied by various researchers, its application is expected to even increase in popularity due to
the decreasing cost of UAVs on the one hand, and the digital (r)evolution of photographic
cameras on the other. Because the latter offer instantaneous feedback, LAAP results contain way less imponderables with results that become virtually predictable, while the current capacity of memory cards allows the airborne time to be maximise.
CHAPTER 5

Helikite Aerial Photography

A New Means of Unmanned, Radio-Controlled Close Range Aerial Archaeology

The content of this chapter is accepted for Archaeological Prospection [ERIH (B); SCIE, AHCI – IF 2007: 0.660]

Abstract

During the past century, various devices were developed and applied to acquire archaeologically interesting aerial imagery from low altitudes (e.g. balloons, kites, poles). This chapter introduces Helikite Aerial Photography or HAP, a new form of close range aerial photography suitable for site or defined area photography, based on a camera suspended from a Helikite: a combination of both a helium balloon and kite wings. By largely overcoming the drawbacks of conventional kite- and balloon-based photography, HAP allows for a very versatile, remotely controlled approach to Low Altitude Aerial Photography (LAAP). Besides a detailed outline of the whole HAP system, its working procedure and possible improvements, some of the resulting imagery will be shown to prove the usefulness of HAP for several archaeological applications.
5.1 Introduction

In 1908 L.P. Bonvillian took the first photograph from an aeroplane near Le Mans (France), although it was actually one single frame of a motion picture [243, 573]. A few years later and largely due to the technological catalysis of World War I, aerial photography by means of an aeroplane became a standard practice [230]. To date, active aerial photography still largely depends on these manned, heavier-than-air motor-driven aircrafts. In practice, individual archaeologists often acquire their own data from the cabin of a small, relatively low-flying, conventional fixed-wing aeroplane, utilizing 135 mm format (or slightly larger or smaller) hand-held still cameras to acquire imagery that is mostly oblique in nature. On some occasions, however, this conventional way of image acquisition is impossible (e.g. forbidden by the military), inconvenient (e.g. imagery must be generated twice a day) or unsuited to reach particular goals. As an example of the latter, one might think of beyond visible imaging (e.g. UltraViolet photography) or large-scale photography (e.g. 1:250), in which case the forward movement of the aeroplane is far too fast to compensate for the situation-specific shutter speed needed.

To deal with these issues and still be able to obtain qualitative imagery, archaeologists often resort to unmanned devices such as balloons, kites, model aircrafts, blimps, and poles to acquire imagery from the air. Generally, these devices allow the (digital) still or video camera to be more or less stationary over specific spots of interest at particular altitudes, difficult or impossible to achieve through all kinds of manned aerial platforms such as aeroplanes, powered parachutes, helicopters, balloons, ULMs (Ultra Légers Motorisés)/ultralight aircrafts, gliders, and paramotors. As most of these devices have a restricted operation height (e.g. 100 m), they are ideal to perform Low Altitude Aerial Photography (LAAP), also called close range aerial photography.

The system presented here is a new approach to such low altitude aerial archaeology, developed to acquire highly detailed (digital) aerial imagery in most occasions by means of a Helikite. Before outlining the system, the ground-based approaches mostly used nowadays will be shortly reviewed, as their characteristics will prove essential in showing the advantageous use of Helikites.

5.2 LAAP Platforms

There are several means used in archaeology and other scientific fields to lift Radio Controlled (R/C) or action delayed photo (or video) cameras and acquire large-scale imagery (see Chapter 4 for a detailed overview). In general, the following unmanned camera platforms are in use to
capture low altitude aerial imagery in archaeology, each with its distinct advantages and drawbacks:

- masts, poles or booms: although these platforms are cost-efficient, very portable and stable, they are limited by a moderate maximum operation height of 20 m;
- unmanned Aerial Vehicles or UAVs – Encompassing mostly R/C model aeroplanes and helicopters, this category is generally characterised by superior navigation possibilities, but problems with induced vibrations, cost and less straightforward operation still allow kites and balloons to be the most widespread LAAP platforms;
- kites: since the 1970s, Kite Aerial Photography (or KAP) is practised by many individuals and archaeological teams, as these highly inexpensive and portable platforms can accommodate a few kilograms of payload. Moreover, only wind is needed to make it work. This dependency is also its largest drawback, as irregular winds are not suited for KAP and the size of the kite is dependent upon the wind speed. It goes without saying that ‘KAPing’ is not possible in windless situations;
- balloons and blimps: these lighter-than-air devices fill in the gap characteristic for KAP, as they can be used in windless and very light wind conditions. Moreover balloon photography is extremely flexible in its setup and operation is easy. However, balloons and blimps become difficult to position and hold steady if the wind speed exceeds approximately 15 km/h.

This way, it is clear that a combination of a balloon and a variety of kites is often essential to maximise the conditions for low level aerial archaeology. However, rather than employing two or more separate lifting platforms (e.g. [2, 19, 61, 820, 825, 827, 838]), a Helikite can be used.

- Helikite: this design, patented by Sandy Allsopp in 1993 and currently manufactured by Allsopp Helikites Ltd., combines the two aforementioned constructions. By joining a helium balloon with kite wings (Figure 5.1), this lighter-than-air device combines the best properties of both platforms. The helium filled balloon allows it to take off in windless weather conditions, whereas the kite components become important in case there is wind: first of all, they lift the construction up in the air to altitudes higher than the pure helium lift. Moreover, the Helikite’s lift becomes stronger with increasing wind speed (with an upper lift limit depending on the Helikite’s size). Secondly, the wings counteract any unstable behaviour that is characteristic of balloons and blimps flown in windy conditions, hence stabilising the Helikite [31].
The Helikite’s distinct excellent all-round behaviour has also been reported by researchers of the Center for Transportation Research and Education (CTRE) at the Iowa State University, who compared the photographic conditions yielded by a kite, blimp, Helikite and balloon in several wind conditions (Figure 5.2). In this respect, the Allsopp Helikite is not only more versatile than comparable devices, but it supports more payload for its size when compared to ordinary aerostats, and operates in stronger winds than traditional blimps or balloons [31]. Due to the fact that it can be used in adverse weather (e.g. rain, fog, freezing conditions) it is also more flexible in operation than most UAVs.

Based on these properties, a complete photographic system using a 7 m$^3$ Skyhook Helikite (Figure 5.1) was designed in 2005-2006 by the Classical Archaeology section of the Department of Archaeology and Ancient History of Europe (Ghent University; Belgium), in
collaboration with the Department of Industrial Engineering (KaHo Sint-Lieven, Ghent; Belgium). For obvious reasons, this type of unmanned photography was inaugurated Helikite Aerial Photography or HAP.

5.3 HAP – The System

The requirement was to allow the acquisition of general overviews as well as highly detailed images of specific locations, both in the visible and invisible range of the ElectroMagnetic (EM) spectrum, so this complete system had to be stable, easily maintained and remotely controllable. In this section, the nine major components which are part of the finally assembled HAP system are described separately (and indicated on Figure 5.3).

Figure 5.3. The complete HAP system.

1. The camera-lifting device is a 7 m³ Skyhook Helikite that can lift a mass of about 3.5 kg in windless conditions at ordinary air pressure and temperature. Due to its wings, a 25 km/h wind allows for a buoyance of 100 N: i.e. 10 kg of payload. This value largely surpasses the gross lift of pure helium (He). With a gaseous density $d$ of 0.18 g/l (or 0.18 kg/m³) at 0 °C and 1 atmosphere (i.e. Standard Temperature and Pressure or STP) – compared to 1.29 kg/m³ at STP for air [533, 714, 851] – 1 m³ helium lifts at STP slightly more than 1.1 kg (i.e. 1.29 kg/m³ minus 0.18 kg/m³) of payload, a value that will alter with varying temperature and atmospheric pressure [209]. Consequently, a 7 m³ airship can have a maximum buoyancy of about 77 N (if its own mass was zero). However, the
design of the Helikite overcomes the physical restraints of its lifting gas. By way of comparison, Aber [1] mentions a helium blimp of 7 m$^3$ with a net lift of 2.3 kg, while Summers [747] utilised a 20 m$^3$ helium blimp, yielding 95 N net buoyancy. Table 5.1 gives an overview of the current (February 2009) Helikite product line, with the model specific (lifting) capabilities indicated;

Table 5.1. Helikite models and performance (adapted from [31]).

<table>
<thead>
<tr>
<th>Helikite type</th>
<th>Helium capacity (m$^3$)</th>
<th>Material thickness (Thou’ inch)</th>
<th>Lift in no wind (kg)</th>
<th>Lift in 15 mi/h (kg)</th>
<th>Max wind speed (mi/h)</th>
<th>Max unloaded altit. (feet)</th>
<th>Helikite length (feet)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vigilante</td>
<td>0.15</td>
<td>1</td>
<td>0.03</td>
<td>0.15</td>
<td>25</td>
<td>1 000</td>
<td>3</td>
</tr>
<tr>
<td>Lightweight</td>
<td>0.15</td>
<td>1</td>
<td>0.06</td>
<td>0.18</td>
<td>25</td>
<td>1 300</td>
<td>3</td>
</tr>
<tr>
<td>Skyshot</td>
<td>1.6</td>
<td>2</td>
<td>DSC</td>
<td>DSC</td>
<td>30</td>
<td>2 500</td>
<td>6</td>
</tr>
<tr>
<td>Skyshot</td>
<td>1.0</td>
<td>2</td>
<td>0.3</td>
<td>1.5</td>
<td>28</td>
<td>2 000</td>
<td>5</td>
</tr>
<tr>
<td>Skyshot</td>
<td>1.6</td>
<td>2</td>
<td>0.5</td>
<td>2.5</td>
<td>30</td>
<td>2 500</td>
<td>6</td>
</tr>
<tr>
<td>Skyshot</td>
<td>2.0</td>
<td>2</td>
<td>1.0</td>
<td>4.8</td>
<td>32</td>
<td>5 500</td>
<td>7</td>
</tr>
<tr>
<td>Skyshot</td>
<td>3.3</td>
<td>3</td>
<td>1.2</td>
<td>6.5</td>
<td>35</td>
<td>6 000</td>
<td>9</td>
</tr>
<tr>
<td>Skyshot</td>
<td>6</td>
<td>3</td>
<td>3.0</td>
<td>9.0</td>
<td>40</td>
<td>6 500</td>
<td>11</td>
</tr>
<tr>
<td>Skyshot</td>
<td>11</td>
<td>3</td>
<td>5.5</td>
<td>12.0</td>
<td>45</td>
<td>9 000</td>
<td>12</td>
</tr>
<tr>
<td>Skyshot</td>
<td>16</td>
<td>3</td>
<td>8.0</td>
<td>16.0</td>
<td>46</td>
<td>9 000</td>
<td>13</td>
</tr>
<tr>
<td>Skyshot</td>
<td>25</td>
<td>6</td>
<td>9.0</td>
<td>20.0</td>
<td>50</td>
<td>9 000</td>
<td>16</td>
</tr>
<tr>
<td>Skyshot</td>
<td>35</td>
<td>6</td>
<td>14.0</td>
<td>30.0</td>
<td>60</td>
<td>11 000</td>
<td>22</td>
</tr>
<tr>
<td>Skyshot</td>
<td>64</td>
<td>6</td>
<td>30.0</td>
<td>70.0</td>
<td>70</td>
<td>15 000</td>
<td>26</td>
</tr>
</tbody>
</table>

2. To securely fly the Helikite, an appropriate tether must be used, taking several considerations into account. First of all, higher flying (i.e. above 100 m) causes sag of the line due to gravity. Therefore, the tether’s mass is best kept to a minimum. Additionally, the line is also not allowed to stretch too much (certainly not when 300 m of line is used to reach maximum operation height), as the combination of both sag and stretch will make it very hard to keep tight control over the Helikite. Therefore, Dyneema, an extremely strong polyethylene fiber, is advisable [126], as it has only 5 % stretch and is more than ten times stronger than steel per unit of weight [246]. With a high breaking strain of 270 kg and a diameter of only 2.2 mm, the currently employed Dyneema tether allows the Helikite to fly securely and makes steering easy, as the operator can readily feel the connection with the Helikite. To attach the Dyneema line to the Helikite knots are inevitable. However, together with kinks or angles, knots stress the fibers of the tether unevenly, weakening the strength of the line. The degree of strength loss is largely knot dependent [472, 660], with particular knots weakening the line to 50 % of its rated strength. It is therefore safe to assume that the tether will never perform at more than half its claimed breaking strain [352, 660]. Hence, 270 kg Dyneema is still safe as a tether. Ideally, 500 kg Dyneema line should be used, as resistance to abrasion also needs to be taken into account. However, its mass and diameter would compromise the payload capacity too much in windless conditions and create problems with the amount of flying line that can be held by the reel (see 5.3.8);
3. The camera rig is attached to the tether some 20 m below the Helikite (to separate vibrations and sudden movements of the Helikite), and consists of a camera-supporting frame or cradle and a suspension system. The Picavet suspension applied in this system allows for self-levelling and securing the cradle (see next point). Named after its French inventor, the Picavet suspension system [624] can be applied in several variants [79, 403]. The one used in this project is a large rigid cross (known to be more twist resistant than small crosses – [319]) with each of the four ends (Figure 5.4: 1-4) connected at two anchor points (Figure 5.4: A and B) to the flying line. The latter is accomplished by means of a double pulley block, combined with a dedicated fishing hook that is attached to the line by means of a Brooxes Hangup. The small Ron Thompson DynaCable line (a micro filament fishing line of 0.25 mm diameter and a 20.2 kg breaking strength) which provides this connection is one looping string. Lastly, a ring constrains the two innermost crossing cables. The result is a simple, very lightweight dampened suspension for the cradle, superior to the often used pendulum [79, 403, 570, 772], and capable of minimizing camera swinging when manoeuvring (which changes the angle of the tether) as well as absorbing all kinds of vibrations (e.g. those induced by the wind), the latter being one of the prime requisites for convenient LAAP [266];

4. The sturdy aluminium and carbon cradle, the camera supporting part of the rig, was specifically designed and built by ir. Jo Loenders within the framework of his Master's thesis. Except for the four carbon legs and the carbon Picavet cross – allowing the
construction to stand and take-off independently and protect the still camera in case of a rough landing (Figure 5.5A) – the cradle is completely made of aluminium: a cheap, light, but sturdy, bendable, and easily drilled material that is often used to construct cradles [127, 273]. Due to its design as well as the solid, precisely lasered and bended aluminium frame profiles used, the cradle experiences a very low static nodal stress when loaded [487], allowing for extremely fluid rotations, the latter controlled by three small servo motors (type Graupner C577). Although existing cradles come in all kinds of designs [272, 370], many of them (certainly the older ones) only allow the camera’s orientation to be set before taking it aloft. The more advanced ones enable remote control of the attitude of the camera, generally allowing for rotation (0°-360°) and tilt (0°-90°). However, altering the camera’s orientation with only two Degrees of Freedom (DoF) will always exclude certain compositions to be made. Hence, this cradle was designed as to allow for three, R/C DoF of the camera (see Figure 5.5B): ω = -45° to +45° around X (tilt or roll), φ = -45° to +45° around Y (tip or pitch) and κ = 0° to 360° around the Z-axis (swing or yaw, but also called pan in this context). By implementing these three functions both vertical and oblique pictures can be taken, the latter with every possible orientation in relation to the object/site under investigation and/or position of the sun. Even though still a prototype, the current cradle largely fulfils the initial design goals, as it is: durable, easily operated, and steady with smooth rotations at the joints;

Figure 5.5. The cradle (A; photograph by W. Gheyle – Ghent University) which allows the camera to tilt, tip, and swing (B).

5. The cradle was designed to allow a variety of cameras to be mounted, but not more than one simultaneously. Although the initial testing was performed using a film-based Nikon F70 Single-Lens Reflex (SLR) camera, all currently cradle-mounted cameras are Digital Still Cameras (DSCs) of the SLR-type: the Nikon D50_NIR, D70s, and D80_FS. While the first SLR is a converted Nikon D50 that allows true Near-InfraRed (NIR) photographs to be taken ([776] or Chapter 8, Chapter 9, and Chapter 10), the Full Spectrum (FS) modified Nikon D80 enables NIR, visible, and Near-UltraViolet (NUV) photography ([777] or Chapter 12). The Nikon D70s is a conventional DSC which is used to acquire ‘normal’ visible photographs. To trigger the shutter of these SLRs, a gentLED is used. When connected to an R/C receiver, these tiny devices can be triggered by the
latter to emit InfraRed (IR) signals that can operate Digital Still and Video Cameras with IR receivers [319]. As all aforementioned Digital SLRs (D-SLRs) contain such a wireless receiver, they can be remotely operated by a gentLED SHUTTER (only one of the several gentLED options) to enable focusing and releasing the shutter. Moreover, D-SLRs suspended from the Helikite have many advantages: they are lightweight (620 g, 679 g, and 668 g respectively, batteries included), there is no restriction to 36 frames, the exposure can be calculated automatically (with aperture or shutter speed priority), and a wide choice of lenses with different focal lengths is available (at a range of qualities and prices). All four are essential features for convenient R/C photography [296];

6. To control the shutter and enable the steering of the camera, a six channel proportional R/C hand-held transmitter (type Graupner X-412 UNIT 35 MHz – Figure 5.6) uses radio signals of 35 MHz to wirelessly send the commands given by the operator to a proportional six channel receiver (type Graupner R700 miniature SUPERHET) mounted on the cradle. This receiver, for its part, controls the three small Graupner servo motors to make the camera rotate in all possible directions, while a fourth receiver channel is used to trigger the gentLED and allows the DSC to focus and take a photograph;

7. Being unable to directly observe the area photographed is one of the biggest disadvantages of certain close range aerial photographic solutions [375], as it is very hard to estimate what the still camera will exactly photograph [159, 381, 587, 747]. To counteract this, a direct video link was established using the Pro X2. This very handy plug & play video system consists of a tiny 4.8 V Hi Cam EO5-380 CCD camera (WxHxD = 30 mm x 25 mm x 28.6 mm) attached to look directly through the camera’s eyepiece; a connected micro FM (Frequency Modulation) transmitter to send the video signal wirelessly to the ground; a ground-based audio/video receiver with a patch antenna (8 dBi) to pick up the signal and feed it to a small monitor (1440 pixels x 234 pixels). In this way, the TFT (Thin Film Transistor) screen, which runs on a 12 V battery, instantaneously displays the area seen by the camera-lens combination (Figure 5.6), allowing the camera operator(s) to correctly orientate the D-SLR, compose the shot and decide whether or not to take the image. Moreover, as the viewfinder display also shows useful information on the focus, the number of exposures remaining, the battery status, aperture, and shutter speed, the camera operators can check the DSC’s normal operation and verify when the memory card is full. Thanks to its compactness and low mass (about 65 g, the 4.8 V battery needed to feed both the video camera and transmitter not included), the Pro X2 system is ideal for R/C aerial photography. Moreover, the 2.4 GHz transmitter’s output power of 200 mW allows the video signal to be sent over about 300 m line-of-sight [386];
8. One can imagine the tractive force of the Helikite when flying even in moderate winds. Using a big game fishing lever drag reel (type Shimano Tiagra 80W) and accompanying carbon sea-fishing rod (type Shimano Tiagra Trolling 80AX), these forces can be managed reasonably well and allow the Helikite’s operator (i.e. the navigator) to freely walk around while letting out tether smoothly and quickly, using the rod to guide the line. Fixing a large and solid winch to the ground could be more convenient to pull down the Helikite, but severely restricts steering and risks the penetration of archaeological layers (e.g. floors). With a mass of 3.2 kg, a big winder handle and a two speed Gear Ratio (1:2.5 to take in line fast and 1:1.3 for high power retrieves, a feature which is essential for flying kites/Helikites [271]), this reel is capable of holding at least 300 m of the specified Dyneema line. When the spool is completely filled, the reel’s maximum drag is about 18 kg and therefore sufficient to operate in winds of at least 30 km/h. The reel’s construction also counteracts the faster winds that can be expected at higher altitudes, as its drag increases with reduced line level. In practical terms, the reel’s drag force will be doubled to some 36 kg when approximately 225 m Dyneema line (about 75 % of the full spool) is off the reel, although these figures should not be followed too strictly, as many external factors can affect drag performance [710]. This means the reel can always be slowed down and even stopped completely by the Helikite’s navigator, as 35 km/h is determined to be the maximum ground wind speed to safely and conveniently perform HAP (hence covering about the same operating range as a kite and blimp combination). The fishing rod and reel are attached to the navigator’s body using a big game fishing harness and a Tsunami TS-A-1 Gimbal Utility Belt (Figure 5.7). Although this combination is primarily designed to allow the body to deliver maximum pulling leverage when the rod is drawn downwards – rather than
upwards as is the case in HAP –, it is still very useful when walking around, pausing or reeling in the Helikite, as the rod will always rest in the gimbal belt, while the shoulder harness makes sure the reel and rod stay securely attached to the navigator’s body in every situation;

Figure 5.7. Fishing rod and reel attached to the navigator’s body (photographs by W. Gheyle and D. Van Limbergen respectively).

9. As a fast running tether (and even Dyneema line under tension) can severely cut exposed skin, gloves are of the utmost importance. Therefore, both the navigator and the persons assisting in attaching the cradle to the tether (generally the camera operators), always wear Marigold Industrial Kevlar gloves (FB20PD).

The combination of all these nine elements forms the complete HAP system.

5.4 HAP – Cost, Maintenance, and Operation

Apart from the initial purchase cost of the Skyhook Helikite with accompanying Dyneema tether (circa € 4000 and funded by the UTOPA Foundation (Voorhout, The Netherlands)), the fishing gear (€ 1350), and the Pro X2 (about € 550), the price tag of other necessary parts (as TFT screen, servos, R/C transmitter, batteries, flight case, gloves, etc.) was very moderate. In the end, the building of the complete HAP system – with its prototype cradle – was about € 8000. Besides being affordable, HAP’s running costs are low, because all equipment is driven by rechargeable batteries and the construction is very cheap to maintain. As the buoyancy comes from helium (He), the only additional cost when applying HAP is the need for this completely inert, non-toxic, colour-, taste- and odourless noble gas. The non flammable helium is transported in a pressurised B50 cylinder containing 10 m³ of this gas, priced at about € 150 (+ additional rent for the cylinder). Once inflated, the Helikite only needs to be topped up with a minimal amount of helium twice a week. As a result, one B50 cylinder proved to be largely
sufficient for one month of HAP over one area. When different locations have to be photographed, a delivery van is rented to store the Helikite partly deflated and transport it to the area of investigation, as it is not feasible to recover the helium from the balloon in the field, while completely refilling a 7 m³ Helikite daily would be too expensive. In practice, one week of intensive HAP (on six different sites) was completed with only one 10 m³ volume cylinder. To cut down costs further, the purchase of a large trailer that enables the storage of the whole, inflated system is considered. Moreover, the trailer could act as a protective hangar. Since a Helikite crash is almost out of the question, additional expenses due to broken equipment can largely be prevented.

To apply HAP in practice, the Helikite is first inflated to its desired pressure using a tapered foil filling outlet mounted onto the helium cylinder and slid into a plastic plug which is connected to the Helikite’s non-return valve. Afterwards, the DSC and lens are mounted onto the cradle and all mechanisms thoroughly checked twice. Once the maximum wind speed is verified with a hand-held anemometer and the Dyneema line attached to the Helikite using a karabiner, the Helikite operator attaches the reel and rod to his/her body by means of the harness. While the aerostat is sent skyward, the R/C transmitter and receiver as well as the Pro X2 are activated and tested once more by the camera operator(s). As soon as the Helikite is about 20 m aloft, the rig is attached to the tether and finally more line is let out to get the complete system skyward. To avoid any punctures in the Helikite’s delicate surface, all aforementioned operations take place on a large and thick canvas.

Once the system is completely and safely airborne, the Helikite’s navigator walks around to establish the right position and height for image acquisition, while at least one camera operator (preferably two) determines the DSC’s angle and decides to ultimately shoot the aerial imagery. Constant communication and coordination between both teams is enabled by two-way radios and is deemed absolutely crucial for accurate positioning of the camera, flight planning and signifying the presence of power lines and potential conflicts with occupied aircrafts – an issue not to be taken too lightly [75, 74] and indicating the crucial need for this second camera operator. Finally, some training and experience are vital to yield above average results and to make sure efficiency and reliability in image acquisition remain constantly high.

5.5 HAP – Possible Improvements and Drawbacks

Even though HAP completely meets most points identified by Walker and De Vore [807] and Schlitz [688] for effective LAAP (i.e. low velocity, small take-off and landing space, portability, low cost, minimal operational staff, low vibration, reliable power supply, fast to set up and employ, low risk and low impact) and its advantages over conventional kite and balloon photography are evident, the system is not perfect. Aerial imagery with a high spatial and temporal resolution can be generated in several wind conditions, but the camera cradle is not
(yet) fully weather proof, making it impossible to use in case of rainfall (although one can doubt the usefulness of aerial photographs taken during such wet conditions).

Secondly, as walking the Helikite around and allowing it to ascend or descend to various altitudes is the only way this aircraft can be moved into position, it remains rather challenging to accurately establish a precise camera location and/or take photographs along a previously outlined track – a problem encountered with all forms of photography using kites and blimps [432, 562, 718, 756]. Currently, the position of the D-SLR is largely determined by looking both at the Helikite’s location and the transmitted video image on the ground-based monitor; by passing the required instructions (i.e. higher, lower, left, right) on to the navigator, the latter decides how to move with the Helikite – taking the direction of the wind and local topography into account – in order to get the DSC where it should be. This approach has already proven to be a very convenient way of working. However, in cases where ground conditions are very monotonous (e.g. a very extensive corn field), the camera operators will struggle to get properly oriented. In an attempt to counteract such issues (and improve the positioning in general), the signal of a very small, cradle mounted GPS receiver with WAAS/EGNOS (Wide Area Augmentation System/European Geostationary Navigation Overlay Service) capabilities will in the near future be transmitted to the ground. Its accuracy of geolocation, about 3 m at $2\sigma$ RMSE (Root Mean Square Error), should be very helpful in establishing the required geodetical position of the DSC, whereas the complete flight path (which is continually logged) will be available afterwards to geocode the photographs and visualise the camera’s 3D-position through time.

Thirdly, the possible places of survey can be limited by objects that may conflict with the tether: trees, high tension power lines, houses, scrub, etc. Furthermore, there must be a suitable place for the operator to stand for the desired shots given the specific direction of the wind, the length of the tether and general topographic setting. Consequently, the positioning capabilities of UAVs still remain superior for R/C LAAP.

Finally, the helium dependency might in some situations be the largest drawback: besides its cost, it might not always be possible to purchase helium locally and get it afterwards on site; in some situations, this can be a reason not to opt for such helium-filled devices (e.g. [28, 40, 83, 98, 587, 589, 696, 827]), although some authors also point to the fact that helium allows for very silent operation and lets the vehicle to be aloft for extended periods of time [510], while Aber [2] favoured helium blimps over hot-air blimps due to reasons of field operation and handling, dimensions and cost.

5.6 HAP – Archaeological Applications

Notwithstanding some inevitable drawbacks, HAP photography has been rigorously tested since 2006 and proved to be a stable, easily maintained, and versatile radio controlled system.
From 2007 onwards, a large amount of close range archaeological image data has been generated at several geographical locations in varying weather conditions. Although HAP was initially developed to allow for analogue NIR site photography [781], its application became much wider and now allows for various types of aerial archaeological applications.

5.6.1 General Overviews and Small Area Reconnaissance

With a maximum operating altitude limited – for the moment – to about 200 m, a vertically oriented DSC can record approximately 240 m x 160 m when fitted with a 20 mm lens, yielding a scale of 1:10 000. If some obliqueness is allowed, the area captured can be much larger. Such obliques are suited to the overview of a large site and/or its direct environment. If climatic and environmental conditions are favourable, these overviews can also yield new archaeological information, as was the case in Figure 5.8. While acquiring imagery to illustrate the direct relationship between the Roman mausoleum (1) and the western gate (2) of the Roman city Potentia (Adriatic Italy, Regione Marche), the trajectory of the decumanus maximus’s extra muros prolongation became largely apparent as a very distinct negative crop mark. Given the fact that this feature had previously been unnoticed in the grassland, two new conventional reconnaissance flights were initiated. Hence, HAP also gave clear indications about the information one could expect when stepping into a small aeroplane. The latter method thus still remains mandatory, because – just like all unmanned systems but the very heavy, military based UAVs – HAP is impractical for surveying geographically extended areas.

Figure 5.8. Decumanus maximus leaving the western gate of Potentia (HAP with Nikon D70s + Nikkor 20 mm f/3.5 AI-S).
5.6.2 Site Overview for Documentation and Interpretation

On a much larger scale (ca. 1:500 to 1:5000), highly detailed overviews of archaeological sites can be produced (Figure 5.9), imagery that can suit several purposes: documenting the progress of ongoing excavations and the several field phases (e.g. the different layers excavated), generating imagery to use in presentations, folders and books, revealing minute aspects about individual features that cannot be seen in plans or from traditional images as well as aiding in the interpretation of a site. In case the site is too extensive to be caught in one frame, a multitude of overlapping photographs can be used to generate site encompassing mosaics (e.g. [19, 587]). This way, HAP bridges the gap between the lowest conventional aerial photography and the highest ground supported pole photography. Although one could acquire such imagery without a video live link, the ability to see what the D-SLR will capture is a very welcome means in establishing a cost-efficient workflow, because it allows the operators to frame and compose in a convenient way instead of shooting dozens of frames approximating a workable image of the subject under consideration.

![NIR photograph of Potentia's temple area (HAP with Nikon D50\textsubscript{NIR} + Nikkor 20 mm f/3.5 AI-S).](image)

5.6.3 Site Mapping and Photogrammetry

Even though such close range photographs have not always to result in plans, HAP is well suited to the acquisition of stereophotos to subsequently generate topographic surfaces and accurate planimetric information by means of photogrammetry – the latter process already being explored since the nineteenth century [85]. In a similar way, stereoscopic photography has already been performed several times with kites, blimps, and balloons (e.g. [163, 705, 802, 810, 815, 820, 821, 827, 852]). Although there are people that fly stereo camera rigs (e.g. [7, 83, 471, 629, 630, 631, 684]), the baseline provided by these constructions is often too small to yield stereo imagery for precise height measurements [67]. Therefore, the mono-DSC of the Helikite has to be moved from one point to another, so as to generate a multitude of stereo
image pairs with a certain overlap. In order to perform the generation of complete excavation plans or intra-site maps as smoothly as possible, the cradle is set to acquire (near) nadir photographs (i.e. vertical). Because the appropriate **Ground Control Points (GCPs)** – marked prior any excavation and measured by a total station survey – must be incorporated into the imagery (Figure 5.10), such very-low altitude mapping needs rather precise framing [99], making the video live link a necessity. However, the end products will allow the generation of maps much faster and more consistently than most other, low cost techniques [399], while the textural and colour information provided by the orthoimages remains essential to complement expensive 3D laser scanning [705]. This way, HAP is consistent with the view of Żurawski, who stated that all ongoing excavations should have a “handy, reliable and inexpensive vehicle capable of shooting aerial pictures at a particular moment of the fieldwork” ([852]: 244). The fact that this approach is not restricted to conventional sites, has been illustrated by several scholars who used similar unmanned lighter-than-air constructions to map underwater remains (e.g. [419, 562, 563, 820, 821, 824, 825, 827, 828]).

![Image of excavation area with GCPs](image)

**Figure 5.10.** Near vertical photograph of an excavation area with GCPs indicated (HAP with Nikon D70s + Nikkor 20 mm f/3.5 AI-S).

### 5.6.4 Monitoring

Due to its ability to generate imagery with an extremely high **temporal resolution** (i.e. the ability of a system to record images at a certain time interval – e.g. one day versus one month), a fast response to events and detailed site-based monitoring at short time intervals is possible
with HAP (Figure 5.11). As an example, one could base a flying strategy on the outcome of such multi-temporal, sequential data: as soon as the first crop marks start appearing in the HAP imagery over a known crop mark-sensitive site, the urgency to begin conventional reconnaissance flights is indicated.

Figure 5.11. Monitoring excavation phases (HAP with Nikon D70s + AF Nikkor 50 mm f/1.8D).

### 5.6.5 Multispectral Sensing

Being a stable system, HAP is suited for those situations where long shutters speeds are inevitable: low light conditions, narrow-band, and non-visible remote sensing. As previously mentioned, the initial aim in developing HAP was to perform close range film-based NIR photography [781]. However, it soon turned out digital NIR photography is far less cumberstone, with much shorter exposure times compared to the analogue technique (see Chapter 8). As an example, Figure 5.12B shows an NIR record of a tower and connected piece of wall belonging to the central Italian Roman city of Septempeda, while Figure 5.12A displays a conventional aerial frame of the same location, taken one day later but with the anomalies imaged in a less distinct way. Since the summer of 2008, the Helikite-based system has also been applied as a research instrument for digital NUV photography by means of the modified Nikon D80s ([777] or Chapter 12). The same D-SLR will also be used in the near future (i.e. summer of 2009) to explore narrow-band photography, together with NUV imaging aimed at better revealing sub-surface structures.

### 5.7 Conclusion

Fitting a (modified) DSC to a Helikite allows for low level photography during conditions in which a tethered balloon or a kite would fail to work properly. By holding a 7 m³ unmanned, helium-filled Helikite aloft with a tether, different low level aerial photographic methods can be employed as the device is largely scale-independent. Although the solution is not perfect, HAP enables photography at particular times of the day, in varying weather conditions, is
Helikite Aerial Photography

uncomplicated to deploy, has a large range of operation altitudes, and is easy to maintain. Hence, it is often more cost effective and flexible than other individual approaches in yielding high spatial and temporal resolution coverage.

Figure 5.12. Visible (A; photograph by F. Vermeulen – Ghent University) and pure NIR record (B – HAP with Nikon D50_{NIR} + Nikkor 20 mm f/3.5 Ai-S) of a Roman town wall and connected tower.
"There is nothing worse than a sharp image of a fuzzy concept."

Ansel Adams (1902-1984), photographer
Cropping the Aerial View

On the Interplay between Focal Length and Sensor Size

The content of this chapter is published in *Archeologia Aerea*

6.1 Introduction

The last decennium has been very remarkable. With the advent of digital photography, a new world has opened for many people. Although photography can almost celebrate its second centennial, there has never been a moment in history that so many people actually had and used a photo camera. New technologies often introduce a totally new technical jargon, and photography has not been an exception. Although photography already had its own language, digital photography added many new terms.

In spite of the overwhelming amount of – supposedly – informative literature in the form of books, articles in magazines, and items on the WWW (World Wide Web), most of these terms often stay mysterious for most practitioners. The fault lies not always with those seeking information, as a large portion of this information flow is just misinformation floating around (not in the least on the WWW). One of these often heard and read delusions is the fact that the focal length of lenses changes when used on a digital camera, leading to strongly magnified images as indicated by the digital magnification factor. By outlining the concepts of focal length and image sensor size as well as their combined effect on field angle of view, it is the purpose of this chapter to show why the above statement is not correct. Furthermore, the quantification of angle of view will prove useful in the classification of lenses.

6.2 Film Format Does Matter

The amount and diversity of both film and digital photography formats are enormous, ranging from extremely small to awfully large. In order not to end up comparing apples to oranges, it is expedient to restrict this chapter to the formats (both film and digital) aerial archaeologists mostly use(d).

In oblique archaeological aerial reconnaissance, a very common photographic emulsion definitely is/has been 135 general-purpose transparency or negative film (Figure 6.1). Introduced in 1934, the 135 film format has 16 perforations per 72 mm and mostly uses frames which can hold an image of 36 mm by 24 mm (approximately 1.42 in x 0.94 in), having an aspect ratio of 3:2 (i.e. width/height). Because the film is 35 mm wide, the creation of images by cameras and lenses suited for such 135 format emulsions became commonly known as 35 mm photography. Moreover, its popularity also gave rise to the term standard format photography.
Chapter 6

Figure 6.1. The dimensions of a 35 mm frame.

As film only indicates the specific combination of a base or support covered with a photographic emulsion [416, 417], it is not limited to this one format. Although many people are not aware of this due to the fact that 135 film is the most sold film format worldwide, smaller as well as larger formats exist. Compared to the subsequently outlined larger film formats, the small dimensions of the emulsion caused 35 mm photography also to be termed ‘small’ and ‘miniature format’ photography. Other existing film formats that were commonly used in the last decennium are given below [417, 812]:

- **110** or **sub-miniature format**: introduced by Kodak in 1972, such films of exactly 16 mm wide contain 13 mm x 17 mm frames;

- **APS** or **Advanced Photo System**: although it was launched rather recently (1996), this 24 mm wide film never became really popular, despite some advantages over usual 135 film. Firstly, it allows changing rolls mid-roll as the negative film never comes out of the cassette. Moreover, most APS cameras enable the recording of three different formats on the frame of 30.2 mm by 16.7 mm. Besides the format which utilizes the whole frame and has an aspect ratio of 16:9 (‘H’ format), cropping permits the recording of ‘classic’ (C) pictures (25.1 mm x 16.7 mm and 3:2 as aspect ratio) and P or ‘panoramic’ photographs (with a dimension of 30.2 mm x 9.5 mm and an aspect ratio of 3:1). Furthermore, most APS cameras record metadata as date and time of image capture besides exposure data as aperture and shutter speed;

- **120, 220** and discontinued **620 format**: these films enable so-called **medium format** photography and encompass a wide variety of frame sizes with different aspect ratios: 6 cm x 4.5 cm, 6 cm x 6 cm, 6 cm x 7 cm, 6 cm x 8 cm, 6 cm x 9 cm, 6 cm x 12 cm and 6 cm x 17 cm. Although all frames always have one size of 6 cm (2.36 in) wide, the most common have always been the square 6 cm x 6 cm and the rectangular 6 cm x 4.5 cm frames (which is still often denoted as ‘645’). In fact, 6 cm x 6 cm are dimensions only used for convenience, as the actual film frame is 56 mm x 56 mm – which is rather logical if one knows the whole film is 60 mm wide. Accordingly, 6 cm x 12 cm is 56 mm x 112 mm. The 120 film – introduced by Kodak in 1901 – differs from the 220 film (launched in 1965) in length, the latter enabling twice the amount of exposures per roll.
In fact, medium format photography uses any film size between standard 35 mm and the following large format films:

- **4 in x 5 in, 5 in x 7 in, 8 in x 10 in, 11 in x 14 in, and 20 in x 24 in:** The dimensions of these films – which are always expressed in inch – cause it to be only available as sheet film and allow large format photography with superb, unsurpassed results. Moreover, each sheet can be exposed and developed individually. A film which could also be classified as large format is the 230 mm square format film (i.e. 230 mm x 230 mm or 9 in x 9 in). Although not available for ‘normal’ photographic purposes, this film is commonly loaded in metric mapping cameras used in vertical aerial survey. The total width of these particular films is 240 mm with a possible length of up to 120 m.

As small format cameras commonly enable 36 frames to be taken without reloading the camera and allow a maximum freedom of movement in handling the camera and lenses, 35 mm photography is ideal for aerial oblique imaging, where easiness of use and flexibility to work under varying light conditions are primordial. The main disadvantage often quoted for this format is the fact that the small negatives are more difficult to enlarge successfully compared to their larger format counterparts.

Instead of emulsions, digital photography uses a digital image sensor to record the scene: in most cases a solid state electronic device such as **CCD (Charge-Coupled Device)** or **CMOS (Complementary Metal Oxide Semiconductor)**. Although also medium and large format cameras can be equipped with digital sensors, 35 mm-like digital cameras are most popular with amateurs and a lot of professionals for the reasons mentioned before. This chapter will therefore focus on standard 35 mm analogue photography and the small format in digital photography. Besides, all cameras considered are **(Digital) Single-Lens Reflex cameras**, abbreviated (D)-SLRs.

![Figure 6.2. The path of a light ray in an SLR body with mounted photographic lens.](image)
SLRs got their name from the fact that in-camera reflection permits using only one lens for both the creation of the photograph and viewing the scene (Figure 6.2). Hence, what is seen through the viewfinder is exactly what will be photographed. Moreover, SLRs allow switching lenses, which is important for this chapter as a whole range of focal lengths will be tackled. Both features are unavailable with compact point and shoot and hybrid/bridge/SLR-like camera models. Although small format (D)-SLRs will be used to prove the initial statement incorrect, all principles discussed also hold for other image formats and camera types. The restriction is just for convenience when choosing examples.

In this context, it is important to note that SLR is not only applicable to the 35 mm format. Throughout the 20th century, they were produced for most film formats, enabling also 110, APS, medium, and large format SLR systems. From the 1990s onwards, also Digital SLRs (D-SLRs) were produced.

### 6.3 Small Format Image Sensors

Figure 6.3 displays different sensor sizes. The upper row belongs to compact and hybrid cameras and is not of concern here (note the 4:3 aspect ratio of all). It is obvious that these simple point-and-shoot cameras have various sensors at the very low end of the possible sensor dimensions (till the 2/3” format). Their strange kind of designation is rooted in the size of TV camera tubes of the mid-20th century. These glass tubes had typical outer diagonal sizes like 1/1.8” (i.e. 0.56 in or 14.11 mm), 1/2” (0.50 in or 12.70 mm), 2/3” (0.7 in or 16.93 mm), etc. The usable image area of such a tube was less, being more or less two thirds of the tube’s diameter. In the case of a 1/1.8” tube, the image area was 7.18 mm x 5.32 mm with a diagonal of 8.93 mm (2/3 of 14.11 mm would yield 9.41 mm). An image sensor which equals these dimensions is nowadays called a 1/1.8” image sensor. It is strange that, even today, sensor manufacturers still insist on using these outdated glass tube diameters to indicate the format of the sensor.

The lower part of Figure 6.3 depicts different small format D-SLR sensors with dimensions starting from 4/3” onwards. Although their sizes vary to a rather large extent, the majority of D-SLRs uses digital sensors with dimensions that are smaller than the frames of the 135 format. Nevertheless, due to the larger dimensions (compared to the compact cameras), the costs of producing sensors increase significantly [12], hereby partly explaining the price difference between D-SLRs and other, smaller cameras.

Generally, three main categories of small format D-SLR sensor sizes can be discriminated (note that the camera types mentioned date from 2006-2007; because camera manufacturers constantly update their product line, it was deemed unnecessary to include newer models, as this would not change the aim of this list):
- **the Four Thirds (4/3”) format**: the whole sensor measures 18.0 mm x 13.5 mm, thus covering about 25% of the 135 format. It was introduced by Olympus and Panasonic to cut down the costs of sensor manufacturing. In August 2008, these companies also announced their Micro Four Thirds system which shares the same sensor size, but allows smaller bodies to be designed [586];

- **the APS-C format**: as mentioned before, an APS ‘classic’ (C) negative is 25.1 mm x 16.7 mm large. Digital sensors approximating this size are called APS-C sized digital sensors. Most D-SLR camera manufacturers stick to this standard. As they all approximate the APS-C dimensions, these sensors are characterized by a lot of size variation [134, 310, 451, 577, 608, 713, 729]:
  - 20.7 mm x 13.8 mm, the dimensions of the Foveon X3 sensor used in Sigma’s SD9 and SD10 D-SLRs;
  - 22.2 mm x 14.8 mm in the Canon EOS 350D;
  - 22.5 mm x 15.0 mm for both the Canon EOS 20D and EOS 30D;
  - 22.7 mm x 15.1 mm in Canon’s EOS 300D;
  - 23.0 mm x 15.5 mm in Fujifilm’s FinePix S3 Pro;
  - 23.3 mm x 15.5 mm by the Nikon D2Hs (most sensors used in Nikon’s D-SLR line have approximately the same dimensions, called the DX format by Nikon);
  - 23.5 mm x 15.7 mm in the Pentax K100D, K110D and *ist DL;
  - 23.6 mm x 15.8 mm by Sony’s α 100 as well as the Nikon D80 and D200;
  - 23.7 mm x 15.6 mm in both the Nikon D50 and D70s along with the Konica Minolta Dynax 5D and 7D;
  - 23.7 mm x 15.7 mm in Nikon’s D2Xs;
  - 28.7 mm x 19.1 mm in the Canon EOS 1D Mark II N;

Figure 6.3. Different digital sensors with their characteristics.
• the **Full Frame (FF)** format, which equals the dimensions of classic 135 format film: 36 mm x 24 mm. To date (2009), several D-SLRs have sensor sizes that match a 35 mm film frame.

Figure 6.4. The Pythagorean Theorem.

Figure 6.3 also gives the diagonal distance of each of the sensors. In the case of 135 mm film, a frame has a diagonal of 43.27 mm. This can easily be calculated by using the Pythagorean Theorem (Figure 6.4). The value of this diagonal has two meanings. First of all, a lens with the same focal length as this diagonal is denoted a standard lens for that particular format (see *infra*). Secondly, this numeric value indicates the diameter of the circle which completely circumscribes the rectangular image frame. In this particular example of 135 film, a circle with a minimum diameter of 43.27 mm can completely encompass the whole frame. This circle is also shown for all different sensors in Figure 6.3. The tinier the sensor is, the smaller will be its encompassing circle.

### 6.4 Some General Lens Properties

Before delving deeper into the world of image sensors and the importance of their encompassing circle, it is beneficial to present some basic principles and terminology of optical image formation and lens systems (for more information on this topic, consider [234, 292, 424, 433, 443, 475, 490, 537, 554, 649, 651]). In a photographic lens, image formation is achieved by different optical lens elements which alter light in different ways. It is the combination of all these elements – usually made of glass – that form the photographic lens, which therefore truly can be seen as a multi-element system (see the lens in Figure 6.2). However, to establish an understanding of some theoretical principles and definitions in image formation, a **simple lens** model will be utilized instead of such a ‘real world’ **compound lens**. The simple lens employed here is classified as a **thin lens**: an imaginary, theoretical lens with zero physical thickness. Consequently, light rays refract along a single plane. It must be clear that this thin lens model does not provide the complete picture, as real lenses do have a finite thickness and two refracting surfaces. This thin lens model is therefore not an exact description of a real photographic lens, but it is sufficient to show the most important lens principles and terminology.
Cropping the Aerial View

By using Figure 6.5 which depicts a **single lens system** with a thin lens, the most fundamental lens constants can be sketched. The centre of such a system is the thin lens (here depicted with a certain width, to improve the visualisation as a lens) in air – so the amount of refraction is identical on both sides of the lens. In front of it, the **object space** can be defined. The lens is located at a certain distance from the object: the **object distance** $s$. This lens will form an image in the so-called **image space**, the region behind the lens. Accordingly, the distance between lens and image is referred to as the **image distance** $s'$. 

Through the centre of the lens runs the optical axis on which two **focal points** can be found: $F$ (primary or object-space focal point) and $F'$ (secondary or image-space focal point), respectively lying in object and in image space. The distance in mm from these points to the centre of the lens is the so-called **focal length** of the lens. As every lens has a pair of focal points, two corresponding focal distances $f$ and $f'$ exist (equal in size and opposite in sign). When one buys a lens, $f'$ is quoted. Together with the focal length, both the image and object distances (measured parallel to the optical axis) are related by the thin lens equation:

$$\frac{1}{f'} = \frac{1}{s'} - \frac{1}{s}$$

This equation clearly shows that the focal length of a lens equals the distance (expressed in mm) from the lens to the film or sensor plane when focused on an object at infinity. In such a case, $s$ is equal to infinity, yielding $1/s = 0$ and leading to $f' = s'$. Otherwise stated: focal length equals image distance for a far subject. To focus on something closer, the lens is moved further away from the film, producing an image distance $s'$ which surpasses focal length $f'$. In aerial photography, objects are so far away that the object distance is quasi infinite. Consequently, the image distance may be accepted as equal to $f'$ (i.e. all images are formed in the same plane,
one focal length behind the lens). To illustrate this with an example, consider a lens with a focal length of 50 mm and an object distance (earth surface to aerial photographer) of 300 m. The thin lens equation gives: 1/50 mm = 1/s' - (-1/300 000 mm) \( \Rightarrow \) 1/s' = 1/50 mm - 1/300 000 mm \( \Rightarrow \) s' = 50.01 mm. Notice that the image distance is marginally greater than the focal length. The object distance is negative, because the straightforward graphical sign convention for image formation is used here (Figure 6.6).

Figure 6.6. The graphical sign convention.

The linear magnification \( m \) of the real-world object is defined by the ratio of the image/object distance or the proportion of image size to object size: \( m = s'/s = h'/h \).

As can be seen in the illustration, the image is upside-down here, so the magnification \( m \) will be negative (due to the negative object distance). Because \( s' \) equals \( f' \) in aerial photography, \( m \) becomes \( f'/s \). In other words: dividing focal length by flying height delivers the scale of an aerial image.

Figure 6.7. A thick lens system in air.
It was already mentioned that real imaging systems like photographic lenses do not consist of a single thin lens, but can be considered compound lenses: a combination of multiple converging and diverging lens elements having a finite thickness, all mounted concentrically with the optical axis of the device. To characterize the whole lens system, two special refracting planes perpendicular to the optical axis, called principal planes, have to be located (Figure 6.7). The principal planes are conjugates and coincide with so-called nodal planes for a lens system in air. Where the optical axis intersects these planes, the principal points $P$ and $P'$ as well as the nodal points $N$ and $N'$ are found. Although the focal length of such a thick lens is defined as the distance from the rear or secondary nodal point $N'$ to the focal point $F'$, the earlier given definition for focal length will still be used here, both for the sake of simplicity and the fact that the thin lens in air provides a good approximation to the behaviour of real lenses in many cases. Moreover, a thin lens in air can be considered a thick lens where the planes $N$, $N'$, $P$ and $P'$ are coincident and go through the lens’s optical centre.

### 6.5 Tracing Rays to Reveal Focal Length Principles

Tracing the path of the light rays from the object through the lens (a process called ray tracing) allows to determine where the rays intersect to form the final image [433, 554, 651]. More specifically, the location of the generated image in such a thin lens system is defined by three key rays from the object. The first of these rays is the parallel ray (P-ray), a ray from the object, parallel to the optical axis. In the lens, this parallel ray will refract and be redirected on a path which takes it through the image focus (see Figure 6.8). The second key ray is the ray directed to the centre of the thin lens and will therefore not deviate from its path. This ray is called the chief ray. These two rays are sufficient to locate the image, but often a third ray is also included to better identify the image location. This F-ray (focal point ray) is drawn by extending a line from the object, through the primary focal point and to the lens. Since this ray passes through the object focus, it will emerge parallel to the axis after refraction. The two outer rays that pass through the edge of the lens are called marginal rays.

![Figure 6.8. Ray tracing for a thin lens in air.](image)
A differentiation can be made between converging and diverging lenses, both having a different effect on light. In particular, the latter will not form an image behind the lens, but yield the projection of a **virtual object** in front of it. As a converging lens is used in the illustrations, the image is said to be **real**. (Note that, to perform ray tracing in a compound lens, only one lens element at a time is considered. After the image position relative to the first lens element is found, this image will act as the source for the subsequent lens element, and allows determining the next position of the image with respect to this second lens element. The object is simply traced through the succession of different lenses).

Figure 6.9. Using ray tracing to show the influence of object distance on image distance and magnification.

Besides locating the final image, ray tracing also permits to visualize the fact that focal length equals image distance for a far subject. The larger the object distance from a lens with a particular focal length, the closer to the focal point $F'$ the image is created (Figure 6.9). This way, one can imagine that an object at infinity yields an image that is created in the plane perpendicular to the optical axis at the location of the focal point, or – as already concluded
before from a non-drawing perspective – the focal length is the distance from the optical centre of the lens to the principal focal plane (film or digital sensor) when the lens is focused on a subject at infinity. If an object comes closer, the lens must be moved further away from the imaging area to create a sharp image on the latter. Most lenses therefore extend when focussing on objects close at hand. The same illustration also indicates that great object distances (relative to the focal length $f$) lead to a very small magnification $m$.

Focal length is one of those primary physical characteristics of a lens that can be measured accurately and remains the same no matter what camera the lens is mounted on. A lens with 50 mm focal length will always be a 50 mm focal length lens, while a 600 mm focal length lens constantly remains a 600 mm focal length lens. The longer the focal length, the more ‘tele’ effect it has, meaning it has the ability to display distant objects large. This is illustrated in Figure 6.10, where an object is imaged twice: once by a lens with a particular focal length (say 40 mm) and once by a lens with twice that focal length (80 mm). In the second case, the distant scene appears much larger on the photograph. This fact thus explains why long focus lenses magnify distant objects to a great extent, while wide-angle lenses (indicated by a short focal length) do not. As a result, lenses with a shorter focal length enable more of a scene to be captured on the same image area.

![Figure 6.10. The effect of increased focal length on subject magnification.](image)

6.6 Field of View – Putting It All Together

Field Of View (FOV), field angle of view, angle of view, angular field of view, picture angle or angle of coverage all indicate the same thing: the angle in object space over which objects are recorded in a camera. Often quoted as a fixed lens characteristic, FOV is a result of the combined play between both the physical size of the camera’s sensor or film frame and the focal length of the lens attached. Therefore, it must always be quoted in relation to the imaging area. To explain this, imagine the cone of light rays from a scene reaching the sensor
or film after refraction by a photographic lens (Figure 6.11). As the lens elements are all circularly shaped, the picture-forming light passes circular openings and delivers a sharp image which is circular itself. This circle cast by the lens is called the **image circle**, its size not being an inherent property of the focal length of the lens, but completely determined by the optical design [649, 651].

Although a circular image is projected by the lens, different film frames and sensors all capture and store a rectangular or square image (Figure 6.11). From an optical point of view, the solid state sensor or film frame is the field stop of the optical system. Stops are devices that prevent the light rays from reaching the imaging area. They come in two forms: **aperture** and **field stops**. An aperture stop simply limits the diameter of the light beam that passes through the photographic lens. Such a limit can basically be the size of the first lens element or a hole deliberately incorporated, such as the iris diaphragm. Apart from this, a field stop exists, being the stop that confines which object points are imaged [292, 433, 475, 490, 554, 651]. The edge of the detector (film or digital sensor) limits which part of the image circle is used. Essentially, four curved areas at the edges of the detector are cropped out of the whole projected image.

![Figure 6.11. Digitally cropping the image circle.](image)

When it comes to the standard 35 mm film, only a rectangular zone of 24 mm height and 36 mm width is captured. Using exactly the same lens on a Nikon DX sensor, even a smaller portion of the whole image is stored: only a rectangle of 23.7 mm x 15.7 mm. This indicates a very important photographic issue: lenses are designed to project an image circle, but one must be sure that the corresponding diameter of the image circle at least equals the diameter...
Cropping the Aerial View

of the light-sensitive area. In other words, the image circle created by the lens must be larger than the circle which just encompasses the film frame/sensor. In the case of the 35 mm frame, a lens must project an image circle of at least 43.27 mm in diameter. If not, completely black corners will be the result. In practice, the light capturing area will be somewhat smaller than the image circle to assure reasonable image illumination, as the periphery of the image circle will always suffer from reduced illuminance. This phenomenon is called **vignetting** and results in a radial decrease of image illumination, yielding corners that are less bright than the centre of the image [292, 433, 490, 651]. Buying a lens that is designed for (‘optimized for’) 35 mm film SLRs and using it on a medium format camera will not work. Even if the lens could physically be connected by a lens-to-camera adapter, the image circle would not be large enough to adequately cover this larger imaging area due to the fact that the lens manufacturer designed the lens with only an image circle large enough to cover the 35 mm film format. Without completely tackling the issue whether digital cameras should best be fitted with specifically designed ‘digital lenses’ or not, it is important to note that all 35 mm lenses can be used on today’s small format D-SLRs as far as the image circle is concerned.

![Figure 6.12. Diagonal, horizontal, and vertical FOV.](image)

This explains why FOV is a combination of both focal length and imaging area. Although the lens of Figure 6.11 is in both cases the same, the amount of the scene that is recorded depends on the imaging area. The tinier the sensor is, the smaller is the amount of the scene that will be covered. As seen before, changing the focal length will also change the captured portion of the scene. It is common practice with optical instruments to state the complete FOV in terms of
the angle $\alpha$, expressed in degrees ($^\circ$). This value represents the angle in object space over which the scene is imaged. The semi-field or half-field angle $\theta$ is defined as the angle made by the optical axis with the extreme light rays falling on the sensor (i.e. $\alpha/2$) [433, 651]. The complete field of view can be measured by three parameters: horizontal, vertical and diagonal FOV (see Figure 6.12).

The formula to mathematically calculate FOV depends on the type of lens used. In essence, two types exist for photographic use: rectilinear and fisheye lenses [651]. The latter are less common, displaying straight lines as curves (except for those running through the centre of the frame) and sometimes able to project the scene as a circle within the imaging area (an example from such a circular fisheye is nicely illustrated by Figure 6.13 which was generated with a 6 mm fisheye lens on a D-SLR).

Figure 6.13. A fisheye image (photograph by D. Slater [722]).

Fisheyes are not further considered here. The more common rectilinear lenses render straight lines in the scene as straight lines in the image (although these lenses mostly produce nearly rectilinear images due to several distortions) [292, 651]. The FOV $\alpha$ of such a rectilinear lens can be calculated using simple trigonometry, the complete formula – and how to derive it – being outlined in Figure 6.14.
The image distance $s'$ is important in the formula. More specifically, the closer an object is, the larger the image distance and the smaller (although slightly) the FOV will be. The three examples in Table 6.1 show that the object-at-infinity-formula gives a good indication of FOV in most circumstances, making the more extensive formula only of real importance in macro photography (because the value of $m$ is significant in that area).
Table 6.1. Diagonal FOV for three types of photographs.

<table>
<thead>
<tr>
<th>Type of photograph</th>
<th>s (m)</th>
<th>f' (mm)</th>
<th>s' (mm)</th>
<th>m</th>
<th>FOV (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object at infinity</td>
<td>∞</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>24.42</td>
</tr>
<tr>
<td>135 format</td>
<td></td>
<td>100</td>
<td>100/∞</td>
<td></td>
<td>2*arctan (43.27/200)</td>
</tr>
<tr>
<td>Aerial photograph</td>
<td>250</td>
<td>100</td>
<td>100.04</td>
<td>0.0004</td>
<td>24.41</td>
</tr>
<tr>
<td>135 format</td>
<td>111.11</td>
<td>111.11/1000</td>
<td>43.27/200</td>
<td>2<em>arctan [43.27/(200</em>(0.0004+1))]</td>
<td></td>
</tr>
<tr>
<td>Macro photograph</td>
<td>1</td>
<td>100</td>
<td>111.11/1000</td>
<td>22.04</td>
<td>2<em>arctan [43.27/(200</em>(0.111+1))]</td>
</tr>
</tbody>
</table>

Certainly in the field of aerial photography, the calculation of FOV for lenses and imaging formats can therefore be executed by using the simpler equation for objects at infinity. This formula will be used here to calculate some values and bring a few issues to the notice. As mentioned before, three kinds of FOV are possible. That is why the frame size is incorporated into the formula: the height of the frame or sensor allows the vertical FOV to be calculated, whereas the width is needed for the horizontal FOV and the diagonal for the diagonal FOV (see again Figure 6.12).

As an example, consider a 50 mm lens on a 35 mm SLR:
- the horizontal FOV equals $2 \times \arctan \left( \frac{36}{100} \right) = 39° 35' 52''$;
- the vertical FOV equals $2 \times \arctan \left( \frac{24}{100} \right) = 26° 59' 29''$;
- the diagonal FOV equals $2 \times \arctan \left( \frac{43.27}{100} \right) = 46° 47' 33''$.

When the same lens is mounted on a Nikon DX sensor, the values become:
- horizontal FOV = $2 \times \arctan \left( \frac{23.7}{100} \right) = 26° 39' 58''$;
- vertical FOV = $2 \times \arctan \left( \frac{15.7}{100} \right) = 17° 50' 43''$;
- diagonal FOV = $2 \times \arctan \left( \frac{28.43}{100} \right) = 31° 44' 27''$.

### 6.7 Crop Factor

#### 6.7.1 Equivalent FOV

From the last numeric example and Figure 6.11, it is clear that FOV changes with the imaging format. Due to the popularity of the format, most FOVs associated with a particular focal length are based on 35 mm film photography. If one uses such a lens on a digital small format camera, a so-called focal length factor, focal length multiplier, digital multiplier, digital magnification factor or digital crop factor needs to be applied to calculate the 35 mm equivalent focal length and FOV.

Consider an analogue 135 format lens of 24 mm fitted onto a Nikon specific DX format sensor (23.7 mm x 15.7 mm). The image yielded by the camera will look more ‘tele’ (i.e. the object is apparently brought closer), because a smaller portion of the image circle is used (see also Figure 6.11). To find the 35 mm equivalent focal length which would produce exactly the same
image on a 35 mm film frame, one needs to use a multiplier which can be calculated for the horizontal, vertical or diagonal dimension:

- horizontal: 36.00 mm / 23.7 mm = 1.519;
- vertical: 24.00 mm / 15.7 mm = 1.529;
- diagonal: 43.267 mm / 28.429 mm = 1.522.

Any of these three factors can be used to multiply any given lens focal length used on the DX sensor to obtain the 35 mm equivalent lens:

- horizontally: 24 mm * 1.519 = 36.46 mm;
- vertically: 24 mm * 1.529 = 36.70 mm;
- diagonally: 24 mm * 1.522 = 36.53 mm.

So, using the mentioned lens with a focal length of 24 mm on a digital still camera, the same image will be captured as a 35 mm film frame-36.5 mm lens combination would do. Therefore, 36.5 mm is denoted the equivalent focal length for the 135 format. Although there are always three possible outcomes, the camera manufacturers mostly give one value as a conversion factor: the one based on the diagonal of the image plane, to tackle the issue of different aspect ratios. For this camera, Nikon gives 1.5 as a multiplication factor.

Besides focal length, also an equivalent FOV exists. From this particular example, it must be obvious that the FOV set by a focal length of 24 mm and across the diagonal of the DX sensor equals the FOV across the diagonal of the 35 mm film plane with a lens of focal length 36.5 mm. Mathematically this can easily be confirmed:

- diagonal FOV of a 36.5 mm lens on the 135 format = 61.27°;
- diagonal FOV of a 24 mm lens on a DX sensor = 61.27°.

Although one will not commonly read this, it is only possible to accurately divide the FOV by the crop factor to get the result for the new FOV if the focal length is very large. For example: the diagonal FOV for a 50 mm lens on a 35 mm frame is 46.79°, while the diagonal FOV for the same lens on a DX sensor equals 31.74°. Dividing both outcomes gives 1.47 and not the expected 1.52. However, the diagonal FOV for a 500 mm lens on a 35 mm frame is 4.96° while the diagonal FOV for the same lens on a DX sensor is 3.257°. Dividing both results yields 1.52, which is very close to the crop factor of 1.52. Considered mathematically, the FOV of a certain focal length \( f \) on a 35 mm frame must equal 1.522 times the FOV of the same lens on a DX sensor. Otherwise stated: \( 2 \times \arctan \left( \frac{43.27}{2f} \right) = 3.044 \times \arctan \left( \frac{28.43}{2f} \right) \). Thus \( \arctan \left( \frac{21.63}{f} \right) \) needs to equal 1.522 * \( \arctan \left( \frac{14.21}{f} \right) \), which only is possible if \( f = \infty \).

### 6.7.2 Correct Terminology

Equivalent FOV and focal length can be calculated for every lens-sensor combination whenever the sensor size is smaller (or larger) than the 135 format frame. This gives cameras with a full frame sensor the advantage of not having to take this ‘changed’ FOV in mind. It is,
however, very important to understand that the actual focal length of the lens remains unaltered, no matter how big or how tiny the sensor's dimensions are. Mounting a 50 mm lens on a digital camera with a Four Thirds sensor does not deliver a lens with another focal length, neither does the lens behave differently, as can often be read. 50 mm stays 50 mm. In this respect, terms as 'focal length multiplier', 'focal length factor', and 'digital multiplier' are misnomers. Also the expression 'digital magnification factor' can be interpreted erroneously. Even though the circular image field can be larger than the field stop (i.e. sensor or film frame), circumscribe it or be wholly within it, reducing or enlarging that field stop just alters the area recorded and does not change the magnification of the object.

Therefore, all lenses of a particular focal length produce the same image magnification at the plane of focus when focused at the same distances. However, cropping the imaging area consequently results in a narrow FOV, making the terms digital crop factor or FOV crop factor the only valid ones. As a rule of thumb: the smaller the imaging area, the shorter the focal length lens is needed to yield the same FOV as a lens with a longer focal length on the 135 format. Figure 6.3 also shows the crop factor exercised by the image sensor. Common crop values range from slightly larger than one to two for D-SLRs, whereas compacts crop the FOV by at least a factor four.

6.8 Wide-Angle, Normal, and Long Focus Lenses

"Moderate wide-angle lenses are characterized by their short focal length which lies in the 24 mm to 35 mm range". These and similar statements hold true, but only for the 135 format. Wide-angles for smaller formats will have shorter focal lengths for example. Using FOV on the other hand, a meaningful and unambiguous method is given to classify all lenses into the three main categories of wide-angle, standard, and long focus lenses.

6.8.1 Standard Lens

A wide-angle lens must provide a wide FOV (e.g. 90°), while long focus lenses have to deliver a narrow FOV (say 7°). A standard, ideal or normal lens on the other hand should give a normal FOV with a correct perspective rendition in the photograph. Therefore, the human visual system must be considered. As the static human eye has a sharp visual FOV of some 50°, this angle can be taken to calculate the focal length of a standard lens which creates a corresponding angle on the 135 format. Mathematically, this becomes: 50° = 2 * (arctan (43.27 mm / 2X)) or 50°/2 = arctan (43.27 mm / 2 * X). This gives tan 25° = 43.27 mm / (2 * X) or X = 43.27 mm / (2 * (tan 25°)) = 46.40 mm.

Because this value of 46.40 mm is close enough to the diagonal of the 35 mm frame, the diagonal of the imaging format was suggested to be a good way of defining a standard focal
length in general [649]. Applied to the 35 mm frame with a diagonal of 43.27 mm, the corresponding focal length yields a theoretical diagonal FOV of circa 53°. As FOV changes with imaging area, the value of a standard lens in case of a 230 x 230 mm roll film frame (the often utilized standard in analogue vertical survey photography) is 325.27 mm, as this is the value of the frame diagonal. It is obvious that these values are out of the ordinary. Therefore, all focal lengths that fall within a certain range are called normal lens. For the 35 mm format, focal lengths can vary from about 35 to 58 mm, with 50 mm the most common value found for a standard lens (a standard which dates back to 1925 when it was introduced on the Leica camera [651]). The idea is that they all come close to the theoretical angle of 53°. Once the focal length is significantly smaller or larger than the diagonal of the format in use, one deals with wide-angle and long focus lenses respectively.

6.8.2 Wide-Angle Lens

For wide-angle lenses (i.e. a diagonal FOV larger than 63°), a division is sometimes made between wide-angle (WA; 63° to 84°), very wide-angle (VWA; 84° to 100°) and extreme wide-angle (EWA; 100° to 120°). 120° to 140° is about the possible limit for rectilinear lenses (due to the diminishing peripheral illumination of the image circle), therefore also called the fisheye limit. Fisheye lenses no longer use the rectilinear central projection, but apply equidistant, orthographic or other projections to enable a hemispherical (i.e. 180°) or larger subject field (i.e. hyper-hemispherical) to be recorded.

<table>
<thead>
<tr>
<th>type of lens</th>
<th>appr. field of view (° on diagonal)</th>
<th>appr. ( f ) (in mm) for the 135 format</th>
<th>appr. ( f ) (in mm) for the DX sensor</th>
<th>appr. ( f ) (in mm) for the 4/3° format</th>
</tr>
</thead>
<tbody>
<tr>
<td>extreme wide-angle</td>
<td>120</td>
<td>12</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>very wide-angle</td>
<td>100</td>
<td>18</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td>wide-angle</td>
<td>84</td>
<td>24</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>standard</td>
<td>63</td>
<td>35</td>
<td>23</td>
<td>18</td>
</tr>
<tr>
<td>medium long focus</td>
<td>41</td>
<td>58</td>
<td>38</td>
<td>30</td>
</tr>
<tr>
<td>long focus</td>
<td>18</td>
<td>135</td>
<td>90</td>
<td>70</td>
</tr>
<tr>
<td>very long focus</td>
<td>8</td>
<td>300</td>
<td>200</td>
<td>155</td>
</tr>
<tr>
<td>extreme long focus</td>
<td>5</td>
<td>500</td>
<td>330</td>
<td>260</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2000</td>
<td>1300</td>
<td>1050</td>
</tr>
</tbody>
</table>

Figure 6.15. Classification of lenses which apply a rectilinear projection (data from [649, 651, 812]).
6.8.3 Long Focus Lenses

On the other part of the lens spectrum long focus lenses can be found, sometimes subdivided in medium long focus lenses (MLF; 18° to 41°); long focus lenses (LF; 8° to 18°) with focal lengths up to six times the format-dependent standard focal length, very long focus (VLF; 5° to 18°) and extreme long focus lenses (ELF; 1° to 5°). Often, long focus lenses are erroneously called telephotos. Instead, telephoto refers to a specific type of lens construction. The telephoto design is indeed used in many long focus lenses, but the latter can also be built by other configurations [649, 651]. Figure 6.15 gives a classification of lenses and their matching diagonal FOV. As the corresponding focal lengths of the 135 format are so well known, their approximate values are given in the third column, followed by the focal lengths that deliver an equivalent FOV on smaller digital sensors. As one can see, a standard lens has a diagonal FOV between 63° and 41°, which results in a 35 mm to 58 mm on the 135 format. The same FOV is given by a lens with focal lengths in the range of 18 mm to 30 mm on a Four Thirds sensor. Once more, full frame sensors do not suffer from crop factors, making the calculation of an equivalent FOV unnecessary.

![Figure 6.15](image)

Figure 6.15. Classification of lenses and their matching diagonal FOV.

6.9 Ground Coverage

Finally, it is appropriate to end the chapter the same way it started: with the World Wide Web. It needs to be mentioned that the WWW is not all sorrow and misery (which might have been the impression one got from the introduction). This is proven by f/calc, a nice free tool which can be downloaded from tangentsoft.net/fcalc/ and allows to calculate FOV along with other photographic quantities as Depth Of Field (DOF), hyperfocal distance, etc. (see Figure 6.16). To

![Figure 6.16](image)

Figure 6.16. Using FOV in the calculation of field coverage.
calculate FOV, one can choose a film format and enter a focal length, after which the programme instantaneously displays the resulting horizontal, vertical and diagonal FOV for an object at infinity. Beware that the creators made a distinction between FOV and angle of view, the tab “Field of View” comprising a means to calculate the effective width and height of the real-world scene being imaged. As in this chapter, both FOV and angle of view are mostly considered to be synonyms. What the programme calculates is usually denoted field coverage [651] (or ground coverage in the context of aerial remote sensing [482]). Figure 6.16 shows how to calculate field coverage when the FOV is already known. Moreover, an easy formula to determine the ground coverage of an aerial image is provided. From the drawing, it is obvious that both FOV and field/ground coverage are heavily related. Luckily, this is at least one firmly established fact which does not need to be reconsidered with the advent of the digital (r)evolution.
CHAPTER 7

It's All about the Format

Unleashing the Power of RAW Aerial Photography

The content of this chapter is accepted for International Journal of Remote Sensing [SCIE – IF 2007: 0.987]

Abstract

Current one-shot, hand-held Digital Still Cameras (DSCs) generally offer different file formats to save the captured frames: JPEG, RAW, and/or TIFF. Although JPEG is the most commonly used file format worldwide, it is incapable of storing all original data, something that goes to a certain extent also for big TIFF files. Therefore, most professional photographers prefer shooting RAW, often described as the digital photography’s equivalent of a film negative. As a RAW file contains the absolute maximum amount of information and original data generated by the sensor, it is the only scientifically justifiable file format. Besides, its tremendous flexibility in both processing and post-processing makes it beneficial from a workflow and image quality point of view. Large file sizes, the required software, and proprietary file formats remain, however, hurdles often too difficult to take for a lot of photographers.

Aerial photographers who shoot with hand-held DSCs should be familiar with both RAW and the other file formats, as their implications cannot be neglected. By outlining the complete process from photon capture to the generation of pixel values (additionally illustrated by real-world examples, the advantages and particularities of RAW aerial photography should become clear).
7.1 Introduction

Since the advent of the first truly affordable small format Digital Single-Lens Reflex (D-SLR) Canon EOS 300D/Digital Rebel in August 2003, a worldwide increasing number of (aerial) photographers have been seeing the light and converted to the digital approach of 35 mm photography, applying one-shot Digital Still Cameras (DSCs) in a rich variety of photographic solutions [619]. Though this process is not often without striking a blow, yet it is save to state that the majority of digital shooters (both small and medium format) already know more about enhancing a digital image in photo editing software than they have ever known about darkroom techniques. The direct approach (there is no ‘preview’ button in the darkroom), the ability to work in daylight with ‘clean’ computers instead of juggling with toxic products in darkrooms, and the relative easiness as well as low cost are only some of the advantages digital image acquisition and (post-)processing benefit from. Nevertheless, both film and digital photography perfectionists are trying to accomplish the same thing: getting the maximum out of their initially acquired data. The latter is what this chapter is all about: the originally captured or RAW information. Besides being beneficial for the image quality, it will be shown that RAW is the only format (remote sensing) scientists should use in their research, as it offers quantitative and qualitative possibilities in-camera generated JPEGs and TIFFs do not.

7.2 RAW – A Definition

In accordance with common terminology from the digital world, one would think RAW is an acronym. However, the word is an exception and signifies just what it sounds like: raw data. Although most texts describe RAW as the unprocessed data from the sensor of the DSC (e.g. [36]), it is more accurate to consider a RAW file as the analogue sensor information which has been amplified and converted to digital data, without being subjected to any major processing by the camera’s embedded software (i.e. firmware). Because this RAW file holds all data with only a minimal change compared to the data coming from the camera’s digital sensor, RAW can be seen as the digital negative: it will never degrade and allows an infinite number of digital prints (as JPEG or TIFF files) to be made in the future. It might even be better to consider it the equivalent of the latent image, as a RAW file holds all captured information without any digital development done afterwards.

To completely understand the nature of digital RAW capture and master its full (remote sensing) potential, it is best to delve a bit deeper into the process of actual image capture, which completely takes place inside the one-shot DSC, the latter to be defined as a photo camera equipped with both a digital image sensor for capturing full photographic data in one
exposure as well as a storage device for digitally saving the obtained image signals [760]. Anno 2009, all small and medium format D-SLRs offer the possibility to shoot RAW, while even most hybrid and some compact models can (the latter two being known as the so-called consumer DSCs).

7.3 RAW – The Creation

7.3.1 Photodiodes

Whether a digital sensor is one of the CCD, CMOS, or JFET type, all image sensors of today’s one-shot DSCs are silicon chips containing a two-dimensional array of photosites in order to produce the final image (Figure 7.1). Each of those photosites contains a light-sensitive area made of silicon, a photosensitive detector or photodiode [395, 569, 752, 843]. When the number of effective pixels in a DSC is for example said to be 2560 x 1920, the camera’s sensor has at least 2560 by 1920 photodiodes. As typically one photosensitive element of the array contributes one pixel to the final image, the result is an image with 2560 x 1920 or 4.9 million pixels (also denoted as MegaPixels or MP).

Figure 7.1. The layout and working of a Bayer photodiode array.

From the moment the exposure begins, these photodiodes will start to collect photons (Figure 7.1) that are gathered by the lens. After the exposure, each diode contains a certain number of these photons just as buckets would contain a certain quantity of raindrops after a rainstorm [421]. By collecting these photons, DSCs sample ElectroMagnetic (EM) radiation both in a
It's All about the Format

spatial (the location of the photodiodes) and tonal way (the amount of photons captured) as well as in time (the exposure time).

However, photodiodes simply count photons; they are monochrome devices, unable to tell the difference between different wavelengths. Consequently, such a construction would only be able to create greyscale photographs without further adaptations. Therefore, the most widespread method to give colour sensitivity to a one-shot DSC image sensor is the use of a Colour Filter Array (CFA). This mosaic pattern of coloured filters is positioned on top of the sensor, allowing only particular spectral components of the incident EM radiation to be collected (Figure 7.1) [395, 569].

Almost all DSCs use a three-colour Red-Green-Blue (RGB) pattern in which the coloured filters are arranged as shown in Figure 7.1. This arrangement, called a Bayer pattern, typically features a repeating group of four photodiodes, in which two have Green filters – to mimic the higher sensitivity to Green light of the Human Visual System (HVS) and enlarge the perceived sharpness of the digitally recorded scene [404, 601] – while the remainder are either Red or Blue [62]. Notwithstanding this Bayer arrangement is almost constantly used in digital photography, other RGB patterns exist as well, as do designs with the three complementary colours Cyan-Magenta-Yellow [260, 395, 495] along with four-colour systems like the RGBE (E indicating Emerald) CFA introduced by Sony (Figure 7.2).

A completely different approach was patented by Foveon Inc. With their innovative X3 direct image sensor, this privately held corporation created in 2002 a particular kind of three-layered CMOS image sensor, enabling the capture of all three Red, Green, and Blue wavebands at the same location (Figure 7.9) by exploiting the wavelength-dependent absorption of silicon [297, 502]. As this unique design is only implemented into Sigma’s D-SLR SD9, SD10, and SD14 as well as into Polaroid’s x530 point-and-shoot camera [297], the remaining of this chapter will focus on the abundant Bayer CFA approach – unless otherwise indicated.

Figure 7.2. Different CFAs.

7.3.2 From Analogue to Digital

In spite of the different CFA designs, every single photodiode will capture only one spectral band (one colour component in the case of visible imaging), which is stored by a digital
intensity value that is proportional to that particular incident EM radiation. As an example, consider a Red-filtered photosite. Only the Red part of the incoming light will pass through the filter, subsequently creating an according charge in the silicon photodiode due to the photo-electric effect [420]. This photo-electric effect, explained by Albert Einstein (1879-1955) in 1905, makes the silicon release electrons when exposed to EM radiation [806], the latter process obeying a linear relationship [256, 395, 420, 752, 843]. Even though only a fractional number of incident photons – denoted by the term Quantum Efficiency (QE) – will effectively be converted by the photo-electric effect, more photons will always generate more free electrons. Coming back to the bucket analogy, it’s this electrical charge that gets trapped and collected in a potential well as long as the integration time lasts [421]. As soon as the shutter closes, this electrical charge is shifted to the output sense node (Figure 7.3) and converted to a voltage [260, 395, 420]. Afterwards, these small analogue voltages are amplified by the read-out amplifier with a certain gain to correspond to the specific ISO value set [453, 816]. It goes without saying that a higher ISO setting (e.g. ISO 3200) needs more amplification than a lower value (e.g. ISO 200).

Finally, once the real-world signal is sampled by the diodes and captured in the form of voltages, it must be quantized to Digital/Data Numbers (DNs) or Analogue-to-Digital Units (ADUs) by the Analogue-to-Digital converter (A/D converter or ADC). The ADC therefore classifies the total possible range of continuously varying analogue voltages into a finite number of levels/gradations, subsequently assigning a DN to each level. The total range of different tones or quantization values an ADC can create is termed tonal range [93], and completely determined by its sample/bit depth: quantization with $N$-bits rounds all possible voltage levels to these $2^N$ values [344]. As each additional bit results in a doubling of the number magnitude, more bits used in quantization means that more shades can be encoded, which in turn leads to a smoother transition between each tone (Figure 7.4). A typical

![CCD Sensor and CMOS Sensor Diagram](image.png)
consumer DSC therefore uses eight bits/one byte, hereby allowing $2^8$ or 256 distinct values. Most high-end D-SLR cameras utilize 12-bit, 14-bit or 16-bit ADCs [11, 811], yielding a wide tonal range of $2^{12}$ (i.e. 4096), $2^{14}$ (i.e. 16384) or $2^{16}$ (i.e. 65536) gradations respectively. These high bit depths are important in avoiding posterisation/banding, a phenomenon where abrupt changes between tones become apparent, often due to seriously post-processing (e.g. histogram stretching) and first discernable in regions with gradual tonal transitions as skies and clouds (Figure 7.4).

At this stage, the DN originating from a particular filtered photodiode still only refers to a greyscale radiation intensity value. Taking all diodes into account, a complete array of DNs is the generated outcome of each digital sensor, notwithstanding the sensor-related differences in this whole electron-to-DN chain [274]. Some DSCs also apply a so-called pre-processing or camera compensation step on these DNs. Even though there is no real convention in the execution of this operation (neither in the steps executed nor the algorithms used), a few simple operations might be applied: defective pixel correction (to estimate the value of the defective diode), a linearization step to counteract any non-linearity introduced by the DSC’s electronics, and some noise compensation [11, 646]. Ultimately, this minimally processed array of DNs is sent to the camera’s local buffer (Figure 7.3) together with important information about the RAW file, the so-called metadata.
7.3.3 Metadata

In addition to the DNs that encode the real-world scene, metadata are generated as well. Literally meaning ‘data about data’, these metadata describe the content, quality, condition, owner rights, and other characteristics of data. In the world of digital photography, different standards are used to store information about digital frames, the Exif (Exchangeable image file format) metadata standard probably the most commonly known one. Created by JEIDA (Japan Electronic Industry Development Association), this Exif specification provides a rigid format to record shooting data (e.g. the serial number and model of the DSC, the aperture, shutter speed, focal length, possible flash compensation, the colour space, the date and time of shooting, etc.) in mandatory, recommended, and optional tags stored in a separate segment of the file.

If the camera is GPS-enabled, tags can also hold the latitude, longitude and altitude of the geographical location the particular photo was taken in. Moreover, also new vendor-defined metadata can be added [423, 600]. These Exif defined tags are created and stored simultaneously with the DNs, making it possible to analyse them afterwards. In addition, RAW files also hold some tags to define the CFA data (e.g. the pattern used), and additional image reconstruction parameters like white balance, sharpening and noise settings [600].

7.4 RAW – File Details

Once it is created, such a RAW file can be seen as a container, holding two separate parts: the metadata stored in a separate header and a bunch of samples in X and Y direction, with every sample characterised by one DN (except in Foveon’s solution), and its location expressed in the image coordinate system. In this collection of DNs, each individual number represents one spectral value, generated by one photodiode. Consequently, this RAW image still has a greyscale character [36, 304] with embedded CFA pattern (apparent in Figure 7.5). Besides the aforementioned amplification, A/D conversion, and possible pre-processing step, no further adjustments are performed, making the amount of data processing really small. A RAW file thus truly is the most pure form of generated digital photographic data.

When RAW files of different manufacturers are compared, it becomes obvious that this file type lacks a general standard. Although this absence of a common structure is often considered to be a serious drawback, all RAW files store the original DNs and the metadata, as was outlined above.
7.5 RAW – Processing

Initially, every DSC takes RAW pictures, but whether these are directly saved for processing afterwards or instantaneously ‘developed’ by the camera to a JPEG (Joint Photographic Experts Group) or TIFF (Tag(ged) Image File Format) file depends on both the photographer and the DSC – as the latter needs to offer this option. When option two is chosen, the DSC’s firmware shall process the RAW data based on a mix of default parameters and certain user settings like sharpening, brightness, **White Balance** (WB), exposure adjustments, etc. By contrast, the first choice allows the processing of the data at a later stage and gives the photographer almost total control over the further processing of the original DNs, only requiring a computer and suited software, all in exchange for more visual quality (with the extra benefit of a data source that can be revisited and reprocessed endlessly without any quality loss) and the scientifically very important possibility to address the most pure data form generated.

In comparison with the film-based approach, a JPEG or TIFF file that is written on the memory card equals the development and enlargement of the latent image by a photo lab inside the camera. Using the RAW file and subsequently processing it on the computer comes down to performing all darkroom work yourself (although now in a digital environment), with the additional benefit to read out the initially captured values, of the utmost importance in scientific image processing and analysis.

By unfolding some of the individual processing steps the firmware performs, the opportunity is seized to compare this development procedure with the choices one has in a computer-based RAW conversion (Figure 7.6 depicts a flowchart of all individual development steps, hereby
serving as a kind of visual guideline for the whole processing chain). Real-world remote sensing examples will allow the important differences between both approaches to become clear.

7.5.1 White Balance

Because the channel specific DNs are generally unequal when photographing a spectrally flat object (white, black or grey), the values in each channel must be multiplied by a certain scaling factor to yield the expected identical channel numbers and tackle the unequal spectral
It's All about the Format

responses [326, 461, 741, 741]. This is visually shown in Figure 7.7. The RAW 12-bit values are yielded by photographing a grey, spectrally neutral WhiBal White Balance Reference Card [626], displayed on the lefthand side. To generate a perfectly neutral grey card on screen, the Red and Blue channels are normalised here to the Green channel.

However, it is also clear from the illustration that these multipliers change according to the EM source used to illuminate the object. A perfect white wall might reflect more Blue wavelengths than Red radiation when photographing on a cloudy, overcast day. On the other hand, several artificial light sources abundantly generate Red wavelengths [601]. This large variety of generated radiation attributes to each EM source a certain so-called **Correlated Colour Temperature (CCT)**: a number expressed on the Kelvin temperature scale, relating the specific spectral output of that EM source – which is perceived as a certain colour by the HVS – to the same colour perceived by heating a blackbody (i.e. an idealized dense object which absorbs all incident energy). The higher the temperature at which a blackbody is heated, the more its colour shifts to shorter wavelengths (from red to orange to bluish white) and the more intense the emitted light is [806]. A good example to illustrate this is heated iron. In a first stage, there will be a deep red glow. By raising the temperature, the iron shall radiate brighter, reddish-orange light. Increasing the temperature even more yields a brilliant blue-white light. Otherwise said: the emitted spectral radiation of a blackbody is only a function of its absolute surface temperature (as described by Planck’s law [806]), hence the term **Colour Temperature (CT)**.

As a blackbody is an idealized object and most EM sources are far from ideal blackbody radiators – apart from the sun (ca. 5800 °K), halogen tungsten lamps (ca. 3200 °K), and tungsten filament lamps (ca. 2850 °K) – these sources can not be described solely as a function of their temperature. This led to the concept of CCT: the blackbody temperature that yields the same chromacity experience as the EM source under consideration [103, 305].

![Figure 7.7. The channel specific DNs and calculated normalised multipliers retrieved from a WhiBal White Balance Reference Card photographed under different illumination conditions.](image)

Human eyes constantly adjust to such CCT changes and will therefore be able to tell a wall is white, irrespective of the illumination conditions [326, 402, 484]. Digital sensors and film are unable to do so. In the analogue era, one had to change the type of film and/or use
appropriate filters to avoid colour casts. In digital photography, the DSC only needs to know the wall is supposed to be white so it can accordingly calculate the correct multipliers [461]. This is also explained in Figure 7.7. Within the WhiBal, four differently coloured patches are displayed, showing the WhiBal’s spectrally flat grey surface, but photographed under different illumination conditions (flash light, incandescent bulb, cloudy and open sky) and without any WB applied. By reading out the particular Red, Green, and Blue values of these reference pictures, the different channel multipliers were calculated by normalising everything to the Green channel. Both these patches and multipliers (which are also graphically displayed on the right) obviously demonstrate incandescent light to emit much more Red wavelengths than the other sources, indicated by the much lower Red multiplier and the orange-yellowish colour cast of the patch. A cloudy sky on the other hand creates a bluish cast, indicated by the high DN in the Blue channel (relative to the Red channel) and the low Blue multiplier.

At the time of capture, the WB setting is determined using the DSC’s automatic or manual WB setting and stored in the metadata. It has no effect on the generated DNs until the specific normalisation values are effectively applied in the final calculation of the complete pixel values. In the case the RAW file is processed by the firmware, the multipliers are used to recalculate the initially generated DNs of all channels, hence making sure the spectrally neutral zones – and by extension also all the other colours in the digital image – appear without serious colour casts, irrespective of illumination condition. In case of the WhiBal, all three channels will ultimately have almost identical DNs for the spectrally flat surface [11], yielding a picture of the grey card that looks very neutral to the HVS. Therefore, this card – and similar utilities – can be of great benefit to correctly calculate the WB, as even the automatic WB determination of professional DSCs can be fooled to a certain extent. In such cases of incorrect WB, the in-camera processed JPEG or TIFF will have a colour cast and although this can be mostly dealt with in post-processing, the quality of the picture will degrade to a certain extent and the cast is sometimes difficult to remove completely.

When storing RAW files, the WB can be set after the frame has been taken. Because this information is only stored as metadata, normal RAW conversion software reads this tag and applies it to the image when opening it. The user can always override this setting by applying other correction values, using a dedicated WB tool as the WhiBal (and others) for accurate determination or arbitrarily choose values (in which case the captured WB serves as a reference point to adjust the colours further). No matter which approach is used, the WB is altered without altering the original RAW file or destroying any initially captured information. RAW therefore is an ideal solution when the photographer is not completely sure about the CCT of the light (as in aerial photography) and/or maximum possibilities in post-processing have to be maintained. Additionally, utilizing software more suited for scientific purposes (e.g. The MathWorks’ MATLAB, dcrw, IRIS) the image can be processed without any WB being applied, hence yielding the originally captured DNs (which in fact boils down to multiplying every channel with a factor of 1.0). This is extremely important when the spectral response of a DSC needs to be determined (see [548] or Chapter 10) or in case non-visual imaging is performed. Chapters 8, 9, 10, and [776] describe the use of a modified D-SLR, capable of taking pure Near-
InfraRed (NIR) aerial photographs. As there is no need to get true colour in this invisible range, white balancing can be omitted, generating files that clearly show the different spectral response of every channel. In Figure 7.8A, an NIR NEF (Nikon Electronic Format, Nikon’s proprietary RAW format) aerial photograph was linearly processed (see 7.5.3) to a white balanced 16-bit TIFF using dcraw, a free RAW decoder [184]. Figure 7.8B shows the same image without any WB applied. The included histogram clearly indicates the different response of the three filter sets, showing that the Red filters are most transparent to NIR (hence the image’s accordingly dominant colour). The different response in the Red and Blue channels of Figure 7.8A and Figure 7.8B respectively is again shown in the lower part of Figure 7.8. These results show that omitting WB can be very important in case only one or two spectral channels are needed or absolute spectral intensity measures have to be taken.

Figure 7.8. (A) A linearly processed NIR NEF file with custom channel multipliers; (B) the same image without any WB. The lower part shows the Red and Blue channel of Figure A and B respectively.

### 7.5.2 Demosaicing

Apart from the unique, three layered Foveon X3 sensor, most DSCs are single-shot models using one CCD, CMOS or JFET image sensor where each photosite contains its own spectrally selective coloured filter. Because each photodiode senses only one spectral component, an algorithm is needed to fill in the corresponding DN for the other two bands (Figure 7.9). To end up with a valuable three-channel image, the incomplete RGB values of the RAW file need to undergo a double operation.

- the greyscale DN – which corresponds to the intensity of a certain waveband – has to be converted to a matching colour value;
- the other two primary colours must be approximated to achieve a complete RGB image.
Both operations are done simultaneously in a process called **demosaic**ing, CFA interpolation, colour reconstruction or de-Bayering (in case a Bayer array is used). To accomplish this process, a very important piece of information is read in the beginning. This information, in fact another piece of metadata included in the RAW file, is called the decoder ring [303]. It is crucial in the processing of a RAW file since it stores the arrangement of the CFA, enabling the link between each intensity value and one of the three primary colours [600]. Once the software is aware of this, the missing spectral components can be filled in.

![Diagram](image.png)

**Figure 7.9.** Difference in diode-specific spectral information acquired by the Foveon sensor and the CFA-solution (adapted from [297]).

As a general guideline, a better reconstruction can be obtained if more actual spectral measurements are taken into account to estimate the specific missing channels [681], although the situation is slightly more complex. To tackle the various artefacts that can be introduced during demosaicing, an immense range of algorithms (linear and nonlinear), all varying in complexity and sometimes specified for particular CFA patterns, have already been proposed in recent years (e.g. [12, 30, 111, 166, 173, 221, 355, 477, 492, 496, 523, 557, 599, 645]).

Most DSCs apply a Bayer CFA with a ratio of 2:1:1 among Green, Red, and Blue filters. Besides the abundance of information in the Green channel, important correlations between Red,
Green, and Blue DNs exist \([355, 440, 476, 492, 842]\). While old, well-known linear techniques as nearest-neighbour, bilinear, and bicubic interpolation do not use these characteristics (hence yielding rather bad to mediocre performance by blurring fine detail and producing artefacts around edges), more sophisticated, adaptive techniques do exploit the diode’s spatial and/or spectral correlations. These methods use algorithms that completely interpolate the Green channel before tackling the remaining Red and Blue spectral components, apply edge-directed interpolations as well as pattern recognition, perform pattern matching and even combinations of those techniques, all in complexity and computationally varying methods.

However, less straightforward demosaicing requires more processing power to create a full colour image in an acceptable time span, power that often can not be delivered by DSCs due to practical reasons: the processor must be small, light, and may not ask too much of the batteries [811]. Therefore, the best interpolations can only be implemented in dedicated (proprietary) RAW converters run on modern computers, while DSCs (certainly compacts) apply quality-compromising algorithms.

To compare demosaicing quality, a minimally compressed, in-camera generated JPEG file is compared to a post-processed JPEG, generated out of the simultaneously stored NEF file.
By using Capture NX as a proprietary RAW converter, all parameters were kept equal, so to only compare the difference between demosaicing methods. It is obvious from enlarged portions of the in-camera generated JPEG that fewer details are present in both the structure of the leaf and the words ‘DIGITAL’ and ‘Start’. Although this might not be a serious issue in case of illustrative pictures, identifying small objects on aerial imagery certainly benefits from a RAW workflow.

Besides the additional advantage to work with the uninterpolated data – which may be beneficial if only true spectral measurements are to be made – more technically oriented software packages might allow to choose/implement a specific demosaicing algorithm according to the data source. Consider aerial imaging in the invisible wavebands again. All the existing CFA interpolations are applicable in the visual domain, but modification might be expected to obtain a complete, reconstructed or demosaiced UltraViolet (UV) or NIR image with a minimum of various artefacts. As an example, Miao et al. [534] concluded that spectral correlations based on colour ratio [440] or colour difference [492] are less in multi-spectral images compared to colour images. To solve this problem, they proposed the BTES approach (Binary Tree-based Edge-Sensing method), a generic demosaicing technique that establishes the interpolation order of different spectral bands after which it determines the interpolation order of pixel locations within each spectral band, hereby using a binary tree-based scheme and edge-sensing [534]. A quantitative comparison to assess reconstruction accuracy (using Root Mean Square Error – RMSE) of mosaiced images created by their BTES method with the algorithm proposed by Lu and Tan [492] showed the latter to perform better when only three channels (one visible and two IR) were taken into account. As a result, the Adaptive Homogeneity-Directed demosaicing algorithm (AHD [389]) was chosen to demosaic the author’s NIR aerial imagery. Although BTES has not yet been directly compared to AHD, the latter algorithm yields less artefacts when compared to the Lu and Tan method, making it even more suited for NIR interpolation.

AHD demosaicing selects the direction of interpolation in order to minimize artefacts by applying a homogeneity map. Tested against even more recent methods, this nonlinear iterative procedure still has to be considered an extremely well performing algorithm, very good in reducing noise, hereby yielding sharp edges [173, 221, 356, 477, 523]. A possible drawback is the unidirectional interpolation (only in horizontal or vertical directions), sometimes yielding artefacts in high frequency components [167].

Given this, AHD was tested and proved to be the best in a range of possible algorithms for digital NIR imagery (Figure 7.11). Using dcraw, demosaiced images where used in arithmetic channel operations in an attempt to calculate a Vegetation Index (VI) with the response of only one modified D-SLR (see Chapter 10). Hence, a faithful calculation of the missing channel values was of the utmost importance and would be impossible without a RAW-based photography workflow.
7.5.3 Tonal Curve

When it comes down to the response to EM radiation, there is a big difference between shooting film and shooting digitally: film tries to mimic the light response of the HVS, while digital image sensors do not. The HVS, just like all sensations, functions in a non-linear way [740, 739]. This can be illustrated by many examples: two light bulbs (stimulus) do not make the room seem twice as bright (sensation); moreover, it is easy to discriminate a 30 W light source from one of 60 W, but almost impossible to perceive the difference between a 500 W and 530 W lamp, although both differ by the same amount. Ernst Heinrich Weber (1795-1878) was the first who tried to describe the relationship between physical magnitudes of stimuli and their perception [839].

Some thirty years after Weber, the German psychologist Gustav Theodor Fechner (1801-1887) elaborated on Weber’s law and came up with a logarithmic relationship to relate the magnitude of perceived sensation to the intensity of the stimulus [288, 739]. His law, which is known as Fechner’s law – sometimes also denoted as Weber-Fechner’s law – states mathematically that:

\[ S = k \log(I) \]

with \( S \) being the magnitude of the perceived psychological sensation, \( I \) the physical intensity of the stimulus and \( k \) a specific sensory constant previously defined by Weber [288]. In the middle of the last century, Stanley Smith Stevens (1906-1973) showed Fechner’s law to be sometimes inaccurate. He therefore proposed a different relation between sensory magnitude and stimulus magnitude: not a logarithmic one, but one that followed a power law. According
to this power law, sensory or subjective magnitude (brightness in the case of light) grows in proportion to the physical intensity (luminance) of the stimulus raised to a power. Stevens therefore introduced a formula with one additional parameter ($\gamma$)

$$S = kl^\gamma$$

where $k$ is an arbitrary constant determining the scale units used and the power exponent gamma ($\gamma$) is a constant dependent on the sensory dimension [740, 739]. For the sensation of brightness, the exponent in Stevens’ Law (also called Stevens’ Power Law or the Psychophysical Power Law) varies between 0.33 and 0.5 according to the conditions [740, 739], therefore resulting in a concave curve (Figure 7.12). To double the brightness sensation of a light source, a considerable amount of light intensity must be expended: more exactly eight times ($2 = r^{0.33} \rightarrow I = 2^8$). Due to this perceptual compression, the HVS perceives smaller steps in dark than in light as the absolute stimulus increase in the latter must be a lot bigger to perceive a brightness variation, compared to the stimulus intensity needed in dark areas. A DSC’s image sensor lacks this build-in compression. It just functions linearly: twice the amount of captured photons produces twice the sensor response. Digital sensors are therefore said to be linear recording systems, with a gamma equal to 1:

output pixel value = $k$ (input value)$^\gamma$, with $\gamma = 1$ (see Figure 7.13).

![Figure 7.12. Stevens' Power Law illustrated for brightness.](image)

This linear response yields two unpleasant effects:

- the darker areas – to which the HVS is most sensitive – are described by just a few tones [304, 453]. As an example, consider the graph and table in Figure 7.13. The x-axis of the graph indicates the whole light intensity range a particular sensor is capable of displaying (i.e. its Dynamic Range or DR). This input is related to an 8-bit output by two
curves. The linear curve indicates the way an image sensor maps the intensity range to 256 different tones. The brightest possible value the sensor is capable of will be captured by DN 255. If this maximal amount of light (100 %) is halved (i.e. the quantity of photons divided by two), the remaining brightest value will correspond to a DN of 127 (due to the linear input-output relationship). As halving the light quantity is known in photography as a photographic stop, it is obvious that the brightest stop uses 128 or 50 % of all available tonal values, while it only corresponds to the range where human vision is least sensitive. As additional halving yields 64 levels, 25 % of all DNs is used to describe the consequent photographic stop [304, 453]. Correspondingly, the third stop is represented by 12.5 % of all levels (32 tones), etc. The higher the photographic stop, the smaller the amount of light. However, these smaller amounts of light indicate zones where the HVS becomes more sensitive; the more sensitive the human vision, the smaller the amount of available tones. In this example, the darkest shadows are captured in the eighth stop and represented by only one tonal value: black;

- the uncorrected output of a digital sensor looks very dark to the non-linear working human eye [93, 304]. Because humans perceive the first 12.5 % - 20 % of a complete intensity scale as middle grey (a value that depends upon the particular gamma value used), all remaining intensities (>80 %) are perceived as middle grey to white. In a linear converted RAW file, the lower part of this intensity range (<20 %) will be converted to very dark tones. Here, middle grey is only reached at an intensity value of 50 % (Figure 7.13). Due to the linear relationship, all pixels falling in this 0 % - 50 % intensity interval are now attributed with tonal values ranging from black to middle grey, whereas they should be perceived as black to quite bright. Consequently, a linear processed RAW image is perceived as very dim (see Figure 7.8 and Figure 7.13), additionally revealed by the histogram which is seriously skewed to the right (Figure 7.13).

![Graph showing linear and gamma corrected RAW data](image)

Figure 7.13. Gamma correction of RAW data.
To solve these issues, RAW converters will apply a **gamma curve** (mostly $\gamma = 1/2.2 = 0.45$) to redistribute all tonal values and mimic the HVS [304, 453, 453, 495]. This non-linear correction allocates more tones to the shadows and fewer to the brighter areas, yielding a more equal distribution of levels. In the end, only 69 different levels (or 27.1\%) out of the same range of 256 discernable tonal values will be attributed to the first stop while the shadow areas – stop eight – are now captured by eight levels (Figure 7.13).

However, rather than a pure gamma curve, RAW converters and DSCs apply a **gradation/tonal/tone curve** to compensate for the non-linear human perception [93, 844]. To a large extent, such a tonal curve equals a gamma curve. The difference between both can be seen in a log-log plot (Figure 7.14): the gamma curve forms a straight line while the tonal curve is more S-shaped, enlarging the overall contrast of the image [11]. As a matter of fact, this tonal curve strongly mimics the characteristic curve (D-LogE curve) known from film, hence making the digital image look as ‘normal’ as frames shot on conventional emulsions.

![Figure 7.14. Comparison between normal gamma curve and generally applied tonal curve (adapted from [93]).](image)

Although in-camera processed images are subjected to a curve determined by the manufacturer (typically contrasty S-curves), setting the optional contrast parameters of the DSC (usually with the low, medium and high as options) allows this tonal curve to be altered, generating a different output (e.g. reduction or enhancement of the midtones). This is where the advantage of RAW comes into play. RAW conversion within the DSC gives the photographer only a limited amount of control over the final tonal curve, whereas storing RAW files yields enormous possibilities concerning the redistribution of the available tones afterwards, as most RAW converters have several tonal curves embedded: one for emphasizing the shadows, another for better general contrast, etc. Additionally, each tonal curve can be
It's All about the Format

further modified to a great extent by advanced settings such as contrast, brightness, and shadows (the specific terms and settings all depending on the software used), thus creating the best tonal distribution for a particular image. As RAW can be developed over and over again, one can even go back to the original file and use particular tonal curves of other RAW converters. The possibilities are almost endless, which is not the case when shooting TIFF or JPEG, as these file formats already have an in-camera tonal curve been applied to. Of course, an additional tonal redistribution can be executed in image editing software, but will degenerate the final image quality and invite posterisation, certainly in 8-bit images [304].

One major option offered by some software packages is the linear conversion [394], hereby displaying the dark image by omitting any tonal curve. It is believed by some photographers that such a linear converted frame yields even better results, since maximum flexibility is offered in retaining all possible highlight and shadow details [830]. Besides better image quality, a completely linear converted RAW is scientifically very important in research that needs the originally generated DNs: spectral characterisation of the photographed features or the DSC itself (as in Chapter 10), mathematical operations between channels, spectroscopy, and astronomical photometry (e.g. [400]).

![Figure 7.15. Relative (%) and absolute (DN) reflectance increase in stressed vegetation canopy collected from a single, but differently processed RAW frame.](image)

Figure 7.15 shows an aerial image of a faint negative crop mark, indicating a possible Roman road in central Adriatic Italy. Three versions of the same NEF file are displayed. All were converted to an 8-bit TIFF, but conversion settings differed on two points: the WB – which was only applied in the last example (C) – and the tonal remapping, which was suppressed in (A). From the larger absolute differences in DNs – the latter acquired by taking the mean value of a zone in the healthy and one in the adjacent stressed canopy – it is obvious that most tonal curves seriously increase contrast in the middle input value range (DN divergence rises from 7 to >17), while the relative reflectance difference decreases (from 17.5 % to 11 %). Although the Red and Blue values are often scaled to the Green channel, this example shows that white balancing might in some occasions also alter the values of the Green channel to yield equal
RGB values for neutral objects. However, scaling does not have an influence on the relative reflectance increase (the small difference between tile 15B and 15C is due to rounding errors in the 8-bit space). Consequently, a linear converted RAW file (WB being omitted or not) is the only image that can give accurate reflectance information (if the DNs are expressed in absolute units), whereas visual interpretation is generally best performed on imagery that took the WB as well as the dissimilarity between the sensor’s response and the HVS into account (example 15C). Only this way, the yellow discoloration of stressed plants (called chlorosis) due to the leaves’ lost chlorophyll dominance over the carotenoids [13, 383] becomes visibly apparent.

7.6 RAW – Saving

Apart from these three major processing steps, capturing RAW allows to assign a specific colour space such as sRGB or Adobe RGB 1998 to the image (note that most of these colour spaces use a gamma of $1/2.2$, which explains why this particular gamma value was used in the previous step). Besides, there are particular possibilities to tackle noise (rather than some automatic noise abatement inside the DSC) and perform sharpening – a process that alters DNs again, with the additional risk of generating image artefacts (jaggies, halos, etc.). But finally, both the in-camera and computer workflow end with the choice of file format to ultimately save the processed RAW file in.

File formats are containers, created to hold digital data permanently and securely in files by offering a particular way of encoding the information. In order to define the format used for a specific file, a filename extension is often utilized, being a text string of usually three or four letters that comes in the file name after the final period. As there are a lot of different data types into circulation, many file formats exist. In essence, three fundamentally different graphical data types occur: raster data (as photographs), vector or geometry data (e.g. CAD-drawings), and latent image data (like the sensor’s RAW information). As mentioned, the latter stores both incomplete intensity information as well as metadata that need to be processed into a raster image file format, holding complete colour or greyscale data. Both TIFF and JPEG are raster file formats specifically designed to store still (or static) images as photographs.

7.6.1 JPEG

Rather than being a specific file format, JPEG is a large and complex compression standard defined in 1992 by the in 1986 created Joint Photographic Experts Group [607]. Instead of encompassing one algorithm, JPEG can be seen as a toolbox. In this toolbox, several algorithms reside together with optional capabilities [808]. When an application wants some of this standard to be implemented, a particular subset of the JPEG standard is selected to meet the requirements best.
As a matter of fact, the JPEG standard allows for two compression methods: **lossy** and **lossless**, meaning respectively ‘throwing away information’ and ‘storing all information’ [344]. Lossless compression always yields smaller files than the original ones, but the compression gained by a lossy technique can be much higher. Because it typically can only compress full colour data by 2:1 (i.e. the resulting file is two times smaller), lossless JPEG was never that popular. Lossy JPEG on the other hand is a completely other story. With a typical compression of 10:1, files can become really small when compressed with the **baseline lossy JPEG algorithm**, hereby almost visually indistinguishable from the input file. More compression can be achieved, but as more data need to be thrown away, the quality of the file will gradually deteriorate. Due to the possibility to trade off quality against file size, this baseline JPEG subset (which also features a few optional extensions) became worldwide implemented in most applications dealing with photographs [607, 808].

To store this JPEG encoded stream of pixel data together with the header containing all compression parameters (i.e. quantization tables and Huffman coding tables), a file format is needed. The **JFIF standard** (*JPEG File Interchange Format*) was therefore created by Eric Hamilton [363, 607]. When one talks about a JPEG file – recognisable by one of the possible extensions jpe, jpg, jpeg, jif, jfif and spf – in reality a JFIF or SPIFF (*Still Picture Interchange File Format*) file is meant, the latter being a more advanced substitute for JFIF, but without ever achieving significant adoption [558, 600]. However, this situation changed in 1996 when JEIDA approved version 1 of the **Exif (Exchangeable Image File format)** standard, a file format that was specifically designed for storing image data originating from DSCs. In essence, this format stores fully processed images using the baseline subset of the JPEG compression standard together with metadata embedded in various tags. On the memory card, these **Exif-JPEG files** are organised in particular directories and named according to rules defined by the **DCF (Design rule for Camera File System)**.

The Exif-JPEG image files themselves use TIFF tags to store the required metadata (such as camera model, capture date, and time) in one or more specific segments near the beginning of the image file, together with optional information on exposure values, lens settings, and GPS coordinates. Besides, the first segment also contains a thumbnail image of 160 pixels by 120 pixels [600]. The structure of the file was cleverly designed, so existing JPEG/JFIF readers can process these new files without a problem. Consequently, virtually all consumer DSCs currently (i.e. beginning 2009) store JPG compressed images in this standard Exif-JPEG file format with the JPG extension [600].

However, besides throwing away original information by the provided lossy compression, Exif-JPEG is also a scientifically unjustifiable file format because its limited tonal range has often implications on the displayed **Dynamic Range** (**DR**). To explain this, the dissimilar, but often intermingled concepts of DR and tonal range must be explained. If a DSC is capable of capturing a DR of eleven photographic stops, it means that about 2000 different light intensities can be recorded by this sensor in a single exposure, the smallest detectable amount of light intensity being eleven stops or $2^{11}$ times weaker than the strongest light intensity [615].
The bigger the difference between the faintest and brightest objects a sensor is able to capture simultaneously in one exposure, the larger its DR [421, 844]. A DSC with a high DR is preferable, in order to capture weak light signals without washing out the highlights. Outside this DR, everything becomes absolutely white or hidden in noise. The tonal range on the other hand is nothing but the number of tones a digital image has to describe the whole DR. A useful analogy may be a staircase: while the height of the staircase is its DR, bit depth equals the number of staircase steps [304].

Consider again the example of the sensor with approximately eleven stops of DR. In order to render these 2000 different intensities, the ADC needs to have at least a bit depth of eleven, enabling the discrimination of 2048 (i.e. 2\(^{11}\)) tones. However, if the signals coming from the same sensor would be quantized by an 8-bit ADC, only 256 tonal levels would be possible in the final image. Although the DR could be almost unaltered, the image would suffer from serious posterisation due to the loss of tones that could not be stored [176]. In this example, all eight units of linear intensity (2000/256) will be grouped into one of these 256 categories, disregarding 87.5 % (7/8) of all tonal values.

By depicting two sets of images with a different DR, Figure 7.16 illustrates these concepts. In the upper images, the tones range from dark shadows on the forested slopes to detailed highlights in the clouds. Only a small portion of the clouds is completely white without any detail. In the second row of images, a lot of highlight detail is gone and large parts of the clouds are pure white. Both narrow and wide tonal ranges can, however, be found in small as well as high DR imagery. So notwithstanding its intrinsic properties, using the aforementioned DSC to capture an eleven-stop scene by saving a JPEG, will cause tonal values to be lost,
because a JPEG file can maximally store eight bits/colour channel/pixel. The final DR will, however, depend on the workflow followed.

Dealing with an in-camera generated JPEG, firmware generally sacrifices a portion of the DR originally present in the RAW data by using a highlight-clipping tonal curve, consequently attributing enough tones to the shadow areas where the HVS can discern most gradations. What remains is a smaller (e.g. 7 stops to 8.5 stops) but better rendered DR, yielding a smoother image with good detail in the darker areas using the available tonal range. The largest DR most in-camera produced JPEGs can attain is about nine stops for ISO 100 (even when applying user defined curves), a value that will decrease at higher ISO settings [178, 233].

When shooting in RAW mode, no tonal curve is applied to the data, thus storing the sensor’s full DR. This is possible because most digital imaging sensors in current DSCs have a DR of about ten to twelve stops [408], while almost all prosumer and professional DSCs utilize a 12- or 14-bit ADC. As a result, a RAW workflow can extend the DR by at least one stop [177]. In this respect, capturing RAW enables some ‘exposure’ flexibility as shooting JPEG would ask for two different exposures of the same scene (i.e. bracketing) to capture the RAW’s file DR. This is visualised in Figure 7.17, which displays two enlarged portions (A and B) of the same low altitude oblique aerial photograph. Although a standard JPEG conversion (Figure 7.17A) disregarded all highlight details, the latter were brought back (and selectively darkened to make them more pronounced) by processing the RAW file (Figure 7.17B), using the same WB, demosaicing algorithm, and JPEG settings as (A).

Besides the possibility to recover highlight detail that would be lost forever if the camera was set to JPEG mode, the wider tonal range RAW possesses, offers more freedom to play around with tones and lowers the posterisation risk in case of image manipulation [304, 708, 735]. In both respects, JPEGs (certainly in-camera generated ones) truly reduce the creative and scientific options. Besides disregarding most of the acquired tonal values – which is scientifically indefensible –, the saved image will always be worse than the original data (perceivable or not), making JPEG totally unsuited for aerial imaging. Moreover, resaving a JPEG file (e.g. after post-processing) generally introduces compression errors, meaning the file’s quality is diminished every time it is opened and saved again [378].

Therefore, Exif-JPEGs should only be used at the end of the complete imaging workflow, serving purposes as small preview files and e-mail attachments, imagery to be embedded in databases and/or in presentations. In the case the DSC is only capable of initially storing JPEG files, one must make sure to use the largest file size and least compression possible, subsequently converting every JPEG into TIFF from the moment all frames are downloaded from the memory card.
7.6.2 TIFF

First published by Aldus Corporation in 1986 but currently maintained by Adobe Systems Incorporated, this Image File Format describes and stores raster images using Tags [15]. All the pixels which make up an image are stored in the body section of the TIFF file, while these so-called tags (which are also used in the Exif standard) hold information on width and depth of the image, acquisition date and time, copyright data, colour profiles, etc. TIFF was designed with the flexibility to define new tags in the future [600], which has led for instance to the development of a tagset for carrying georeference information, enabling the TIFF to be located somewhere on the earth’s surface. This new standard was called GeoTIFF [669].

In contrast to JPEG, TIFF can handle 16-bit/colour channel data and serves as a container for both uncompressed as well as compressed images. In the latter category, often the lossless LZW (Lempel-Ziv-Welch, named after Abraham Lempel, Jacob Ziv and Terry Welch) is offered [15, 344], besides the lossy JPEG compression algorithm. The virtue of a TIFF file (recognizable by the extension tiff or tif) is its ability to store all data in the original order, hereby containing all captured colours and other pixel related information (if no lossy JPEG compression is applied). Uncompressed TIFFs are insensitive for the aforementioned accumulative data loss.

Figure 7.17. Comparison between highlight details of an in-camera generated JPEG (A) and a JPEG resulting from a RAW workflow (B).
Additionally, its support for 48-bit imagery and the fact it is portable (i.e. supported across different platforms like Windows, Macintosh, UNIX and no favour for particular file systems such as HFS, FAT, NTFS, etc.) make it the world’s number one preservation format for master copies – and hence the standard to save developed RAW frames (although the latter should never be thrown away). These characteristics of course have their drawbacks. Due to the fact that all possible data are stored, large file sizes occur (with a maximum of four gigabytes). What is more, the flexible set of information fields or tags sometimes leads to problems concerning the correct opening or interpretation of the image file, because TIFF-enabled software packages do not always support the new tags added by the wide variety of (scientific) users. Although this issue created a new interpretation of the acronym – Thousands of Incompatible File Formats –, it is a problem not that often encountered.

Finally, shooting in-camera generated TIFFs is far from ideal, as most DSCs only output a 24-bit image [736]. Even though most of the sensor’s DR can be captured and the tonal range offers large editing headroom in case a 16-bit/colour channel TIFF can be saved, this file format presents at the capturing state no advantages to RAW, as its file size will be much larger and the interpretation of the scene is already performed (just as in the case of JPEG), with no chance of addressing the originally captured DNs. Consequently, few DSCs provide TIFF as an option.

### 7.7 RAW – Workflow and Software

Under most circumstances, the best processed and least compressed JPEGs are comparable to RAW images from a perceptual point of view. As long as the real-world scene being photographed is characterised by a limited DR, the exposure is spot-on and no (or little) editing is required, 8-bit JPEGs coming straight from the DSC can just be good enough [453]. This is possible in environments as studio’s where the light can be completely controlled, but the question remains whether one can be sure all these demands are fulfilled when shooting aerial photographs with uncontrolled and ever changing lighting? Consequently, a RAW workflow always pays off, even when aerial photographs are only needed for visualisation and interpretative purposes: after all, most RAW developing software enables a fast way to create TIFFs/JPEGs from the stored RAWs. All it takes is tweaking the WB and tonal distribution once, before a batch process can apply all the settings to the RAW frames, converting them to qualitative JPEGs with good demosaiced spectral channels. This approach is hugely superior to the JPEG workflow, even if it would be possible to adjust all camera settings on an image-by-image basis when flying.

An alternative might be the capture of a RAW+JPEG combination, a function offered by most DSCs: the JPEG image is there for immediate use, while the RAW can be processed at a later time and/or serve as back-up in case the JPEG is corrupted (or vice versa). However, the storage space and processing power needed are the major drawbacks of this workflow, with the stored
JPEG even quantitatively inferior to one generated by a batch process initiated a few minutes after the (aerial) shoot has ended. Some RAW formats (e.g. NEF by Nikon) already include a processed Exif-JPEG file which can be extracted by software that is not capable of processing RAW files, hence improving the workflow.

From a scientific point of view, it is impossible to disregard RAW since no other means are available to work with all the acquired spectral information in its most pure form. However, to create something meaningful out of these raw data, both scientists and photo artists need RAW conversion software to construct an image out of the recorded DNs, or at least a programme that can decode the specific data format. Different software packages exist, all with their own specific advantages and drawbacks. Each package uses more or less proprietary techniques for demosaicing, noise reduction and tonal distribution, with the final aim to allow for the best possible image to be created by the data available. Therefore, different converters produce different results.

In the early RAW days, most camera manufacturers provided some basic converter software with the purchase of the DSC. Nowadays, some of the best RAW converters are typically stand-alone programmes or plug-ins created by third-party companies. Due to the amount of software on the one hand and the options available on the other, a detailed comparison will not be given. The list provided in Table 7.1 is just a dry enumeration of a few multi-camera capable converters, mostly optimized for a complete, colour managed RAW workflow, enabling the efficient processing of thousands of photographs. The last software items – UFRaw and dcraw – are worth mentioning, because they are for free. The first package – UFRaw (Unidentified Flying Raw) – is Open Source software. It functions on its own or as a plug-in for the Open Source image editing package GIMP (an acronym for GNU Image Manipulation Program).

**Table 7.1. Overview of various RAW software.**

<table>
<thead>
<tr>
<th>RAW software</th>
<th>Manufacturer</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACDSee Pro</td>
<td>ACD Systems</td>
</tr>
<tr>
<td>Aperture</td>
<td>Apple</td>
</tr>
<tr>
<td>Bibble</td>
<td>Bibble Labs</td>
</tr>
<tr>
<td>BreezeBrowser Pro</td>
<td>BreezeSystems</td>
</tr>
<tr>
<td>Camera Raw</td>
<td>Adobe</td>
</tr>
<tr>
<td>Capture NX</td>
<td>Nikon</td>
</tr>
<tr>
<td>Capture One Pro</td>
<td>Phase One</td>
</tr>
<tr>
<td>DiMAGE Master</td>
<td>Konica-Minolta</td>
</tr>
<tr>
<td>Digital Photo Pro</td>
<td>Canon</td>
</tr>
<tr>
<td>Lightroom</td>
<td>Adobe</td>
</tr>
<tr>
<td>UFRaw</td>
<td>Open Source</td>
</tr>
<tr>
<td>dcraw</td>
<td>Open Source</td>
</tr>
</tbody>
</table>
The second programme, dcraw, is an excellent free tool and the digital child of Dave Coffin who tried to "write and maintain an ANSI C program that decodes any raw image from any digital camera on any computer running any operating system" [184]. In its present form, dcraw supports over 300 different DSCs. As it is completely free and offers the possibility to process RAW frames without any WB and/or tonal curve and with a choice of demosaicing options, dcraw was implemented by many other RAW programmes (ACDSee Pro, BreezeBrowser, UFRaw, and many others), hereby making use of its excellence and offering the possibility to call dcraw from a graphical interface. Besides dcraw, several functions also exist to read RAW imagery into the technical computing environment MATLAB (from The Mathworks), after which an unlimited amount of Digital Image Processing (DIP) techniques are free to explore.

A last, often overlooked advantage of RAW is the possible future improvement of RAW converters. As software mostly gets better with each new version, RAW converters can still be the subject of major development, delivering even improved imagery created by superior demosaicing and noise suppression algorithms. Capturing RAW allows for the exploitation of these improvements, as the ‘digital latent image’ can be developed again in the future. In this respect, RAW presents the greatest long-term flexibility.

7.8 RAW – Disadvantages

There are only three excuses for not using RAW: the DSC is not capable of storing RAW; it is a compact or hybrid model that freezes for some seconds after shooting and saving a RAW image; the lack of storage space. Although the latter should not happen, it can occur. In those situations, there is only one rule: buy more memory, but capture the shot as JPEG, as it is far better than missing the photograph. In the other cases, RAW is the format to go for, even though a RAW workflow might saddle (aerial) photographers with some minor disadvantages. Generally speaking, the problems with RAW are threefold.

7.8.1 File Size

As most RAW formats are not compressed, RAW files are much larger than similar JPEG files. A typical file from a 6 MP camera is about 7 MB to 9 MB (6 000 000 pixels x 12 bits/pixel = 8.6 MB), whereas a similar JPEG file with low compression is about 2 MB to 3 MB. Although the size of a JPEG largely depends on the content of the image and the amount of compression applied, RAW files are often three to five times larger than the same image saved with JPEG compression. To tackle this issue, some DSCs use a compression algorithm to decrease their RAW size. To not mortgage the quality, these compressions are lossless, meaning no information is lost (except for some Kodak and Nikon models, which sometimes create slightly lossy, but visually lossless compressed RAW files). However, RAW compression algorithms can
never achieve the level of compression JPEG can. Resultantly, there will always fit less RAW images onto a particular memory card than JPEGs. Moreover, more RAM and computing power are needed in comparison to working with JPEGs.

Both issues should, however, largely be resolved by the fact that RAM and memory cards significantly dropped in price during the last years, making the file size less of a problem. However, the size of RAWs can still remain a concern in case of a DSC’s maximum frame rate, burst rate, and buffer capacity. The maximum frame rate of a camera quantifies the amount of consecutive images/frames it can maximally take in one second (e.g. 5 fps). Burst rate or burst frame rate equals the number of frames the DSC is able to shoot at the maximum frame rate. Because the DSC is not able to transfer all these images instantaneously to the memory card, they need to be temporarily stored in the buffer. If the latter is full, the DSC will stop shooting or reduce the frame rate significantly. Bigger RAW files will limit this burst rate since they fill the camera’s buffer faster and take longer to be written onto a memory card. Here, the differences between shooting JPEG and capturing RAW can be really significant. To give an example: the Canon EOS-1Ds Mark III shoots in continuous mode up to twelve RAW frames, while it enables the capture of 56 low compressed JPEG images. Finally, a simple comparison with an 8-bit 6 MP TIFF file clearly shows RAW to be the winner concerning file size, as this TIFF will approximately be 18 MB large (i.e. 6 000 000 pixels x 3 bytes/pixel).

7.8.2 Proprietary File Format

As stated before, no two RAW formats from diverse companies are alike. Even though the structure of RAW files is in most cases a specific flavour of the TIFF or TIFF/EP standard, the formats are highly proprietary and unstandardized. Due to constant improvements, even different RAW formats are created by the same DSC manufacturer. As a result, about 300 RAW formats are in existence as of 2008, some of them with their own extension. These proprietary file formats bring along a major drawback: software which has the ability to handle these particular RAW files is needed. Instead of releasing one version every two or three years, all RAW software converters need to be updated every moment a manufacturer brings out a DSC which creates a new RAW file format, just to offer support to a RAW range as wide as possible. Additionally, certain RAW files might not be supported any more by RAW converters within five years, making them impossible to access.

Moreover, most big camera companies as Canon and Nikon even try to hide or encrypt parts of their metadata to make it impossible (or at least harder) for others to decode the format. Nikon’s D2X came in 2005 with encrypted WB data to force Nikon consumers to only make use of Nikon software (which can of course decrypt the data). Just as was the case with the encrypted RAW files Sony used in 2003, it was only a matter of days before programmers cracked these files. However, big companies such as Adobe and Phase One do not dare to implement this crack into their software, because it could expose them to liability.
Table 7.2. The extensions of some RAW formats.

<table>
<thead>
<tr>
<th>Extension</th>
<th>Manufacturer</th>
</tr>
</thead>
<tbody>
<tr>
<td>crw</td>
<td>Canon</td>
</tr>
<tr>
<td>cr2</td>
<td>Canon</td>
</tr>
<tr>
<td>dcr</td>
<td>Kodak</td>
</tr>
<tr>
<td>dng</td>
<td>Adobe</td>
</tr>
<tr>
<td>mrrw</td>
<td>Minolta</td>
</tr>
<tr>
<td>nef</td>
<td>Nikon</td>
</tr>
<tr>
<td>orf</td>
<td>Olympus</td>
</tr>
<tr>
<td>pef</td>
<td>Pentax</td>
</tr>
<tr>
<td>raf</td>
<td>Fuji</td>
</tr>
<tr>
<td>srf</td>
<td>Sony</td>
</tr>
<tr>
<td>x3f</td>
<td>Sigma</td>
</tr>
</tbody>
</table>

As a reaction, the OpenRAW Working Group was founded which proposed two possible solutions to this proprietary file format issue: the adoption of a universal RAW standard or the public documentation of RAW formats. In an attempt to contribute to the former solution (and of course further strengthen their role in the image business), the Digital NeGative (DNG) format was announced by Adobe in 2004. It was their attempt to standardize the RAW file format. To make sure everybody would use the new concept, Adobe launched a free, simple batch DNG converter for both Windows and Macintosh, capable of handling most current RAW formats. Given that companies are always doubtful regarding a standard created and owned by a single corporation, only four DSC manufacturers implemented the DNG format as their native RAW format since 2004: Hasselblad, Leica, Ricoh, and Samsung [16].

Due to the resistance from a lot of manufacturers to even incorporate this DNG specification in their RAW software, DNG often seems to be just another RAW format. The OpenRAW group therefore opts for the other solution, compelling the camera manufacturers to publicly document their RAW image formats [732]. By using only openly documented RAW formats, digital photographs created now and five years ago should also be readable over thirty years.

7.8.3 Time Consuming

The flexibility RAW gives in processing each individual image can often lead to relatively long ‘development’ times. However, the time needed depends on both the final purpose of the file as well as on the photographer. Artists trying to make the best image ever can literally devote hours to this RAW conversion, tweaking every possible setting in the software and combining layers of differently developed RAWs into one final piece of art. These photographers are the Ansel Adams’s of the digital age and their aim to create arty images compels complicated workflows.

However, RAW conversion does not need to be that time-consuming. In case a JPEG file is sufficient for viewing or printing purposes, only some initial settings need to be made, after
which batch processing can be applied. Post-processing RAW images can even be a real time winner in certain cases. Imagine that WB adjustments are required for hundreds of aerial images. Using a RAW workflow, it is much easier and more time-efficient to apply these corrections on RAW images than on any other file format, certainly if this can be integrated in a batch conversion.

### 7.9 Conclusion

With the advent of RAW digital photography, (aerial) photographers are now given the chance to get the maximum from their photography, as a RAW workflow enables enormous control over the final output. Unprocessed files can always be accessed and developed, while converted files may be reprocessed if needed. Even when the high-end results and (scientific) flexibility offered by RAW is not needed, capturing RAW frames can in fact speed up the entire workflow if corrections need to be made. Moreover, batch-processing the whole RAW series with the converter’s default values generally delivers an output that will match at least the result of the in-camera generated developed photograph. For strictly scientific photography such as airborne remote sensing, shooting RAW is mandatory practice because it is the only way to assess the originally captured DNs (which is of the utmost importance in mathematical channel operations, intensity measurements, calibration, and spectral characterisations).

A last statement often heard is that slide film – loved and (still) used by so many photographers – also yielded an image with a limited DR that was completely finished at the time the shutter button was pressed. Although that assertion stands up to scrutiny, it should not restrain photographers – and certainly not scientists – from exploiting new camera techniques and possibilities, neither because of these reminders of an analogue past, nor by the fact some technical photographic knowledge is often a necessary prerequisite to take full advantage of the new imaging technologies.
Pushing Theoretical Limits
Near-InfraRed Aerial Archaeology

« Theory without practice is empty. Practice without theory is blind. »
John Dewey (1859-1952), philosopher
Imaging the Invisible

Modified Digital Still Cameras for Low-Cost and Straightforward Archaeological Near-InfraRed Photography

The content of this chapter is published in

*Journal of Archaeological Science*

[ERIH (A); SCIE, SSCI, AHCI – IF 2007: 1.439]

Abstract

Analogue Near-InfraRed (NIR) photography has already been used a lot in both scientific and medical photography, in which cases NIR radiation was mostly captured by InfraRed (IR) sensitive plates or film emulsions. However, its use in archaeology has remained rather restricted, most likely due to some ignorance and/or lack of knowledge about this kind of photography, while the critical imaging process also severely limited its use. This situation could be, however, changed completely, as the image sensors used in Digital Still Cameras (DSCs) are very sensitive to NIR wavelengths, making the quite lengthy and error-prone film-based NIR imaging process obsolete. Moreover, modifying off-the-shelf DSCs even simplifies this digital acquisition of NIR photographs to a very large extent.

By starting with a general outline of the ElectroMagnetic (EM) spectrum and the specificities of NIR radiation, the base is laid out to explore the practicalities and some possibilities of archaeological NIR imaging, subsequently comparing the earlier film-based approach with the digital way of capturing NIR, showing how the latter can greatly benefit from modified compact, hybrid, and small-format Single-Lens Reflex (SLR) DSCs. Besides in-depth information on the technique of digital NIR photography, examples will illustrate its archaeological potential.
8.1 Introduction

Since the astronomer and composer Sir Frederick William Herschel (1738-1822) discovered in 1800 the InfraRed (IR) portion of the ElectroMagnetic (EM) spectrum, a lot of scientific disciplines have become fascinated by this kind of invisible radiation. This interest even increased after World War II, as Colour InfraRed (CIR) emulsions had shown their capabilities. These days, IR photography is used a lot in forensics, astronomy, aerial survey, (bio)medical, and several other types of scientific photography, although its use in archaeology is rather disappointing. Because this can largely be attributed to a lack of knowledge about the technical specification of Near-InfraRed (NIR) radiation and the concept of EM radiation, the specific character of NIR will first be explained before tackling the possibilities and practicalities of archaeological NIR imaging.

8.1.1 EM Radiation and NIR

The light that makes the Human Visual System (HVS) perceive the world around us is in fact an EM wave, comprising two oscillating magnetic and electric fields perpendicular to each other as well as perpendicular to the direction of propagation [723, 804] and travelling at 299 792.458 km/s in vacuum, a speed that decreases when light travels in air, glass, water or other transparent substances [846]. This explanation, put forward almost two centennials ago by the Scottish physicist James Clerk Maxwell (1831-1879), was based on the findings of the Dutchman Christiaan Huygens (1629-1695), who already declared light to travel in the form of waves [181, 804]. Being a wave phenomenon, the wavelength (\(\lambda\)) is the most important characteristic of EM radiation.

The EM waves humans perceive – the so-called visible light – encompasses a very small portion of all EM radiation: only wavelengths between approximately 380 nm and 750 nm (Figure 8.1), the absolute thresholds varying from person to person and specific viewing conditions. However, on both sides of this extremely small visible spectrum resides EM radiation the HVS is insensitive for, characterised by wavelengths smaller than 400/380 nm or larger than 700/750 nm. Just as visible light, these wavebands were divided into spectral regions and given names as gamma rays, X-rays, and UltraViolet (UV) rays on the short-wavelength side, while IR rays, microwaves, and radio waves can be found in the long-wavelength region (Figure 8.1).
The aforementioned IR portion of the EM spectrum comprises wavelengths between 750 nm and 1 mm, hence spanning three orders of magnitude. As the width of this particular band is exceptional, it is often subdivided into several zones, for which the limits (and also the number of subdivisions) are to a certain extent dependent on discipline and largely varying through literature. In general, the following breakdown can be used (based on [216, 231]):

- Near-InfraRed (NIR) from 750 nm to 1400 nm (1.4 μm);
- Short Wavelength InfraRed (SWIR) from 1.4 μm to 3 μm;
- Mid Wavelength InfraRed (MWIR) from 3 μm to 6 μm;
- Long Wavelength InfraRed (LWIR) from 6 μm to 15 μm;
- Far/Extreme-InfraRed (FIR) from 15 μm to 1000 μm.

Moreover, EM radiation can also be thought of as a travelling bundle of particles. Although this theory was contested by many since its launch by the Englishman Isaac Newton (1642-1727), Albert Einstein (1879-1955) finally demonstrated the existence of such discrete energy packets, now called photons [181, 804]. This wave-particle duality is still one of the key concepts in quantum mechanics, signifying EM radiation exhibits both wave and particle behaviours. Depending on the wavelength of the EM radiation, the energy of the photons differs. Applied
to NIR, it means this type of radiation contains less energetic photons compared to visible light.

### 8.1.2 NIR Imaging

There exists a lot of confusion and misconception about NIR imaging, often linked with images like the ones displayed in Figure 8.2. However, these so-called heat images were yielded by **electronic thermography**, a technique based on a completely different part of the EM spectrum. Heat imaging uses the MWIR and LWIR [659], energy given off by all real-world objects [56]. As a matter of fact, all objects with a temperature above absolute zero (0 °K or -273.15 °C) emit EM radiation, but the type and amount of the latter largely depends upon the temperature of the matter. A healthy human body with a temperature of 310 °K (circa 37 °C) gives off wavelengths with a peak around 9350 nm (LWIR) and no detectable amounts of NIR energy. With a rising temperature, the amount of radiated EM radiation will increase and the wavelength of maximum emittance $\lambda_{\text{max}}$ will be shorter (as described by **Wien’s displacement law**). Practically, objects need to be heated to about 500 °K before they start radiating in the NIR range, while a temperature of at least 800 °K (e.g. an electric stove burner) must be attained before visible red light is emitted [56, 650]. Because it is not possible to photograph IR radiation but the NIR portion with conventional film-based approaches or digital photo cameras, performing NIR photography basically boils down to capturing the particular amounts of reflected NIR radiation emitted by very hot objects such as the sun, incandescent light bulbs or specific extraneous NIR sources, rather than recording the ambient temperature variation. This misunderstanding is also fed by some of the older (archaeological) literature. Simmons ([715]: 94) for instance literally says: “cool objects appear dark, warm objects appear light; hence green vegetation looks very whitish, while water is blackish”. Although water indeed appears black on an NIR photograph, this response has nothing to do with its temperature, but is due to a very high NIR absorption, indicated by its absorption coefficient [210].

![Figure 8.2. (A) LWIR image of a man’s face and (B) MWIR image of a four-fingered hand ([659]: Figs. 6.2 and 2.19).](image)
8.1.3 Two Techniques

In NIR imaging, two techniques exist. The first, and by far most applied, kind of photography uses the already mentioned reflected/transmitted portion of the incident NIR radiation. Every object exposed to NIR radiation will absorb, reflect, and transmit these incident photons to some extent. Recording these particular amounts is generally termed reflected (N)IR photography and should not be confused with IR reflectography, the latter using longer wavelengths up until around 2000 nm [765, 766].

Secondly, there is **NIR fluorescence photography** (sometimes called NIR luminescence [56, 113, 321, 323, 322]), in which the NIR sensitive medium records in fact fluorescence, being radiation emitted by the subject under study in the NIR region. Rather than directly being emitted as in the case of very hot objects, these NIR photons are excited upon being exposed to incident shorter wavelengths (mostly UV or visible Blue and Green wavelengths). Although its application is mainly restricted to the forensic field, the very strong fluorescence of particular minerals and pigments [54] makes this type of NIR photography worthwhile in certain archaeological case studies.

8.2 NIR Imaging in Archaeology

It lasted till the 1930s for NIR sensitive emulsions to become relatively available, allowing photographers to practise this new technique with a certain ease and certainty. From this period onwards, the possibility to visualise an often subtle, dissimilar behaviour of materials in the NIR helped archaeologists to depict certain object characteristics not (or less) apparent to the HVS. However, despite its application and the great deal of work done in a wide variety of scientific research fields, the scale of utilization always remained very small from an archaeological perspective – for instance illustrated by the fact that Cookson’s “Photography for Archaeologists” [193] does not mention one single time the use of NIR. In broad terms, archaeologically inspired NIR photography has been applied over the years to answer questions in all sorts of research areas.

8.2.1 The investigation and decipherment of documents

Charred, dirty, worn, bleached, censured, obliterated, faded or very deteriorated documents often reveal their secrets through recording their NIR reflectance and/or fluorescence. While some inks are largely transparent, others do reflect NIR to a larger extent. Moreover, NIR often can differentiate between dyes and pigments that look indistinguishable to the HVS [118, 672]. Exploiting these spectral signature properties might bring document alterations to light or reveal the underwriting of obliterated passages in case the top ink is less opaque to NIR than the ink below.
This way, direct NIR photography made a faint black Egyptian text from about 1200 B.C. written on very dark brown and aged leather sufficiently readable again [460], while other scholars deciphered original writings on fragments of badly discoloured [63] or charred papyri [158]. Famous is also the work of the Arab photographer Najib Anton Albina, who used reflected NIR photography in the 1950s to retrieve information from the Dead Sea scroll fragments [686], whereas UV induced NIR fluorescence disclosed some of Vindolanda’s wooden Roman writing tablets [676]. The results achieved by Coremans [196] even proved NIR photography to reveal writings and drawings underneath encrustations (Figure 8.3) or a patina.

Figure 8.3. (A) Panchromatic and (B) reflected NIR photograph of the same artefact (adapted from [196]: Figs. 13a and 13b).

8.2.2 The examination of textiles and tapestry

By differentiating between the NIR reflectance and/or transmittance of the particular pigments, dyes, and materials used, NIR photography also furnishes a useful means for the examination of fabrics, even if they appear visually similar [241]. This way, Coremans [196] detected tapestry restoration. Baldia and Jakes [49] even incorporated NIR photography in a complete range of photographic methods for non-destructive research of archaeological textiles.

8.2.3 Inspection of Paintings

The same principles hold for the investigation of all kinds of paintings: on canvas, rock, wood or glass. Being a true pioneer, Marshack [511] utilised NIR photography in some French caves to get a deeper insight into the employed pigment and the dating of the Cro-Magnon man’s cave art, while NIR imaging also proved several times to be indispensable in the detection of painted forgeries. Although such research is situated often in the field of art studies, proving the authenticity and examining the fake or altered state of the canvas also largely rely upon
NIR fluorescence [113] and direct NIR photography [21, 501], as the latter might reveal information which the technique of IR reflectography is unable to unveil [312].

8.2.4 Tattooing research

In several studies, NIR reflected radiation has been taken into account to detect forms of tattooing on human remains that were naturally or artificially mummified [33, 725].

8.2.5 Pottery Study

Traces of pigments can also be found on ceramics, in which case photography by NIR can help enormously in the study of tituli picti, decoration, etc. Used in combination with radiography, Milanesi [538] demonstrated the benefit of NIR photography in dating and classifying prehistoric and protohistoric pottery fragments.

8.2.6 Aerial Photography

Notwithstanding its rather large potential in archaeological aerial reconnaissance, very little non-visual photography has been executed, although it still remains one of the fields where most archaeological NIR-based research is undertaken (e.g. [17, 109, 267, 354, 366, 637, 664, 744, 776] and Chapter 9).

As healthy vegetation appears very bright and the NIR reflection of a stressed, diseased or dead vegetation canopy is much less abundant, crop marks can often be distinguished better in this part of the EM spectrum (Figure 8.4; for a more in-depth discussion consider Chapter 9, Chapter 10 or [776]). Besides, reflected NIR photography enhances certain soil marks because
water largely absorbs incident NIR, while the same principle makes it very easy to discern bodies of water, the latter generally appearing black on the photograph. Thirdly, the long NIR wavelengths largely penetrate the haze often present when imaging landscapes, especially in the case of high altitude and oblique aerial photography, hence increasing the visibility of distant objects [191, 672]. The explanation lies in the fact that atmospheric scattering is much less in the NIR region than it is in the visible part, yielding imagery with enhanced clarity of detail and a larger contrast [648].

8.2.7 Excavation photography

Both in imaging the horizontal and vertical sections, NIR site photography has been proved very useful. Some examples are the work of Buettner-Janusch [123], who claimed features and stratifications to be visible in pure NIR imagery, while not discernable on the panchromatic frames [i.e. Black-and-White (B&W) film sensitive to all visible wavelengths] nor to his eyes. The same conclusion was drawn by Reichstein [656], who effectively acquired B&W NIR imagery to reveal previously indiscernible Bronze Age plough marks. Additionally, Hirsch [391] definitely demonstrated the presence of painted floors at the Cretan Archaeological site of Gournia, as NIR photography revealed the colour of one of the floors to be owed to pigment rather than clay.

This short overview indicates that NIR photography certainly holds several benefits in archaeological research, while it is most likely that much more subjects are very suited for this kind of imaging [191], or as Matthews says: “It so often happens that all other methods failing to record the desired result, that the operator, often in desperation, experiments – and finds his answer in photography by infra-red” ([520]: 131-132). Till only very recently, archaeological NIR photography was performed using films whose emulsion was sensitised into the NIR range. Even though recent advantages in digital technology made this cumberstone process obsolete, it is best to review how NIR imaging was performed in the film era before expatiating on the digital approach and its additional advantages.

8.3 NIR Imaging in the Film Era

Although generally the same cameras and light sources can be used as for imaging reflected visible light, NIR photography still features some peculiarities. In the following overview, the major changes and additional requirements over ‘normal’ photography will be treated (here to be considered small format/35 mm frame photography – see Chapter 6 for this terminology).
8.3.1 Emulsions

NIR sensitive film exists in two variants, being B&W negative NIR film (monochromatic IR film) such as Kodak Professional High-Speed Infrared Film (HIE), or the (False-)Colour Infrared (FCIR or CIR) film, from which the Kodak Ektachrome Professional Infrared EIR film is an example. As it also features only one layer of emulsion, the first film type is to a certain extent similar to a conventional panchromatic emulsion, although it is sensitised up to 900 nm. Due to its non-uniformity in the visible bands, this kind of film is often used for true NIR photography: capturing only wavelengths between 750 nm and 900 nm by fitting an appropriate filter on the lens.

(F)CIR film, which comes in both negative/print and positive/reversal/transparency alternatives, also bears great resemblance in construction to colour film, being an integral tripack material which makes use of dyes added to the silver, spectrally sensitizing the silver-halide grains. However, the three individual emulsion layers react with different portions of the EM spectrum. The upper layer, which contains cyan dyes, only reacts to NIR, while the other two layers, containing Yellow and Magenta, react to Green and Red light respectively (Figure 8.5) [253, 449]. As all three layers also respond to Blue radiation, a Yellow (minus-Blue) filter is used (or implemented in the emulsion) to cut out its image degrading effect, besides limiting the sensitivity of the layer to its intended spectral region. By the effects of exposure and processing, each individual layer produces a dye of a complementary colour (Figure 8.5). Consequently, NIR wavelengths are imaged as Red, while reflected Green radiation is visualised as Blue and Red reproduces as Green. This way, one subtractive colour (Cyan, Magenta or Yellow) controls each primary colour, while the colours of the original object are remapped to pseudo-colours, attributing to this kind of film the term false-colour.
No matter which type of film is used, the NIR sensitivity of most films dramatically drops around 900/925 nm (their specific upper cutoff wavelength depending upon the dyes added), with an emulsion as Konica Infrared 750 even responding to only 820 nm.

### 8.3.2 Lenses

Just as in ‘normal’ photography, about any lens can be used in NIR photography [449] as the majority of optical glasses and polymers freely transmit NIR [651], although the design of the lens still counts when one wants to avoid a **hot spot**: a brighter area in the centre of the image, produced by internal light reflections caused by the lens coatings. Moreover, because NIR radiation features longer wavelengths than visible light, the waves are less refracted and focus to another point which lies behind the visible focal plane (a phenomenon called **longitudinal chromatic aberration** [361]). To counteract, the lens has to be focused on an object that is closer than the actual object, a process known as **short focusing**. This explains why many pro and/or older lenses have a red dot/line/letter “R” indicating the NIR focus offset when focused at infinity [575]. Newer lenses seldom display such a supplementary focusing index, nor do they have an NIR Depth Of Field (DOF) scale (the latter indicating which parts of the subject, expressed as a distance in meter or feet from the lens, will be rendered sharp). Besides short focusing, a smaller diaphragm setting (e.g. f/8 or f/11) – with corresponding extended DOF – will largely solve most of these focus errors. Too small an aperture may be counterproductive, as **diffraction effects** start to come into play and shutter speeds might become lengthy.

### 8.3.3 Filters

As all aforementioned films are largely sensitive to visible light, optical filters are needed to exclude parts of this spectrum (or completely in case of pure NIR photography). In a typical photographic situation, a deep red filter (most commonly a Wratten 25, B+W 092 or similar) is attached to the front of the lens, only allowing some Red and NIR radiation to pass. Such filters make sure the photographer can still see a dim image in the viewfinder, enabling him/her to compose and focus, while not completely loose the NIR look [324, 325]. In case a pure NIR image is required, an IR filter is needed. These completely visually opaque filters (type Hoya R72, Wratten 87 or similar) transmit only radiation in the NIR waveband [191], hence prohibiting visible light from being recorded (the cutoff wavelength and exact amount of radiation transmitted being filter dependent). As composing and focusing is now made completely impossible, the lens must be prefocused and the image composed before installing the filter on the lens. This inconvenient situation even becomes worse when using lenses with different threads, as this necessitates filters in different sizes. In studio situations, this could largely be resolved by placing the appropriate filter over the light source instead. As different filter brands exist (B+W, Cokin, Hoya, Kodak, Lee, Schott) all with their specific screw-in or gel type filters, it is advisable to check the particular spectral curves or transmittance chart for each filter, as it is impossible to use only one, ideal filter for all purposes.
The fact that incident radiation generates longer wavelengths constrains two different filters to successful accomplishment of NIR fluorescence photography: first of all, an appropriate *exciter/excitation filter* must be placed over the light source (generally a Blue/Green-pass filter) to make the latter only emit shortwave visible light and prevent any NIR photons to hit the subject. Secondly, the aforementioned opaque NIR filter (in this case called *barrier filter*) on the lens must make sure only the NIR excited photons are captured, excluding the reflected visible light [321].

### 8.3.4 Disadvantages

This film-based approach has, however, some major disadvantages, which make the complete workflow rather critical and NIR imaging difficult to master.

#### 8.3.4.A Exposure

Being generally the biggest obstacle to getting a good NIR photograph, a solid knowledge of NIR principles is needed to make the right decisions about exposure. The problem lies in the fact that normal exposure meters do not take NIR into account. Consequently, the exposure reading one gets is based on the visible light hitting the exposure sensor rather than the NIR radiant flux. Moreover, NIR sensitive films only have a recommended kind of film speed instead of a fixed ISO value, as the latter is always defined using the visible part of the EM spectrum [257]. Consequently, the exposure reading has to be considered a starting point, after which over- or underexposure is needed according to the subject and the lighting conditions photographed in (time of the day, time of the year, geographical location, weather conditions, NIR source). Only with enough experience, the photographer can estimate the quantity of NIR radiation reflected from or emitted by particular objects. Certainly in aerial work, this is a big disadvantage, as the amount of NIR strongly varies from scene to scene. Consequently, *bracketing* – i.e. taking a series of photographs with different exposures – is advised.

#### 8.3.4.B Sensitivity

Notwithstanding the improvements made, the current emulsions are only sufficiently sensitive to be used in hand-held cameras with rather large amounts of reflected NIR radiation. Moreover, the application of filters additionally extends the exposure time, meaning aerial photography is for instance impossible but in very sunny circumstances [366], as other conditions would considerably increase the risk to acquire blurred imagery.

#### 8.3.4.C Storage, handling and processing

Although the films cannot detect thermal radiation, their extended sensitivity makes them more susceptible to *fogging*. NIR emulsions therefore have to be stored and transported refrigerated [255]. Storage in less than ideal conditions can easily lead to sensitometric
changes. Not only opening the black plastic film canister, but also loading and unloading the camera have to be executed in absolute darkness [324]. After exposure, the film must be developed as soon as possible, but not without the necessary precautions: e.g. using NIR-safe processing tanks, turning off any IR sensor in the labs, etc.

All these elements made NIR imaging a rather tricky business, and although experienced photographers certainly could get very decent results, the final outcome could never be entirely predicted. Consequently, NIR (archaeological) photography was usually only attempted by skilled photographers, scientists, and technicians with a particular purpose in mind. Add to this the cost of the NIR film, one can already imagine the huge advantage and democratization a digital approach could mean to NIR photography. In spite of this, the use of NIR imaging with digital consumer or SLR cameras was till now never fully explored in archaeology.

8.4 Digital-Based NIR Imaging

Since the 1990s, much has changed in the photographic world. Certainly since the advent of the 21st century, there is an ever increasing growth of digital shooters due to the large availability of sophisticated but affordable Digital Still Cameras (DSCs) and major advances in computer technology. Unlike video or silver halide photographic cameras, a DSC equals a camera equipped with both a digital image sensor for capturing photographs and a storage device for saving the obtained image signals in a digital way [760]. Besides all the well-known benefits in acquisition, storing, manipulating, and retrieving digital imagery, DSCs offer an additional major advantage, making NIR photography much less awkward.

8.4.1 Sensor

Being it the often used Charge-Coupled Device (CCD) and Complementary Metal Oxide Semiconductor (CMOS) or the less implemented Junction Field Effect Transistor (JFET) and CMOS-Foveon X3 sensors, pure silicon (Si, atomic number 14) is the basis for essentially all digital image sensors of current DSCs: from low-cost consumer point-and-shoot and hybrid cameras to expensive, professional small-format D-SLR (Digital Single-Lens Reflex) cameras. Pure silicon is a semiconductor: a material whose physical properties lie somewhere in-between electricity conductors and insulators. It exists in the form of a crystal lattice with each silicon atom bonded covalently to four other silicon atoms [714].

Because this arrangement does not allow silicon to conduct electricity very well, small amounts of impurity are introduced into digital image sensors to increase the silicon’s conductivity – a process called doping [395, 662, 714, 752]. When light enters the sensor, photons impinging on and penetrating into the silicon crystal can promote electrons to a
higher energy state as long as their energy level is higher than the so-called bandgap (i.e. the energy difference between the conduction band and the valence band) of silicon, which is approximately 1.12 eV at room temperature [395, 420, 662, 752]. This means that any photon of radiant energy till about 1127 nm (the so-called cutoff wavelength $\lambda_c$) has enough energy to promote an electron from the valence band to the conduction band in crystalline silicon, allowing for the photon-to-signal charge conversion to take place [569, 767]. This physical fact precludes normal digital sensors for detection of wavelengths beyond about 1100 nm, but implies digital cameras to have high **Quantum Efficiency (QE)** in the shorter wavelength-part of NIR [555]. Since four broad types of image sensors exist and different manufacturing processes are used, the general spectral response of digital image sensors tends to vary (see Figure 8.6 for the typical response of two monochrome sensors), although the spectral sensitivity and response increased to a very large extent when compared to photographic film.

![Spectral sensitivity characteristics of Sony’s ICX205AL CCD (adapted from [731]) and a Kodak CMOS Image Sensor (adapted from [258]: Fig. 1).](image)

**Figure 8.6.** Spectral sensitivity characteristics of Sony’s ICX205AL CCD (adapted from [731]) and a Kodak CMOS Image Sensor (adapted from [258]: Fig. 1).

### 8.4.2 Filters and Microlenses

Although it is the heart of every digital camera, it is not the intrinsic NIR response of the imaging sensor alone that determines the final spectral response of the DSC. Whatever the type, all image sensors consist of a 2D-photosite array to generate a digital photograph. The light-sensitive area, called **photodiode**, collects the photons during the exposure time [395, 420, 569, 752]. A digital camera which yields 12.2 million pixels has a sensor built-up by at least 12.2 photodiodes (e.g. 4288 columns x 2848 rows), as one photodiode generally contributes one effective image pixel. The percentage of the photosite that is sensitive to the incoming light is called the **fill factor (FF)**. Improving this FF is achieved by an array of on-chip **microlenses**, collimating incident photons to the photodiode [232, 395, 569, 752]. Below this array of small lenses, an on-chip **Colour Filter Array (CFA)** is positioned. This mosaic pattern of thin, coloured filters is positioned on top of all but the Foveon sensor (which captures all three
basic colours at every location – see Figure 7.9), each optical filter transmitting only specific wavelengths of the incident radiant energy [395, 569]. Every individual photosite of the array is covered by such a filter, allowing only one – but broad – particular range of wavelengths to be transmitted and subsequently captured by the photodiode (Figure 8.7).

Generally, small-format DSCs use a Red-Green-Blue (RGB) pattern with a repeating group of four photodiodes, in which two have Green filters and the remaining are either Red or Blue (Figure 8.7). This arrangement is called a Bayer pattern, after Bryce Bayer of the Eastman Kodak Company who patented it in 1976 [62]. The occurrence of twice as many Green filters as Blue or Red filters mimics the HVS, whose spatial resolving power for luminance information is larger than for chrominance data. As this critical luminance information is mostly carried by Green light, the higher occurrence of Green filters improves the spatial sampling of the luminance signal and hence the perceived sharpness of the digital image [62, 601].

The specific characteristics of both CFA and microlens array (e.g. their thickness and opacity to particular wavelengths) makes the final response to NIR (or to EM radiation in general) varying from camera to camera according to the used matrix, but it can be stated that DSCs are in general very sensitive to NIR radiation. This is confirmed by the examples shown in Figure 8.8A and B. The former depicts the response of the photodiodes covered with Red (R), Blue (B) and the two types of Green glasses (GRr and GBr) of the KAF-8300 Kodak CCD sensor, previously being used in the Olympus E-300 and E-500. In Figure 8.8B, the wavelength-dependent Quantum Efficiency (QE) for the Foveon X3 sensor is plotted as a function of wavelength. Note that this particular kind of CMOS sensor, which is incorporated in Sigma’s D-SLRs (SD9, SD10 and SD14) as well as in Polaroid’s x530 point-and-shoot camera [297], does not use a CFA. Because of its layered design, this imager measures every visible colour at every photosite [297, 502].
8.4.3 Low-Pass and NIR-Cut Filter

The spectral curves in Figure 8.8A and B are, however, not the result one would get when determining the spectral response of all three channels in the final manufactured DSC. To cut down the image-quality-degrading NIR and enhance the true colour rendition, camera manufacturers place an \textit{NIR-blocking/cutoff/cut filter} (also called \textit{hot mirror}) in front of the whole imaging array. By reflecting and/or absorbing all NIR radiation that passes the lens [128, 454, 651], these filters allow only photons from the intended part of the visible spectrum to be transmitted. The resulting spectral response range of the complete image array to incident EM radiation is given in Figure 8.9A and B, again showing the occurring varieties.

Although it depends on the quality of the NIR-blocking filter, most DSCs coming straight out of the box can still be used for NIR imaging, as most filters still transmit a few percentages NIR radiation, hence allowing a small portion of the incident NIR wavelengths to reach the image sensor. Besides a visibly opaque NIR filter in front of the lens (as in the film-based approach), an additional tripod is still needed, because the very low NIR sensitivity inevitable compels very long shutter speeds. This excludes the use of these cameras in research where accurately and fast composing as well as short shutter speeds is crucial (e.g. aerial photography).
8.4.4 Modification

Although different researchers already used off-the-shelf professional and/or consumer DSCs for pure or false-colour NIR imaging (e.g. [306, 544]), replacing the internal hot mirror of these cameras with a visibly opaque filter (Figure 8.10) makes NIR imaging much easier and more predictable, while hugely increasing the camera’s NIR sensitivity. In fact, the filter that is placed in front of the sensor can be chosen according to the specific needs to capture particular wavebands. Using a clear filter enables the whole spectral sensitivity range of the camera (from UV to NIR) to be used, while inserting an NIR-pass/cold filter will allow pure NIR imaging, as the final image will be formed by NIR radiation only. In the end, this operation boils down to moving the previously lens-fitted filter to the sensor, but with the additional advantages of large NIR responsivity (which is much larger and more extended compared to film) and the possibility to look through the camera’s viewfinder. However, one must take into consideration that clear (or other visible transmitting filters) still require an opaque filter to be fitted onto the lens in the case of pure NIR photography.

Replacing this internal IR-block filter is, however, a delicate job, as high voltages are present inside the camera and a lot of dust can stick to the sensor. To exclude the risk of damaging the camera (and losing the money spent as conversion voids the manufacturer’s warranty), it is advisable to have this modification performed by a dedicated company (e.g. Khromagery, LDP LLC, Life Pixel). Although these companies often modify only a few particular DSCs (mostly D-
SLRs), about any camera can have its NIR-block filter replaced. Finally, it remains important to mention that modification makes DSCs somehow more subject to moiré effects, as most cameras – certainly D-SLRs – combine the NIR-block filter with an optical anti-aliasing/low-pass filter, the latter slightly blurring the incoming signal as to generate less aliasing artefacts induced by the image sensor’s sampling. On the other hand, the removal of the low-pass filter results in pictures with increased sharpness in case no aliasing-prone subjects are shot (everything but meshes, fabrics, etc.).

Fujifilm already has discovered these inherent possibilities of digital cameras back in August 2006, when they released the FinePix S3 Pro UVIR, a modified version of the Finepix S3 Pro in which the sensor is only covered with a clear glass, allowing for UV (partly), visible, and NIR photography [309]. Besides not being sold in all countries, opaque filters are still required for pure NIR imaging, hence losing some of the modified camera’s major advantages such as automatic focusing and through-the-lens view (the latter partly resolved by incorporating live previewing on the rear LCD of the camera). Even though a full warranty comes with all their models (Fujifilm also released the hybrid IS-1 and D-SLR IS Pro in January and July 2007, respectively), the D-SLRs feature only full functionality as long as D and G type AF Nikkor lenses are used, making older lenses less and those from other brands almost completely unusable.

8.4.5 Lenses

Using a dedicated modified D-SLR, the same lenses used for the film approach can be applied, although it is still best to test whether hot spots are generated by a specific camera-lens combination [677, 678]. The chance of creating such a hot spot is, however, much larger when utilising unmodified cameras, as the hot mirror will reflect the radiation that needs to be captured. This NIR radiation is subsequently reflected into the back element of the lens, where the occurring internal reflections largely contribute to flare and the creation of the visually unattractive and scientifically unjustifiable hot spot.

8.5 Real-World Examples

To explore the capabilities of NIR photography in aerial archaeology, a Nikon D50 was acquired and subsequently converted (for an in-depth discussion, please consult Chapter 9, Chapter 10, and [776]). This modification proved very useful in the aerial experiments executed so far. Figure 8.11 shows two versions of the same scene: the eastern part of the Roman town of Trea (43° 18’ 40” N, 13° 18’ 42” W – WGS84), to be situated in central Adriatic Italy (Regione Marche).
Figure 8.11. Contrast enhancement (A2 and B2) on a visible (A1) and an NIR frame (B1) clearly shows the latter to reveal better Roman features as crop marks.

The upper left image (A1) is a pure visible record created by an unmodified Nikon D200, while the converted Nikon D50 – coupled with the D200 on one camera rig and using an identical focal length – generated photograph B1 at the very same moment. A2 and B2 again show the same images, but now after their central portions have been subjected to some image processing (histogram stretching and local contrast enhancement) to expose the crop marks more clearly. Directly comparing these two enhanced versions obviously shows the NIR record to reveal more crop marks. This effect can largely be attributed to the fact that diseased, senescent and heavily nutrient/water deficient vegetation is generally characterised by a large
decrease in NIR reflectance [145, 473, 559, 829]. Very often, this NIR decrease is far more noticeable with respect to the visible band, which witnesses a global increase in reflected EM radiation. Consequently, NIR imagery often has a larger potential to depict crop marks compared to conventional photographs (see also the next three Chapters).

Figure 8.12. A conventional (A1) and FCIR record (B1) of the central part of the Roman city of Trea. The enhanced versions (A2 and B2) point to the fact that FCIR can be advantageous in spotting crop marks, clearly illustrated by the walls of the Roman temple in the insets.

The second aerial example (Figure 8.12) reveals the possibility to recreate the appearance of a FCIR film by using a two-camera system based on a converted DSC coupled with an unmodified DSC, and combining the pure NIR image yielded by the former with the Red and Green channel created by the conventional camera. To perform the latter operation, setting the DSC to save a RAW file (instead of a JPEG – Joint Photographic Experts Group – or TIFF – Tagged Image File Format) is advisable, as RAW enables full control over the originally captured Digital Numbers (DNs), hence making it the only valid choice for scientific photography (see Chapter 7). In this example, the visible and NIR RAW files – again captured above the aforementioned Roman city of Trea – were converted in a completely linear way and without any White Balance (WB) applied, using a free RAW decoder written by David Coffin: dcraw [184]. Afterwards, the images were saved as 16-bit TIFFs and their colour channels split.
A pure NIR channel of the modified Nikon D50 was attributed to the Red channel of this new image, and merged with the Red and Green colour channels of the visible photograph (the latter two placed in, respectively, the Green and Blue channels of the new frame).

The result (Figure 8.12B1 and B2) is a digital FCIR image that is rendered the same way as in the analogue approach, but with a much better spectral fidelity. As can be seen in Figure 8.12A2 and B2 (which are the contrast enhanced versions of A1 and B1 respectively), most features in the conventional and FCIR image closely correspond. However, when looking more in detail at for example the Roman temple (see inset), some of the walls are more distinct in the FCIR, particularly the central division. Using such a recombination of channels thus often allows for a better identification of underground structures as roads, buildings, and ditches.

The haze penetrating capacity of NIR imaging is illustrated in Figure 8.13, displaying a very high oblique visible (A) and NIR photograph (B) from a mountainous area in central Adriatic Italy. The NIR wavelengths clearly suffer less from the atmospheric scattering, resulting in an image with a higher contrast and far more discernable detail in the distant parts of the scene (Figure 8.13B).

![Figure 8.13](image.png)  
*Figure 8.13. The haze penetrating capabilities of the longer NIR wavelengths (B) compared to the visible record (A) of a mountainous area.*

Besides aerial reconnaissance, other archaeological research fields benefit from this modified camera, as the subsequent examples shortly illustrate. By comparing a normal (Figure 8.14A)
with a pure NIR sherd image (Figure 8.14B), it can be noticed that the latter more clearly reveals the places where the clay has been exposed to increased reduction during the final phase of the firing process (1) or due to the incineration of organic material and/or decomposition of calcite inclusions (2).

Figure 8.14. Mutual comparison between visible (A) and NIR image (B) of a North African Late Punic/Early Roman bowl. (1) and (2) indicate zones of increased reduction (sherd by courtesy of K. Ryckbosch – Ghent University).

Figure 8.15 and Figure 8.16 both show how the appearance of certain, visibly attestable paint pigments can be altered in the NIR. However, whereas Figure 8.15B more clearly indicates the painted portions of this Roman handmade rim fragment compared to the visible photograph (Figure 8.15A), Figure 8.16 proves certain inks to become indiscernible if they exhibit the same reflectance properties in the NIR as the outer surface of the recipient. Although the inventory number is obvious in Figure 8.16A (i.e. the visible record), the same late Iron Age sherd seems to be unnumbered in the NIR record (Figure 8.16B). The other paint traces that cover this piece of ceramic are rendered with a similar contrast in both types of imagery.
Figure 8.15. Comparison between the visible (A) and NIR record (B) of a coated rim fragment belonging to a Roman handmade pot (sherd by courtesy of W. De Clercq – Ghent University).

Remark that, contrary to the other NIR pictures, the NIR sherds in Figure 8.15 and Figure 8.16 do not have a greyscale character. As illustrated in Figure 8.8, all colour channels capture NIR. Working on only one of these channels always yields a greyscale result, no matter if one captures exclusively visible radiation or pure NIR wavelengths. However, taking a modified DSC’s NIR information acquired in all three Red, Green and Blue channels into account will yield a photograph with an arbitrary coloured appearance, the latter being dependent on the particular NIR response of each channel and the colour channel it is attributed to in the final image. No matter its colour, the resulting frame still is a pure NIR record.

Figure 8.16. Difference between the visible (A) and NIR rendering (B) of the inventory number on a decorated wall fragment of a late Iron Age handmade pot (sherd by courtesy of W. De Clercq – Ghent University).
In the last example, the strength of NIR to digitally uncover obliterated writing is presented (Figure 8.17). As the upper black ink – which was used to mask the underwriting – is more NIR transparent than the ink applied for writing “original”, the latter becomes clearly visible in the digital NIR photograph (Figure 8.17B). Here, the extended spectral sensitivity of the digital approach largely helps as the contrast is maximised by acquiring NIR wavelengths in the upper 950 nm range [64, 158].

8.6 Conclusions

The combination of its error-prone and awkward workflow together with a lack of technical knowledge, a scant understanding of its potential, and unfamiliarity with its principles has often left (archaeological) NIR photography in the hands of only a few experienced and specialised photographers. Although this can and will not completely change in the beginning of the 21st century, the fact that today’s DSCs are perfectly capable of NIR imaging could only be beneficial to the increasing application of this kind of archaeological photography.

Using a modified, dedicated NIR DSC should deal with most of the earlier film related issues, even cutting down the cost hugely as there is no more need for tons of film, processing chemicals, and printing paper. Besides, geographic coordinates can be embedded into the metadata header, while the viewing capability enables a direct feedback and the digital format eases the storage and dissemination of information. As these solid-state devices have a larger Dynamic Range (DR) and QE [158, 373], they can also be used in far from optimal conditions, with more consistent results due to the omitted development stage, better exposure control and a linear response to incoming radiant intensity. Consequently, NIR photography “is at least worth trying more often than is generally done to unveil the past, for the results are fairly unpredictable” ([191]: 76).
Near-InfraRed Sensing of Vegetation Marks

From Historical Use to Digital Approaches
Abstract

Even though most archaeologists are aware of the crop mark phenomenon and its possible archaeological nature, the information on its occurrence and specific character is in most cases obtained by imaging in the visible spectrum. After the second World War, the occasional use of Near-InfraRed (NIR) sensitive emulsions attributed this kind of invisible imaging with a great potential. However, archaeological NIR imaging always remained restricted due to several reasons, not at least its complicated workflow and uncertain results. This chapter wants to delve deeper into the subject, looking at the conventional film-based approach of NIR aerial reconnaissance and its historical use in archaeological crop mark research, after which a current, straightforward digital approach will be outlined. By explaining the spectral properties of plants and using examples of recently acquired NIR imagery in comparison with visible frames, it should become clear why the detection and interpretation of crop marks can benefit from low-cost digital NIR imaging in certain situations.
9.1 Introduction

It is generally known to most archaeologists that sub-surface archaeological remains can reveal themselves as crop/plant marks, soil and shadow marks as well as less common snow, water or wind marks [834]. Although these marks are usually discovered and recorded with small- or medium-format hand-held photo cameras sensitive to visible light, some aerial archaeologists have been taking Near-InfraRed (NIR) sensitised media in their low-flying aeroplane to capture such archaeological signs utilizing this less conventional invisible waveband. Notwithstanding the response was often reported to be good – particularly for crop marks – the use of such beyond-visible techniques did never reach high peaks in archaeological reconnaissance. Before tackling this issue and looking at the historical and current use of the NIR spectrum, there is certainly the “need for a fuller understanding of the formation processes for cropmarks” ([706]: 181), in order to sensibly evaluate these (pre)historically related vegetation patterns and understand why archaeology could be concerned with acquiring aerial NIR data in the first place. Therefore, both the physical characteristics and corresponding spectral properties of healthy and stressed vegetation need to be understood on both leaf and canopy level, because the successful archaeological application, visualisation, and interpretation of optical remotely sensed vegetation data (more in particular visible and NIR information) is fundamentally based on this understanding.

9.2 Crop Marks and Related Plant Reflectance

Buried features can often be identified because they cause some anomalous growth in the plants that overlie them. Sub-surface archaeological remains such as pits or trenches will often be filled with organic material and/or new soil, having a greater moisture retention (and sometimes still more nutrients) than the surrounding matrix. In periods of drought, these humous soils have a positive, favourable effect on the crops, allowing the plants to grow more luxuriantly and for an extended period of time. These resulting vegetation patterns are commonly called positive crop marks. In unfavourable situations (e.g. plants growing over buried stone walls or floors), weaker and shorter plants might occur, in which case negative crop marks are yielded [27, 42, 280, 281, 430, 666, 667, 697, 834]. These plants’ divergent physiology and morphology – positive or negative – often allow them to be distinguished in the visible and/or NIR domain. Both visible and NIR radiation are small parts of the so-called ElectroMagnetic (EM) spectrum radiated by the sun and other sources. This EM spectrum is a very large continuum of varying waves, the latter characterised by a certain wavelength (\(\lambda\)), but all travelling at about 300 000 000 m/s [723, 804]. In general, the complete EM spectrum is subdivided into several wavebands, in order of increasing wavelength: gamma rays, X rays,
UltraViolet (UV) rays, visible light, InfraRed (IR) rays, microwaves, and radio waves. In fact, visible light covers only a very small EM range, situated between circa 400 nm and 750 nm. The radiation with slightly longer wavelengths is part of the NIR waveband (between 750 nm and 1400 nm), comprising the first, short wavelength part of the broad IR spectrum (750 nm to 1 mm). It remains important to note that NIR has nothing to do with heat imaging, which uses completely different portions of the IR waveband (more in particular, the Mid Wavelength InfraRed – MWIR – from 3 μm to 6 μm and Long Wavelength InfraRed – LWIR – from 6 μm to 15 μm; see also 8.1.2).

9.2.1 Healthy Green Vegetation

From the moment visible or NIR energy encounters a healthy green leaf, this radiation can be absorbed, reflected or transmitted, but only the reflected wavelengths will be used to create the aerial image. In the visible range of wavelengths, the photosynthetic pigments chlorophyll \(a\) and \(b\), both residing in the chloroplasts of mainly the long palisade parenchyma mesophyll cells of all green plants (Figure 9.1), absorb the incident radiation to a great extent [330, 383, 478]. By complementing each other, as much as 70 % to 90 % of the incident light might be absorbed, principally in the Blue (centred on 450 nm) and Red (around 670 nm) spectral regions [446, 642, 840]. The slightly decreased absorptivity of chlorophyll in the 500 nm to 600 nm zone makes the Human Visual System (HVS) perceive healthy leaves as green.

---

Figure 9.1. Structure of a plant leaf and its interaction with incident visible and NIR radiation (adapted from [727]: Fig. 32-3, and [748]: Fig. 2).
This behaviour changes completely in the NIR spectral band, where absorption by pigments is extremely low [169] and the leaf's internal cellular structure (more in particular the structure of the spongy parenchyma mesophyll which features a lot of intercellular air space – Figure 9.1) effects a very high and diffuse reflectance in the NIR spectrum [315, 316, 566, 616, 724, 840]. As soon as NIR radiation – to which the upper leaf layers are largely transparent – reaches the spongy mesophyll tissue, heavy scattering takes place at the numerous air-water-cell wall interfaces, due to the differences in refractive index [313, 315, 446, 724, 854]. Quantitatively spoken, NIR reflectance attains values of circa 40 % to 60 % in a healthy leaf. As only about 5 % to 10 % of this incident energy is absorbed internally (to counteract possible plant damage [313, 314], individual leaves still highly transmit NIR to the underlying leaves and ground.

This spectral behaviour is visualised in Figure 9.2. From this graph, it is obvious that healthy vegetation typically reflects low in the Blue band (circa 400 nm to 500 nm), more in the Green band (circa 500 nm to 600 nm), again less in the Red (circa 600 nm to 700 nm), after which it becomes very bright in the NIR. The very steep increase in the reflectance of radiant energy at the edge of the visible spectrum and the beginning of the NIR spectrum (between 700 nm and 750 nm), is the so-called Red edge [398]. It is the most prominent characteristic in the reflectance spectrum of healthy vegetation and one of the most extreme slopes to be found in reflectance spectra of natural materials [180]. Notwithstanding various crops will display slightly different reflectance curves due to their dissimilar cell structure and leaf inclination [331], these spectral features are typical of all mature and healthy green leaves [146, 343, 840].

Figure 9.2. Spectral reflectance characteristics of healthy green vegetation ([425]: Fig. 11-21a).
From a quantitative point of view, reflectance from the canopy might be seriously modified from that of an individual leaf, due to reflectance from non-green canopy components, the anisotropic behaviour of the canopy, the spectral properties of shadows and the variations in soil background reflectance, together determining the overall canopy reflectance [114, 187, 211, 212, 358, 567, 623]. Canopies with a high biomass or leaf area per unit of ground (called Leaf Area Index or LAI) can be characterised by a serious increase in NIR reflectance due to the process of leaf additive reflectance. However, the latter process levels off when the Vegetation Fraction (VF) exceeds 60 %, while a further VF increase can even generate a decrease of NIR canopy reflectance – a process which can be attributed to several reasons and generally occurs at or near the midseason [331].

The spectral contribution of the soil can, however, be negligible in case the canopy features a large number of healthy, mature leaf layers [211, 358, 425, 566]. These phenomena clearly show that individual, leaf-based reflectance spectra are not directly (and as a whole) transferable to the canopy level, although the best correlation generally is found within the NIR region [39, 52], with the strength of this expression being related to LAI and leaf angle [39].

9.2.2 Stressed and Senescent Vegetation

At a certain stage in its growth, the spectral properties of chlorophyllian vegetation considerably change due to senescence or stress – where stress is considered any (a)biotic factor affecting or blocking the growth, development or metabolism of the plant [479] –, a process attributed to a whole variety of anthropogenic and natural stressors. In such cases, the chlorophyll pigment rapidly decays [140, 149, 383, 446, 532, 623], additionally losing its absorption properties [143]. This results in an increased reflectance in the visible region, particularly in the Red chlorophyll absorption band, as the other pigments do not absorb strongly there [532, 845]. Moreover, chlorosis often occurs: a yellowing discoloration of the leaf due to the lost chlorophyll dominance over the carotenoids [13, 383].

In archaeological aerial reconnaissance, this changing spectral quality of the reflected light has always been used as a plant stress indicator. Hence, it is the stress-induced loss of chlorophyll that gives crop marks their altered colour, although the exact physical vegetation properties still depend on the type and amount of environmental stress as well as plant species and soil. Nevertheless, some archaeologists reported a clear relation between a crop mark’s first occurrence and the water content in the topsoil, soil moisture deficit being the chief cause of vegetation’s diminished growth vigour [204, 281, 429, 430, 813] besides factors as vegetation type, soil depth, and type of soil [281, 429, 430, 697].

In the NIR beyond 750 nm, the reflectance values are not related to concentrations of wavelength-dependent absorbing pigments, but might change because senescence as well as biotic (pests, weeds, disease, fungi) and abiotic constraints (drought, floods, high light intensity, chilling, mineral deficiency, ozone) alter the internal structure of the foliage and/or
the water content \([139, 277, 379, 414, 504, 717, 724]\). This means that any spectral change becoming apparent might give clues about the plant’s stress condition and/or senescence. Very often, healthy green leaves respond to short-term, acute (a)biotic stressors with increased NIR reflectance. This certainly holds when plants suffer from dehydration \([107, 139, 140, 358, 840]\) due to changes in direct light absorption by water governed by the water absorption coefficient \([210]\) and internal structural leaf changes: the leave’s increased intercellular air space originating from the dehydration induced collapse of mesophyll cells, generates larger quantities of air, water, and leaf material interfaces. As the cell wall-water interface has a lower angle of refraction than the air-cell wall and air-water interfaces, an increase in multiple reflections inside the leaf is observed \([20, 139, 147, 277, 294, 314, 446, 685]\).

In these early stages of stress, the NIR reflectance gradually changes, while the alteration in the visible region happens abruptly, but only at the time wilting is observed \([294]\). Hence, previsual symptoms might be observed by monitoring the NIR reflectance \([145]\). In some older research, stress damage was indeed reported to be discernable in these invisible wavelengths prior to any pigment-related visible clues \([17, 89, 190, 240, 382]\). However, these results often remain questionable as it is unlikely that such small alterations in the NIR reflectance could be sufficiently captured by the NIR-sensitive films (which is confirmed by several researchers who concluded that NIR imaging is not able to record previsible stress symptoms – e.g. \([759]\)).

The positive NIR results are most likely obtained in case water shortage is chronic and a complete vegetation canopy is considered, as these conditions make NIR reflectance drop significantly in the photographic NIR region \([382, 473, 550, 702, 703]\) (see Figure 9.3), partially related to lower LAI \([145, 473]\).

![Figure 9.3. Reflectance spectra of photosynthetic (green) vegetation, non-photosynthetic (dry) vegetation, and a soil (\([175\]): Fig. 1.18).](image)

Similar decreases in NIR reflectance spectra of plant canopies are to be found in diseased, senescent, and/or heavily nutrient-deficient vegetation \([145, 213, 457, 559, 622, 749, 829]\),
corresponding to a combined effect of internal tissue changes in single leaves, the general canopy geometry, and the additional low soil reflectance due to a decreased LAI [51, 145, 213, 327, 358]. Very often, these very large NIR decreases can reveal the plants’ state of vigour [190, 425, 814], with an absolute change in the NIR reflectance that is far more noticeable with respect to the visible band (Figure 9.3). Considering this, one can imagine that at a certain point, NIR wavelengths make it easier to discern negative crop marks: areas where plants are severely exposed to moisture shortage and/or nutrient deficiency. Conversely, growing plants will reflect more NIR radiant flux, which means that also positive crop marks – due to a large vegetation biomass [448] – might be distinguished in these invisible wavelengths.

9.3 Conventional NIR Aerial Archaeology

9.3.1 Research and Applications

Since its origin more than a hundred years ago, archaeological aerial reconnaissance has been imaging all kinds of marks almost exclusively in the visible part of the EM spectrum, using true colour or panchromatic photography. Although this visually-based data acquisition approach might be justifiable for most types of aforementioned archaeologically related anomalies, it should be strongly reviewed in the case of crop marks, as acquiring non-visible NIR wavelengths could sometimes be a better choice for revealing the vigour of crops. Although this idea is supported by older, mostly non-archaeological literature (e.g. [190, 559]) and more recent research claims the leaf reflectance response to acute plant stress in the visible spectrum to be more consistent [140, 141, 145, 146, 151], the aforementioned physiological and morphological state-related spectral differences in the NIR canopy reflectance clearly indicate that archaeological NIR photography might in some occasions be very fruitful in detecting and monitoring negative and positive crop marks (although it will be hardly possible to correlate changed NIR reflectance to a particular stressor, a problem also characteristic for visual imaging [140, 143, 168, 559].

Besides, NIR photography can also be an important archaeological tool because of its haze penetrating capacity [241, 668] (see also Figure 8.13). As archaeological aerial reconnaissance largely concerns high and low oblique photographs, parts of the image can become obscured due to severe atmospheric scattering in the visible region. NIR photography offers a solution here, yielding imagery with a high contrast ratio and enhanced clarity of detail [648], allowing the application of aerial photography even on early, hazy mornings [700]. Albeit this large potential in aerial reconnaissance, very little, and mostly unsystematic effort has been made to incorporate NIR radiation in archaeological aerial frames, even though NIR photography by means of Black-and-White (B&W) and False Colour InfraRed (FCIR or more commonly CIR) film proved in a lot of different areas its capability to detect archaeological traces.
John Hampton, from 1965 till 1985 the director of the Air Photographs Unit (APU) of the Royal
Commission on the Historic Monuments of England (RCHME), stated in the 1970s that normal
film emulsions and filter combinations can often hardly detect the very small colour and/or
height differences in crops [366]. At early stages of cereal growth and late in the season, he
found Kodak Ektachrome CIR film to be very useful in the detection of crop marks [366].

Besides hypothesising about its advantages [191, 364], other archaeologists also empirically
noted the – sometimes striking – advantage NIR imaging (whether using pure NIR or
conventional CIR photography, airborne multispectral approaches or satellite imagery) can
have on the detection of archaeological (crop) sites [17, 37, 43, 58, 109, 110, 131, 239, 240, 267,
although the objectiveness of some observations needs to be questioned, as a systematic and
elaborated assessment of the true archaeological potential was mostly impossible due to the
lack of simultaneously acquired normal colour photographs and reconnaissance in dissimilar
conditions. Nevertheless, these researchers indicated the potential of monitoring field crops
through the acquisition of reflected NIR radiation. Consequently, it might seem strange that
the archaeological potential of pure or false colour NIR photography was never more widely
explored and generally practised. However, several possible explanations can be identified.

9.3.2 Drawbacks

Although some lack of knowledge about the subject certainly played an important part, the
fact that some scholars – of whom the majority happened to have captured comparative
frames in the visible spectrum – stressed that CIR imaging is not all-revealing (e.g. [205, 247,
354, 366, 376, 413, 518, 744]), most likely made the extremely error-prone film-based NIR
workflow too much of an obstacle to be applied more often: both reversal and transparency
film needed to be stored cooled and developed by specialised labs directly after exposing
them [259, 262]; moreover, determining the right exposure was not as straightforward as with
conventional photography [123], and any inconsistency in emulsion or film development
yielded different results [73].

Being sensitive to about 900 nm, NIR-sensitive film also had a narrow exposure latitude and a
rather small Dynamic Range (DR) [325], whereas its relatively weak sensitivity made it
unusable with poor lighting conditions [89, 92, 116, 366]. Finally, the spectral fidelity of CIR film
was rather low, as the NIR-sensitive band also took large portions of the visible spectrum into
account (infra). Luckily, the digital (r)evolution of photography supplied new tools and
methods to deal with most of these issues.
9.4 Digital NIR Archaeological Reconnaissance

9.4.1 Camera Modification

Since the advent of the 21st century, archaeology has been witnessing an ever increasing role of and significant interest for airborne digital photography [238, 245], as do other disciplines [619]. The advantages are clear: taking more photographs at a lower cost, without the need of loading and unloading film and scanning the resulting negatives/positives; the storage and retrieval is much easier and the resolving power of current Digital Still Cameras (DSCs) is equal or even surpassing the 135 (or 35 mm) format. Besides, such a photo camera or DSC – to be defined as a camera equipped with both a digital image sensor for capturing photographs and a storage device for saving the obtained image signals in a digital way [760] – offers another big advantage: its digital imaging sensor is essentially made of silicon (Si, atomic number 14) and very sensitive to NIR wavelengths till about 1100 nm [256, 569, 767].

To cut out the image degrading effect of this invisible radiation and only allow visible light to create the digital photograph, camera manufacturers place an NIR-blocking/cutoff/cut filter (also called hot-mirror) in front of the sensor, which absorbs and/or reflects most NIR waves that pass the lens [128, 454, 651]. Replacing this filter with an optical element that completely blocks all visible radiation, makes the DSC only responsive to incoming NIR wavelengths (for a technical outline of digital imaging sensors and a detailed overview of this DSC modification, the reader is kindly asked to consult Chapter 8).

The resulting spectral response of such a modified DSC is given in Figure 9.4A, showing a DSC’s sensitivity to NIR. Moreover, such a modification still allows to view through the lens, something that was impossible in the film-based approach of pure NIR imaging (because the visibly opaque filter was mounted in front of the lens). Besides the sensitivity to NIR wavelengths, Figure 9.4A illustrates the channel-specific spectral response, due to the construction of the complete imaging array. The latter comprises also a Colour Filter Array (CFA) between the sensor and the block filter. This mosaic pattern of coloured filters (Figure 9.4B) is the most widespread method to give colour sensitivity to image sensors. Every individual photodiode (i.e. a very small radiation-sensitive area of the sensor which creates one pixel of the final digital image) is covered by a specific coloured filter, allowing only one particular range of EM radiation to be collected. Besides transmitting the Red, Green, and Blue visible waveband, all three Red, Green, and Blue filters are clearly characterised by a specific NIR transmittance, generating a waveband-specific charge in the photodiode.
9.4.2 Benefits

Using NIR-modified DSCs deals with most of the aforementioned film issues, while the inherent properties of digital imaging as well as modification-specific characteristics offer additional benefits. First of all, a DSC’s linear response to incoming radiant intensity, its larger DR as well as the direct feedback in casu accurate exposure and focusing, mean that more consistent results can be generated when compared to a film-based solution. Secondly, the substantial NIR sensitivity allows their use in operational conditions far from optimal, while the possibility to embed geographic coordinates into the photograph’s metadata header should also not be underestimated.

Furthermore, the fact that all three spectral channels are created by taking the reflected EM radiation of a very specific waveband into account, allows for very useful arithmetic channel operations (infra and also Chapter 10). Finally, DSCs are very suited for mapping purposes, as they do not suffer from geometric film distortions [441, 442]. Despite these advantages, the use of NIR-modified DSCs was never fully explored in archaeological aerial reconnaissance.

9.5 NIR Aerial Crop Mark Archaeology – Applications and Examples

To indicate some possible situations where NIR imaging might be very fruitful, some real-world examples are presented. The imagery shown was generated in the framework of the Potenza Valley Survey (PVS), a geoarchaeological research project conducted by the Archaeology and Geography Departments of Ghent University (Belgium) in the Regione Marche, one of the contemporary Italian regions situated on the central Adriatic part of the country. To study the urban and rural occupation patterns in the Potenza river valley through (Pre)history, aerial
archaeology has been playing a very substantial role from the very beginning of the project (for more information, please consult Part I of this thesis or [780, 783, 786, 787, 788, 790, 792, 794, 795, 796, 797, 798]), even though the possible flying time has always been limited due to the practical reason of working abroad.

Till now, three NIR flights have been performed, respectively at the end of April 2007 (23rd), the 15th of May 2008, and the beginning (3rd) of July 2008. In the latter two, perfect simultaneous visible and NIR image acquisition was made possible by the use of a two-DSC rig, whereas the 2007 reconnaissance flight only yielded approximately equal imagery, as both authors photographed individually from the same plane. All three reconnaissance flights were performed around noon, previously being revealed as the optimum period for NIR photography [750]. Besides, NIR imagery has been generated from a Radio Controlled (R/C) low altitude aerial platform based on a Helikite (a combination of a helium balloon with kite wings), allowing for very targeted, and large scale (1/500 till 1/5000) aerial imaging ([781] or Chapter 5). Because the platform currently only allows to mount one DSC, the imagery generated by this apparatus will not be used in the following comparison.

For reasons explained previously [776], most of the presented NIR imagery was captured by a modified Nikon D50, a commercially available Digital Single-Lens Reflex (D-SLR) camera that is hereafter called D50_{NIR}, and whose spectral characterisation (Figure 9.4A) is described in the subsequent Chapter. Only during the July flight of 2008, a Full Spectrum-modified Nikon D80 was taken aloft (denoted D80_{FS}, see 5.3 or 12.3.1 for an explanation), as it generates digital frames with a higher pixel count and accepts larger memory cards. To acquire pure NIR images, the Nikon D80_{FS} was equipped with a Hoya R72 filter.

Before getting airborne, an AF Nikkor 50 mm f/1.8D lens, of which the focus was fixed on infinity, is mounted on both the normal as well as NIR-enabled DSC. Afterwards, the White Balance (WB) of the modified DSC is user-defined by pointing it to a patch of green grass. This operation normalises the dissimilar NIR response of the three channels (see 7.5.1 for an expalantion), as the highly NIR-reflective grass can be considered an NIR-neutral target. To make sure only pure NIR imagery was used for the following comparison, the Blue channels were extracted from both the D50_{NIR} and D80_{FS} imagery, as the Red and Green channels still take small portions of the Red edge region into account (see Figure 9.4A and Chapter 10) – although it needs to be stressed that in normal, daily working practice, the complete RGB image is used because the output is almost identical.

All conventional visible colour photographs were generated by a Nikon D200 or Canon EOS 300D. To compare the two data sets by objective means, both the visible and pure NIR frames were subjected to histogram stretching to optimise the images’ global contrast.
Near-InfraRed Sensing of Vegetation Marks

Figure 9.5. General comparison between the appearance of common ground features in the visible domain (A) and NIR spectrum (B): (1) healthy green vegetation, (2) soil, (3) low density canopy, (4) water, (5) trees, and (6) distant objects. Images acquired in the middle Potenza valley (N 43° 18' 02", E 13° 16' 12" – WGS84) with a Nikon D200 (A) and Nikon D50\textsubscript{NIR} (B) on May 15, 2008 at 11.40 h.

9.5.1 Pure NIR Channel

Figure 9.5 – acquired above the central Potenza valley – gives an overview of a visible frame’s (A) altered appearance in the NIR (B) and illustrates the typical behaviour of ground features. As healthy, green vegetation (1) reflects huge amounts of NIR, it appears very bright in Figure 9.5B, while soils (2) generally reflect much lower amounts of NIR (cf. Figure 9.3). Consequently, they are rendered much darker. When both are combined (3, i.e. a soil with a scarce vegetation cover), an intermediate greyscale value results. Water (4), on the other hand, always appears blackish on an NIR photograph, due to its very high absorption of these invisible wavelengths [210].

The potential of NIR to discriminate between plants on the basis of their species and/or health status, is revealed by the trees (5) at the bottom of the frame, showing a much larger variation in brightness compared to the visible record. Contrary to the visible record where stress is visualised by an increase in brightness (e.g. yellow crop marks), NIR vegetation stress patterns will be darker compared to the surrounding non-stressed plants. Finally, the clarity of detail in distant scenes is also visible to a certain extent. Due to the longer wavelengths, atmospheric scattering is much less in the NIR region than it is in the visible part, allowing NIR radiation to largely penetrate haze [191, 672], yielding imagery with a larger contrast [648].
When focusing more in detail on crop marks, several observations can be made. Figure 9.6, depicting the central part of the Roman coastal colony Potentia and acquired in the middle of May, clearly confirms what was already remarked by [282]: chlorotic vegetation is very difficult to distinguish in the NIR (compare visible frame A with NIR frame B). At moments this stress related loss of chlorophyll pigment is rather moderate, the yellow discoloration of vegetation is not extremely pronounced, but the alteration of the NIR reflectance curve can even be smaller. Hence, the very common negative crop marks are often better discernable in the visible domain, an observation also made by Hampton [366].

Pushing the local contrast in both frames (Figure 9.6A and B) to the limits (Figure 9.6C and D) does not alter this observation: even though the resulting tonal differences clearly prove the spectral NIR responses of chlorotic and healthy green vegetation to be dissimilar (tile D), the amount of features perceptible as well as the distinctness of the archaeological traces still remain superior in the visible spectrum (Figure 9.6C).
On the other hand, pure NIR channels can clearly reveal the more severe drought and nutrient stress in the canopy reflectance [248]. Imagery taken during the hot summer of 2008, again above the Roman city of Potentia, clearly illustrates this physical fact. In the grain field on the right side of Figure 9.7A, little pieces of the street network in the southern part of the city are visible. The NIR record (Figure 9.7B), however, allows much more traces to be seen, whereas the features are also more distinct when compared to the visible record. Because it is ripe, the overall visible reflectance response of the grain is higher and the HVS perceives these crops as yellow-brown.

Therefore, the reflectance increase of grain growing over a rocky sub-surface can only be slightly larger in comparison with the adjacent crops. In the NIR, the global reflectance decreases seriously at this stage of the crop cycle, but Figure 9.7B proves the total reduction of NIR reflectance to be noticeably larger for the extremely stressed plants. After additional contrast enhancement had been performed, the traces that could be mapped were indicated on the right of Figure 9.7A and Figure 9.7B. These results nicely confirm the previous statements of Hampton [366], who also empirically attested NIR imaging to be superior for cereal crop mark imaging late in the season.

Besides pigment variations, smaller plants and a less dense vegetation canopy also can be possible responses to subsoil variations. Figure 9.8 presents a last example of negative crop
marks, showing that these particular canopy situations with a low LAI can benefit from an NIR approach as well. It is evident that both the visible and NIR image of the vegetation canopy uncover hidden ancient hydrographical features, although the most distinct result is yielded by the invisible wavelengths: the low biomass density and related high contribution of low background reflectance produce very explicit, dark traces in the field (e.g. the traces on the left and upper right part of the field).

Figure 9.8. Hydrographical features revealed by negative crop marks in the visible (A) and NIR (B) part of the spectrum. Imagery was acquired in the central part of the Potenza Valley (N 43° 19' 22", E 13° 25' 26" – WGS84) with a Nikon D200 (A) and a Nikon D80FS (B) on July 03, 2008 at 11.23 h.

In addition to these negative crop marks, also positive crop marks might be distinguished in the NIR, as these features are characterised by a larger vegetation biomass and, as a consequence, will reflect more NIR-radiant energy. This principle is illustrated in Figure 9.9, showing a positive, non-archaeological crop mark in the centre of the Roman town of Ricina. The higher and denser amount of plants is perceived in the visible domain as two square, slightly darker green patches (1 and 2 in Figure 9.9A).

In the NIR (tile B), the same squares are brighter, as the larger quantity of plant tissue effects a higher reflectance. When mutually comparing both frames, the magnitude of reflectance dissimilarity is largely equal. Increasing the local contrast to a much larger extent clearly shows the greyscale NIR frame (Figure 9.9D) to have the advantage over the normal colour image (Figure 9.9C).
Near-InfraRed Sensing of Vegetation Marks

Figure 9.9. (A) Visible image of the central part of the Roman town of Ricina (N 43° 19' 41", E 13° 25' 26" – WGS84); (B) NIR image of the same scene. To obtain versions (C) and (D), very extreme local contrast enhancement was applied to (A) and (B) respectively. Images acquired with a Nikon D200 (A) and a Nikon D50NIR (B) on May 15, 2008 at 11.27 h.

Although the above examples clearly indicate the spectral response in the NIR to be unrelated to chlorophyll pigment concentration, the most striking example is given in Figure 9.10, which presents two different photographs of the central part of the Roman town Trea. Figure 9.10A again depicts the visible bands, together with the D50NIR image (9.10B) captured in the middle of April, 2007. Whereas the visible image only shows strange, non-archaeological positive vegetation marks (maybe due to differential manuring or a dissimilar kind of crop), the NIR image – being insensitive for these pigment variations – clearly allows to perceive the outlines of a Roman temple.

Besides the masking effect of the chlorophyll pigment in the strange darker vegetation zones, this striking difference might also be attributed to the anisotropic behaviour of the vegetation canopy. Being a non-Lambertian surface, the reflectance of the vegetation canopy is not equal in all directions, but dependent upon the sun and sensor zenith/off-nadir and azimuth angle [439]. As the canopy reflectance and its anisotropy will generally be higher when the sensor records back-scattered energy [680], the backward scattering component of the incident visible wavelengths – which is imaged in this situation – might be responsible for this lack of archaeological evidence in Figure 9.10A, as reflectance anisotropy reaches a maximum at visible wavelengths. NIR radiation is relatively free of such Bidirectional Reflectance Distributions Function (BRDF) effects [358, 680] due to the multiple canopy scattering of NIR photons [485]. The phenomenon of crop marks being well visible from one specific flying direction is widely known among aerial archaeologists. Concerning this, the less critical
angular view characteristic for the digital NIR approach might prove very useful, certainly in cases one would opt to fly a vertical coverage, trying to detect archaeological features afterwards.

Figure 9.10. A visible (A) and pure NIR image (B) of Roman Trea’s central part (N 43° 18’ 40”, E 13° 18’ 42” – WGS84). The insets are high contrast enlargements of the same area around the left pylon. Images acquired with a Canon EOS 300D (A; photograph by F. Vermeulen – Ghent University), and a Nikon D50\textsubscript{NIR} (B) on April 23, 2007 at 11.01 h.

9.5.2 NIR-Based Vegetation Indices

Capturing NIR by a modified DSC often allows to perform some arithmetic channel operations, as the channels generally acquire dissimilar wavebands of reflected EM radiation. This was illustrated by Figure 9.4, which displays the channel dependent spectral response of the Nikon D50\textsubscript{NIR}. As described in Chapter 10, these characteristic, unequal spectral responses of the D50\textsubscript{NIR} allow for the calculation of a Simple Ratio (SR): a Vegetation Index (VI) that has a large potential to indicate zones with a large amount of green biomass [431, 623, 687] and computed by dividing a pure NIR waveband by a part of the Red spectrum. Applied to a frame of the D50\textsubscript{NIR}, the following simple arithmetic operation can be executed:

\[
F(i,j) = \frac{[R(i,j) - G(i,j)]}{[G(i,j) + B(i,j)]}
\]

in which \(F(i,j)\) is the final pixel and R, G, and B indicate the value of this pixel in the Red, Green, and Blue channels of the NIR image respectively. The output equals a new array of greyscale pixels, of which the intensities represent the amount of chlorophyll pigment present at that particular spot. Figure 9.11 illustrates this principle. On the left (Figure 9.11A), the visible frame indicates some negative, likely geomorphologically related, crop marks. The same traces are only weakly discernable in the NIR (Figure 9.11B) for reasons explained previously. The plot on the lower right of the image is once more a very striking example of NIR’s reflectance.
insensitivity to pigment content: although the red flowers are clearly distinguishable in tile A, the NIR record does not give a single clue about their presence. The SR image (Figure 9.11C), however, indicates that Figure 9.11B has all information embedded to clearly indicate the chlorophyll dense areas: houses and streets are completely white, the red flowers and negative crop marks have some middle grey value, whereas the dark areas correspond to the healthy green zones in the visible picture A. Even though this approach does not always work (see Chapter 10), there is a large chance for this SR to add relevant vegetation information to the spectral signals stored in the NIR photograph.

However, the above approach should be followed in case only one DSC can be taken aloft. In the PVS, it became a common approach (since 2008) to take a two-DSC system – based on an unmodified Nikon DSC and a Nikon D50$_{\text{NIR}}$ or D80$_{\text{FS}}$ – in the aeroplane, a tandem which is rather simple to construct and perfect for hand-held operation [254]. Besides a more rigid and consistent application of the SR – which is also known as the Ratio Vegetation Index (RVI) or Vegetation Index Number (VIN) –, this combination allows additional VIs to be calculated, thus dealing with situations in which the D50$_{\text{NIR}}$-based SR does not prove very useful (e.g. differentiating between moderate and very high chlorophyll content or LAI [331]).

As an example, Figure 9.12 visualises the differences between the SR calculated by using the D50$_{\text{NIR}}$ only (A), while Figure 9.12B displays the output of the D200’s Red channel after division by the D50$_{\text{NIR}}$’s Blue channel. It is obvious that the second approach yields a superior result, due to the far better spectral placement of the Red band (as also discussed in 10.5) – although one must not forget that imagery from two different DSCs has to be coregistered (i.e. the process of geometrically aligning two or more images to allow them to be superimposed) before any mathematical operation can be performed. The latter is, of course, not needed when applying a single DSC.
Figure 9.12. (A) The output of a SR operation, using only the image channels of a Nikon D50\textsubscript{nir}; (B) this SR output was generated by utilizing the D50\textsubscript{nir}'s Blue channel and the visible Red band acquired by a simultaneously operated Nikon D200. The image shows the street pattern in the northern part of Potentia and was taken on May 15, 2008 at 11.16 h.

9.5.3 CIR Film Emulation

Apart from the problems of aligning the optical axis and firing the shutters at the exact same time, creating – and flying – such a two-DSC system is easy and exploits the spectral characteristics of an NIR-modified DSC more fully, offering archaeologists a very affordable and manageable multispectral tool that suits both visual interpretation and mathematical spectral operations, aiding in the revealing of the archaeological subsurface, just by capturing whole parts of the landscape in several spectral bands with the ease of one single exposure.

In the ideal situation, several visible and NIR bands could be sampled by one DSC, hence also omitting the otherwise necessary coregistration step. Some years ago, such an approach was tried by the Eastman Kodak Company with their now discontinued Kodak DCS-200 CIR, DCS-420 CIR and DCS-460 CIR [348]. Although a number of complications clearly proved these models to be initially built for non-scientific work [711], Figure 9.13A and B show that their spectral response approximately matched Kodak’s NIR-sensitive emulsions [91, 347]. As can be seen, both the film and digital approach acquired spectral information in very broad wavebands, hence taking significant portions of the other spectral bands into account [222, 348, 633, 742].

From the response of the Kodak DSC-420CIR (Figure 9.13B), it is obvious that the Red sensitive diodes (and to a lesser extent also the Green channel) take large portions of the NIR into account, a drawback Kodak counteracted by subtracting the NIR’s Digital Numbers (DNs) from the initially captured Green and Red DNs [348]. However, this approach could never yield pure spectral information: because the spectral response curves do not even coincide on the long wavelength side (Figure 9.13B), it is impossible to remove the precise NIR-contributing part of the Green and Red channel.

Moreover, the NIR and Red channel still took a significant portion of the Red edge region into account, which made these broad band imagers less suitable for quantitative spectral analysis.
as they masked unique spectral features to a large extent. Because also the analogue media were characterised by a low spectral fidelity (Figure 9.13A – e.g. consider the amount of visible radiation the NIR-sensitive band takes into account), Kodak’s NIR-enabled DSCs emulated the Kodak CIR film rather well, and was therefore often used in several vegetation studies (e.g. [91, 92, 222, 346, 442, 445, 585, 633, 742]).

Figure 9.13. (A) Spectral sensitivity curves of the Kodak Ektachrome Professional Infrared EIR film (adapted from [259]); (B) spectral response of the Kodak DCS-420 CIR (adapted from [348]: Fig. 5.7); (C) the Green and Red channel response of the Nikon D200 (adapted from [712]: Fig. 6B) with the spectral sensitivity curve of the Nikon D50

\textsubscript{NIR}’s Blue channel. The grey zone indicates the NIR domain.
A two-camera system based on a Nikon D50\textsubscript{NIR} coupled with an unmodified DSC can, however, deal with these issues. Figure 9.13C illustrates the Blue, pure NIR channel response from the D50\textsubscript{NIR} normalised to the sensitivity of the D200’s Green and Red channel. The spectral fidelity of the acquired information is, obviously, superior to the CIR film, and Kodak’s NIR-enabled DSCs. So, given the fact CIR film proved in a lot of different study areas its capability to detect the effects of physiological plant changes (e.g. [282, 748, 754]), means that CIR imagery generated by less broad bands certainly will yield similar, if not largely superior and less debatable results (the latter most likely also one of the reasons for its sparse archaeological application). Moreover, this approach is a cheap solution, cost being the main reason Kodak had to discontinue their models in the late 1990s [348]. The essential coregistration step can be considered a major drawback, slowing down the visualisation of the final output (Figure 9.14B). However, several software packages for (semi-)automatic coregistration exist, meaning this disadvantage can be solved in a rather straightforward way.

A possible solution to capture CIR imagery with only one DSC is presented in the United States Patent 20060066738, in which Hershey and Zhang proposed a CFA pattern consisting of four different coloured filters, three passing the Blue, Green and Red visible bands, while a fourth is dedicated to transmit pure NIR or UV radiation [384]. It is, however, highly questionable whether any company will ever market such a device, in which case a multispectral imaging system utilizing a multitude of DSCs would still offer a higher resolving power and more possibilities in choosing spectral combinations. The latter property is of crucial importance, as it allows the execution of mathematical spectral operations to indicate a number of archaeologically important vegetation parameters (e.g. canopy cover, biomass, chlorophyll and water content, etc.) better than can be revealed by either NIR or Red alone [500, 661]. This clearly proves the highly complementary nature of NIR and visible reflectance data.

![Figure 9.14. Visible record (A) and CIR output (B) of the street pattern in the northern part of Potentia. The images were taken on May 15, 2008 at 11.16 h and generated by a Nikon D200 and a simultaneously operated Nikon D50\textsubscript{NIR}.](image)

Finally, it remains to be seen whether or not digital CIR still has to be considered beneficial (let alone essential) for archaeological purposes. Even though the substantial lack of simultaneously captured CIR (or pure NIR) and visible information made it previously very difficult to correctly assess the added value of such false-colour composites in aerial crop mark
archaeology, Figure 9.14B illustrates that the output does not always yield better results compared to the visible frame (as also noticed by [205, 354, 366, 376, 413, 518, 744]). Moreover, the digital way of working allows the individual examination of each captured spectral channel, whereas the calculation of specific VIs enhances the dissimilarities between these channels far more than a CIR image is able to do. From this point of view, merging several spectral channels into one false-colour photograph appears not to be essential anymore, except for creating a visually pleasing image. Because the analysis of the currently acquired data set confirms this suggestion, it seems more and more obvious that the creation of CIR photographs no longer offers direct archaeological benefits over the abovementioned techniques in revealing information on the vegetation’s physiological and morphological conditions (cf. Figure 9.14A versus Figure 9.14B).

9.6 Conclusion

By using the reflected portion of incident EM radiation, the spatial perspective offered by aerial photography allows to remotely assess the vegetation status over both small and larger areas [73, 571]. Even though accurate and cost-efficient monitoring and mapping of morphological and/or physiological crop changes is essential to study our hidden past, the slight differences of colour and height in crops are often very difficult to record through normal colour photography. Therefore, it might seem striking that relatively few aerial archaeologists have been exploring the limits of the HVS, trying to incorporate NIR reflection in their photographs. This phenomenon can, however, largely be attributed to the error-prone, awkward film-based workflow of pure NIR or CIR imaging, some unfamiliarity with its principles and the fact that such a beyond-visible approach did not always prove successful or even useful – although a thorough assessment of the full archaeological potential has in most cases been problematic due to the usual lack of simultaneously acquired and geographically extended comparison material from the visible spectrum.

Flying with two simultaneously operated DSCs, however, proves to be the way forward. Besides capturing the conventional and familiar visible radiation, a modified DSC can be triggered to record the pure NIR reflectance of the same scene. Due to practical reasons, it has not been possible yet to perform some very systematic testing of such an NIR-visible system’s archaeological capacities (e.g. there was even no time for ground observations to provide the necessary crop identification). Nevertheless, after flying only a few times in a geographically rather restricted area, it goes without saying that the archaeological potential of a modified, NIR-enabled DSC can not be underestimated. Both the use of individual spectral channels, the arithmetic operations performed on a combination of channels (e.g. the calculation of a SR) as well as the combination with simultaneously acquired visible data offer a lot of opportunities to visually enhance archaeologically related anomalies and/or even reveal completely new features. Although archaeological NIR aerial imaging is by no means novel, the advantages modified DSCs can offer in the generation and interpretation of reconnaissance information is
substantial: in addition to simplifying the complete workflow, the possibilities known from the film-based pure NIR or CIR approach are expanded and perfected, without the significant costs of the latter.

After flying several years and quite a few times in dissimilar geographical regions over various sites and in different climatic conditions – all in an attempt to capture growth marks with dissimilar characteristics –, a complementary NIR-visible approach should permit to get a more complete, and conclusive picture of the operational conditions of NIR, indicating the amount to which certain features can become more, less or equally visible in this invisible domain. Obviously, NIR imaging is not solely restricted to crop mark archaeology: the existing soil moisture differences that are characteristic for soil marks [42, 430] can benefit from an NIR approach as well (e.g. [413, 503, 668]).

This way, modified DSCs can be considered low-cost, convenient possibilities to yield additional, beyond-visible archaeological information that supports and speeds-up the analysis of simultaneously captured visible spectral bands, while also aiding in the interpretations of existing data sources collected in previous years with different means – without making these other methods of data acquisition obsolete. As such, every single new piece of information, regardless the instrument it was acquired with, can contribute to improve our understanding of (pre)historic landscapes. Therefore “it is extracting as much archaeological information from as many sources as possible that is the challenge for the coming century” ([81]: 286).
Spectral Characterisation to Enhance Observation

Exploiting the Response Curves of an NIR-Modified Camera to Enable Biomass Visualisation

Abstract

Scholars using still cameras to take (mostly) oblique imagery from a low-flying aircraft of possible archaeologically related anomalies can be defined as aerial archaeologists. At present as well as in the past, aerial/air archaeology has been acquiring data almost exclusively in the visible range of the ElectroMagnetic (EM) spectrum. This phenomenon can largely be attributed to the critical imaging process and sometimes unconvincing results related to the film-based approach of Near-InfraRed (NIR) photography. To overcome the constraints of detecting and interpreting only the varying visible colours in vegetation (so-called crop marks), while still maintaining the flexible and low-cost approach characteristic for archaeological aerial reconnaissance, a consumer Digital Still Camera (DSC) was modified to capture NIR radiation. By its spectral characterisation, more insight was gained into its imaging properties and necessary guidelines for data processing and future improvements could be formulated, all in an attempt to better capture the archaeologically induced anomalous growth stresses in crops.
10.1 Introduction

10.1.1 Aerial Archaeology

The term ‘aerial archaeology’ encompasses the entire process from the acquisition and inventoring of imagery, to the mapping and the final interpretation. It comprises the whole study of all sorts of archaeological remains by using information acquired from a certain altitude: digital or film-based aerial photographs, satellite imagery, LiDAR, RADAR, etc. The majority of source data used by most aerial archaeologists is acquired from the cabin of a low-flying aeroplane using small- or medium-format hand-held cameras with (generally) uncalibrated lenses, capturing mostly oblique imagery. Although this specific type of data acquisition may seem strange to the non-archaeological community, the non-invasive approach yields easily interpretable imagery with abundant spatial detail, is extremely flexible, might be cost-efficient (certainly when compared to other prospecting methods and applied in previously unexplored areas), and is driven by the specific nature of the archaeological anomalies (basic literature on the history, techniques, uses and interpretation of aerial archaeology and its generated data includes [82, 117, 132, 199, 204, 229, 237, 542, 595, 666, 700, 734, 834]).

Archaeological remains such as settlements, graveyards and roads can show up on the surface in a number of ways. Aside from still standing material relics (e.g. churches, bridges, fortifications, etc.) and partly eroded structures (earthen banks, mounds, ditches, etc.), most of the features that can be viewed from above are the remains of buried archaeological sites. While the first type of archaeological features is directly visible, the second type – often referred to as earthworks – is mostly recorded from the air when thrown into relief by low slanting sunlight (sometimes referred to as shadow marks), and in northern Europe by differential snow accumulations or differential melting of snow or frost.

The buried or levelled remains might be disclosed by distinct tonal differences in the (usually ploughed) soil (soil marks) or differences in colour and/or height of vegetation on top of the remains (crop/plant marks), variations in the subsoil being the prime movers in their creation. In other words, archaeological residues must exhibit a certain localised contrast in their surrounding matrix in order to be detected [65]. Although these marks are mostly discovered, photographed and mapped using visible light, this chapter will explore how these anomalies, in particular crop marks, can benefit from detection and interpretation by low-cost digital aerial imaging of Near-InfraRed (NIR) radiation. Consequently, the nature of crop marks needs to be considered first.
10.1.2 Crop Marks and Related Plant Reflectance

Sub-surface archaeological remains such as pits or trenches will often be filled with organic material and/or new soil, which has a greater moisture retention than the surrounding matrix. In periods of drought, these soils might have a favourable effect on the crops, allowing the plants to grow luxuriantly and for an extended period of time. The adjacent plants will be less tall, thinner and ripen quicker, leading to differences in chroma and/or plant size that can be seen from above as positive crop marks (Figure 10.1A).

![Positive crop mark](image1.png)

![Negative crop mark](image2.png)

Figure 10.1. Positive (A) and negative (B) crop marks (adapted from [42]: Fig. 13).

In unfavourable situations (e.g. plants growing over buried stone walls or floors – Figure 10.1B), weaker and shorter plants might occur, in which case negative crop marks are yielded [27, 42, 280, 281, 430, 666, 667, 834]. Speaking in more technical terms, such adverse situations put a certain stress on the vegetation, hence blocking the growth, development or metabolism of the plant. It is the stress related loss of chlorophyll – a green pigment that can be found in all green plants and largely absorbs incident visible wavelengths in the Blue waveband (centred around 450 nm) and Red (around 670 nm) spectral region [446, 642, 840] – which induces an increased visible reflectance in the Green-Yellow-Orange-Red waveband [532, 845]. Consequently, the plant’s dominant green colour disappears in favour of a yellowing discoloration, a phenomenon called chlorosis [13, 338, 383]. By recording the reflected portion of the visible radiation, aerial photographs thus allow the remote assessment of vegetation status [73].

However, aerial archaeologists have sometimes acquired imagery using other parts of the EM spectrum (Figure 10.2), in particular the Near-InfraRed (NIR) waveband (see Chapter 9 for an extensive overview; [17, 37, 110, 239, 267, 366, 750]). In the NIR (700/750 nm to 1400 nm) pigment absorption is extremely low [169] and the leaf’s internal cellular structure (more in particular the structure of the spongy mesophyll) effects a very high and diffuse reflectance [315, 316, 566, 840]. In the case of diseased, senescent and heavily nutrient-deficient vegetation reflectance can drop significantly in the photographic NIR region [382, 473, 550, 702, 703], with an absolute change in the NIR reflectance that might be far more noticeable than the reflectance increase in the visible band (for an in-depth overview of a plant’s physiological and morphological state-related spectral differences in the NIR, consider Chapter 9).
Though imaging reflected NIR has been recognised as potentially beneficial, a conventional, film-based approach has certain inherent drawbacks. These include the requirement for cooled storage and transportation of emulsions, inappropriate exposure determination, narrow exposure latitude, and relatively weak sensitivity, making the complete NIR image acquisition and processing workflow costly and complicated, with a final outcome that is rather unpredictable.

10.1.3 Digital NIR Acquisition

Since the advent of digital photo cameras (also called Digital Still Cameras or DSCs), the acquisition of such NIR imagery has been simplified enormously, because their silicon image sensors are very sensitive to this invisible radiation, with a so-called cutoff wavelength $\lambda_c$ at circa 1100 nm [256, 569, 767]. Besides the digital image sensor, the whole imaging array of most one-shot DSCs also consists of a microlens array – used to increase the amount of photons impinging on the sensor’s photodiode (i.e. the light sensitive area which collects photons, hence creating one pixel of the final digital image) – and a Colour Filter Array (CFA): a mosaic pattern of coloured filters positioned above the photodiodes (Figure 10.3A) [232, 395, 569, 752]. As every photodiode of the image sensor has such a filter, only a specific wavelength range can be transmitted, subsequently generating a charge in the photodiode (Figure 10.3B).

Although both the sensor technology and the arrays of microlenses and coloured filters are responsible for some variation in the spectral responses of DSCs, it is safe to state that most imaging matrices are very responsive to NIR radiation (for a more in depth discussion, consider Chapter 8). To cut out the image-degrading effect of these non-visible wavelengths, camera manufacturers place an NIR-blocking filter in front of the sensor [128, 454, 651]. By removing this optical element and replacing it with a visibly opaque filter, all visible wavelengths are
removed before they reach the sensor, allowing only NIR photons to pass. Such a modification hugely increases the DSC’s sensitivity to NIR, while retaining the facility to view through the lens (impossible in the film-based approach of pure NIR imaging).

Figure 10.3. (A) The layout and principal of a Bayer CFA; (B) wavelength versus absolute Quantum Efficiency (QE) for the Kodak KAF-8300 (adapted from [261]: Fig. 5).

Using a dedicated NIR DSC also deals with most of the difficulties presented by film (supra). Additionally, digital solutions offer enhanced Quantum Efficiencies (QE) and larger Dynamic Ranges (DR) [158, 373] when compared to analogue approaches, which means that the former can be applied in far from optimal operational conditions. Moreover, a DSC’s linear response to radiation as well as its direct feedback on accurate focusing and exposure enables a very consistent output. Finally, DSCs are suited to mapping purposes, as they do not suffer from geometric film distortions [441, 442]. In spite of these major advantages, the application of digital NIR imaging with DSCs was never really investigated in archaeological reconnaissance.

Using imagery generated by such a modified DSC and conventional frames from a simultaneously operated unmodified DSC, Verhoeven & Vermeulen ([784] or Chapter 9) are the first to give an overview of situations in which these easy-to-use digital NIR-imaging instruments might be archaeologically advantageous. Specifically, by comparing both data sources, these authors demonstrated the potential of this approach to overcome the constraints of detecting and interpreting only the varying visible colours in vegetation, while still maintaining a flexible and economic approach (the latter in terms of imaging instruments).

This chapter further explores the possibilities of such a converted DSC in extracting even more meaningful information from an acquired NIR frame, reporting on the evaluation and quantification (as with any scientific measuring tool) of the intrinsic properties of an NIR-modified Digital Single-Lens Reflex (D-SLR) camera. This assessment of the channel-dependent spectral responses and the accuracy of capturing NIR photons might offer significant possibilities in the data processing, interpretation, and quantification of the acquired imagery. Instead of only using the imagery ‘straight-out-of-the-camera’, exploiting the DSC’s individual spectral responses should, ideally, permit the capture of (archaeologically) induced growth stresses in crops even better (i.e. enhance the contrast between the archaeological residue and the landscape matrix [65]).
10.2  DSC Characterisation – Material and Method

10.2.1  Hardware

For reasons discussed by Verhoeven [776], a Nikon D50 D-SLR was employed. The NIR modification of the DSC (hereafter called D50nir) was executed by Jim Chen [172], who placed a sort of cold mirror in front of the sensor to block most visible radiation. The sensor itself, a Sony ICX413AQ APS-C format sensor (called DX format by Nikon) of the Charge-Coupled Device (CCD) type, measures 23.7 mm x 15.6 mm and contains 3008 effective photodiodes in width by 2000 photodiodes in height [576, 730]. Above this sensor, an on-chip three-colour Red-Green-Blue (RGB) CFA is fitted, with the filters arranged in a Bayer pattern as shown in Figure 10.3A. Bayer’s pattern features twice as many Green filters as Blue or Red filters to improve the sampling of the luminance information [62], generating digital imagery with higher perceived sharpness [62, 601].

As the majority of optical glasses and polymers freely transmit NIR [651], most lenses can be used for NIR imaging [128, 241, 677]. On the D50nir, the Nikkor 20 mm f/3.5 Al-S and AF-S DX Zoom-Nikkor 17-55 mm f/2.8G IF-ED are used for Helikite Aerial Photography (or HAP: i.e. remotely controlled photography by means of a Helikite, a helium balloon with kite wings – [781] or Chapter 5) and photography from an aeroplane respectively. While the latter lens is slightly more prone to hot spots (a brighter area in the centre of the image produced by internal reflections) than the fixed focal length lens, it allows for zooming, which is often necessary when flying. The prime lens is, however, a top class performer in the NIR, capable of producing very crisp and extremely sharp images [678]. Moreover, it features an NIR focus mark. This lens was also used in the subsequently described spectral analyses. To verify the consistency of the results, all tests were repeated with another fixed focal length lens, the AF Nikkor 50 mm f/1.8D.

10.2.2  Image Acquisition

To identify the NIR behaviour of the D50nir’s complete imaging system (lens + cold mirror + microlenses + CFA + CCD), spectral response data are very important as they represent the digital output of the image sensor per incident light energy of a certain wavelength. In the procedure followed, a 2800 °K tungsten lamp was used as a reference EM source with known spectral output. A small part of the emission spectrum was selected with a Zeiss quartz prism monochromator (type Carl Zeiss M4 QII) in the wavelength range from 400 nm to 1100 nm. Using quartz prisms for wavelength selection is beneficial as no second order contributions, typical when using a diffraction grating, exist. Nevertheless, it was verified that no spurious light in other than the selected wavelength range was present.
Subsequently, a small entrance slit was fitted on the monochromator to obtain a Gaussian distributed narrow band stimulus. The transmitted waveband was then characterized with a calibrated Ocean Optics QE65000 spectrometer (wavelength resolution of 0.8 nm), to accurately determine the peak wavelength and the bandwidth, which typically had a **Full Width at Half Maximum (FWHM)**; for an explanation see 11.4.1 of 2.8 nm at 600 nm and 5.2 nm at 950 nm. Finally, characterization with the spectrometer allowed the number of photons that passed at each selected wavelength to be determined. The D50 NIR was irradiated with its sensor perpendicular to the output of the monochromator, to minimize as much as possible the angular dependence of the image sensor [547]. Pictures of the transmitted radiation were acquired at monochrome EM levels every 5 nm to obtain sufficient data points. The D-SLR used a lens aperture of f/5.6 and a total exposure time short enough (0.25 s for visual and 5 s for NIR) to make sure no photodiode became saturated, while the integration time was still long enough to generate sufficiently high **Digital Numbers (DNs, also called Analogue-to-Digital Units or ADUs)**, essential for an acceptable **Signal-to-Noise ratio (SNR or S/N)** and related measurement accuracy. For all images, the D50NIR’s default ISO 200 setting was used, yielding a minimal **gain** of 6.57 e-/DN with the 12-bit ADC (Analogue-to-Digital Converter). This value was calculated according to the method described by Berry & Burnell [77] and indicates the number of electrons that will cause the DN to increase by one, hence corresponding to a linear scaling factor $K$ of 0.152 DN/e- ($K = 1/g$).

As it is very important to work with the initially generated integer values, using RAW imagery is crucial. In essence, a RAW file is nothing but an array of DNs, each of them generated by one photodiode and proportional to the EM radiation of a certain wavelength range (determined by the coloured filter on top) plus some offset due to **dark current** and **bias**. Because the D50NIR utilizes a 12-bit ADC, the DNs can vary from 0 to 4095, corresponding to a **tonal range** of $2^{12}$ (i.e. 4096) different gradations. Using a RAW-workflow ensures that imagery for analysis is the ‘pristine’ data generated by the sensor, as these files (which can be created by most consumer, and all professional DSCs) were not subjected to any colour signal-processing algorithms (i.e. white balancing, demosaicing, tonal curve, noise suppression and sharpening) by the DSC’s firmware, unlike in-camera generated JPEGs and TIFFs (for a discussion on the necessity of using RAW in aerial photography and scientific imaging in general, consider Chapter 7).

### 10.2.3 Image Calibration

Subsequently, the RAW images (called **NEF** by Nikon, meaning **Nikon Electronic Format**) were imported to The MathWorks’ MATLAB to measure the DSC’s response to the narrowband illuminations, but not before calibrating the imagery by removing some unwanted signals.

In scientific digital imaging, only the stream of photons that reach the sensor (i.e. the **photon signal**) is of interest. However, the light frame captured by an image sensor always encompasses three particular signals: being the photon signal, the **dark current signal**, and
the bias signal/DC offset [77]. Unlike the photon signal, which is generated by the accumulated EM radiation during the exposure, dark current is a signal that is produced even when the sensor is not illuminated, due to thermally induced electrons. This dark charge accumulates with integration time/exposure and is heavily temperature and ISO dependent. The bias component, a small and mostly steady zero voltage offset which occurs even in the total absence of illumination, is due to the effects of electrical charge applied to the detector prior to exposure [77, 720].

Each of these nonrandom signals has some corresponding random variation (i.e. noise) embedded, all three varying according to the imaging technology used [406]. Besides photon/shot noise (\( \sigma \)) and dark current noise (\( \sigma_d \)), caused by the inherently random process of photon arrival and both obeying the law of Poissonian statistics [232, 263], there is the signal independent read/readout/bias noise (\( \sigma_{\text{ran}} \)): the sum of the reset noise (\( \sigma_{\text{reset}} \), the on- and off-chip amplifier noise (\( \sigma_{\text{amp-on}} \) and \( \sigma_{\text{amp-off}} \)) and the quantization noise (\( \sigma_{\text{ADC}} \)) [395, 655]. In the D50_NIR this minimal noise floor was measured to be about 1.04 ADU (12-bit) or about 6.83 electrons r.m.s. (i.e. 1.04g), an extremely low value that makes the D50_NIR completely photon noise limited when imaging normal signal levels and set to ISO 200.

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Symbol</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Photon signal/count</td>
<td>( x )</td>
<td>electrons</td>
</tr>
<tr>
<td>Signal bias</td>
<td>( b )</td>
<td>ADUs or DNs</td>
</tr>
<tr>
<td>Dark current</td>
<td>( x_d )</td>
<td>electrons</td>
</tr>
<tr>
<td>Photon/photon shot/shot noise</td>
<td>( \sigma )</td>
<td>root-mean-square electrons</td>
</tr>
<tr>
<td>Dark current (shot) noise</td>
<td>( \sigma_d )</td>
<td>root-mean-square electrons</td>
</tr>
<tr>
<td>Readout noise</td>
<td>( \sigma_{\text{ran}} )</td>
<td>root-mean-square electrons</td>
</tr>
<tr>
<td>Quantization noise</td>
<td>( \sigma_{\text{ADC}} )</td>
<td>electrons</td>
</tr>
<tr>
<td>Gain</td>
<td>( g )</td>
<td>electrons/ADU</td>
</tr>
<tr>
<td>Output RAW signal</td>
<td>( S_{\text{raw}} )</td>
<td>ADUs or DNs</td>
</tr>
<tr>
<td>Output RAW noise</td>
<td>( \sigma_{\text{raw}} )</td>
<td>ADUs or DNs</td>
</tr>
</tbody>
</table>

Hence, the DNs making up a NEF picture are the sum of the photon signal (with its corresponding Poisson noise), an unwanted dark current signal (with Poisson noise) and a bias constant (with readout noise), mathematically written as:

\[
S_{\text{raw}} = \frac{x}{g} + \frac{x_d}{g} + b
\]

with the total noise:

\[
\sigma_{\text{raw}} = \frac{1}{g} \sqrt{\sigma^2 + \sigma_d^2 + \sigma_{\text{ran}}^2}
\]

([77]; quantities in Table 10.1)
Due to their randomness, the noise components are difficult to correct for. However, the dark current and bias signals can be removed during calibration. To reveal the dark characteristics of the D50NIR, several sets of five NEF images were shot at dark condition, each set with a different integration time, starting from the fastest possible shutter speed (0.00025 s) up to one second, while the DSC was in thermal equilibrium at a constant room temperature (20 °C). After reading them out linearly (i.e. omitting the non-linear tonal redistribution normally applied by DSCs – see 7.5.3) and disregarding White Balance (WB), the RAW frames were converted to 16-bit TIFFs (one averaged version per set) and both mean and standard deviation of the output values plotted versus integration time. The results are presented in Figure 10.4A, and show this D50NIR to have significantly low dark noise levels at ISO 200.

![Figure 10.4. (A) Dark current (+ bias signal) generated DNs versus exposure time (s) for very short exposures; (B) DNs generated by dark current (+ bias signal) without subsequent median filtering.](image)

However, the sudden drop in maximum dark pixel value makes a particular Nikon characteristic apparent. This is that the firmware runs a median filter when the DSC takes an exposure of ≥ one second, aimed at reducing the effects of hot pixels during long exposures, yielded by particular photodiodes with abnormally high dark current. The French astronomer Christian Buil found a way round this [124], by turning noise reduction on and shutting down the D-SLR immediately after the exposure has completed, thereby aborting the noise reduction job and saving the pure RAW image directly from the buffer to the memory card. When applying this method, it is seen from Figure 10.4B that the mean linear dark current is still not even 5 e-/diode (i.e. 0.7 DN * 6.57 e-/DN) at an exposure of five seconds, which means its error contribution is still negligible (apart from a few hot pixels).

After using this method in the data acquisition, a dark frame was subtracted from all RAW images. Mathematically:

\[ S_{image} = S_{raw} - S_{dark} = \left(\frac{X}{g} + \frac{X_d}{g} + b\right) - \left(\frac{X_d}{g} + b\right) \]

with the resulting total image noise:

\[ \sigma_{image} = \sqrt{\sigma_{raw}^2 + \sigma_{dark}^2} . \]
This expression clearly shows the noise to slightly increase by dark subtraction. Therefore, rather than generating a single frame, a high S/N master dark frame yielded by averaging ten stacked 5 s-dark frames (or 0.25 s dark frames) was subtracted from the original image to average the random noise. As the master dark frame also contains the bias component $b$, this operation corrects for both unwanted signals, making the use of a bias-frame obsolete [77, 400, 837]. Thirdly, this approach also accounts for the possible amplifier glow resulting from a response of the photodiodes to radiation emitted by the readout amplifiers every time the detector is read out [720], although the latter was not attested visually.

Besides dark subtraction, calibration also involves the removal of a multiplicative component by flat fielding [77, 263, 400, 837]. This process corrects the image for Photo-Response Non-Uniformity (PRNU) by dividing the dark subtracted light frame with a master flat frame: an average of several dark current corrected images taken from a uniform/‘flat’ field of light, hence recording dust particles on the lens and sensor, optical vignetting and Photodiode Non-Uniformity (PNU), the main cause of PRNU [497].

Finally, all calibrated RAW images were analyzed with a purpose-written MATLAB programme. Once the spectral and intensity response of both Green filter sets were verified to be identical, a DN for the Red, Green, and Blue sensor responses was extracted by averaging over a rectangular section of some fifteen pixels by a hundred pixels in the centre portion of every image. The resulting set of three measured intensities allowed plotting of the colour filter-dependent relationship between the captured wavelength and the ratio of the DN to the intensity of the emitted radiant energy.

However, accurate measurement of such a spectral sensor response requires the output signal to be linearly proportional to the incident light intensity over a large range of input levels. Although this is known to be mostly the case [120], and certainly to be expected for modern DSCs [801], a coefficient of determination $R^2 > 0.99$ (calculated for both the complete CFA and all three colour channels) confirms the almost perfect linearity of the photometric response below saturation for this CCD, an observation that was also reported by Clark [179].

### 10.3 DSC Characterisation – Results and Processing

#### 10.3.1 Spectral Response Curves

Figure 10.5 displays the relative spectral sensitivity response of the different photodiodes in the D50nir to the 2800 °K lamp as measured with the procedure explained above. The graph describes the way in which the whole imaging matrix responds to particular wavelengths. By repeating the same procedure with an AF Nikkor 50 mm f/1.8D, it was verified that the impact of the photographic lens can be ignored to a large extent. Only from 740 nm onwards do the eleven lens elements of the Nikkor 20 mm f/3.5 AI-S [474] slightly decrease the NIR
transmission rate [507] compared to the 50 mm lens (which consists of only six lens elements [578]). This fact confirms that normal photographic lenses are highly transparent to NIR radiation, although – strictly speaking – they also have a specific spectral absorption response.

Besides transmitting radiation in specific spectral bands of the visual spectrum, the coloured filters thus also function as wavelength-specific filters in the NIR range, allowing the photodiodes to capture information in particular spectral bands. From the curves, it is clearly seen that the spectral sensitivity is almost negligible for visible light with wavelengths below 650 nm, corresponding to the cuton frequency of the NIR-pass filter in front of the CCD. Starting at about 660 nm, the Red photodiodes are most sensitive for deep-Red to NIR wavelengths, reaching a maximum at 730 nm. Above this value, the QE drops markedly due to generated electrons often recombining before reaching a sensor’s depletion region where they are stored [395].

The Blue filter locations are, however, totally insensitive for the entire visible part of the EM spectrum, as their sensitivity onset lies at 780 nm, rapidly increasing to a maximum response at around 815 nm. The spectral range of 795 nm to 875 nm at half maximum indicates most information is gathered before the moisture-sensitive NIR trough starting at about 940 nm [621, 751], making the Blue filtered diodes in particular sensitive to vegetational density or biomass [446, 762, 840]. Because the general spectral response in the Blue channel is much weaker than the Green and Red responses, it is best to expose with a somewhat longer-than-normal integration time. This will effectively counter high noise levels, as the SNR increases with the square root of all photons captured by the diode:

\[
SNR_x = \sqrt{x} \quad [232, 263]
\]
Finally, the Green diodes show an intermediate spectral behaviour, being responsive to EM radiation from 680 nm onwards, till they also reach a maximum at about 815 nm. On this long wavelength side (> 820 nm), the similar response of the particular diodes indicates that the RGB filters become nearly completely transparent to the incident radiation, until the imaging matrix becomes the perfect equivalent of a monochrome detector at around 850 nm, meaning that all filtered photodiodes are equally sensitive to the incoming radiation.

For wavelengths longer than 1000 nm, the D50\textsubscript{NIR}’s QE becomes extremely low, due to the inherent wavelength-dependent low absorption coefficient [395]. On the other side of the spectrum, the sensitivity in the wavelength range from 400 nm to 650 nm is extremely low as one would expect from a good visible-blocking filter. Only the Green and Red photodiodes show a very small response, with Green peaking spectrally at 565 nm. Yet, the contribution of these wavelengths to the final output can be safely ignored.

### 10.3.2 New Spectral Bands

NIR imagery generated by the D50\textsubscript{NIR} has already been used in archaeological research (see [776] or Chapter 8, Chapter 9, and Chapter 11). However, the spectral characterisation described above allows one to go beyond the initial approaches in which the default output was used. Because this analysis has clearly revealed the unequal spectral responses of each photodiode type, spectroscopic information can be extracted by differentiating between the Red, Green, and Blue channels. The normalised spectral response after subtraction and addition of particular channels is shown in Figure 10.6.

These mathematical operations make sense, as all three diode types have the same transmittance on the long wavelength side, while the Blue and Green spectral responses completely fit within the response ranges of the Green and Red diodes respectively. This way, the Blue, pure NIR component can effectively be filtered out of the Green channel, whereas subtracting the Green from the Red channel seriously narrows the bandwidth of the latter. Adding the Green to the Blue band on the other hand, creates a new spectral range that peaks around 815 nm, with a better response in the 750 nm to 900 nm range, where a plant’s maximum NIR reflectance lies [219].

Table 10.2 gives an overview of all primary and newly-created bands that can be worked with, and their close resemblance to particular spectral bands acquired by satellite sensors, though for the purposes of this study only the archaeological potential of the bands displayed in Figure 10.6 is exploited (see 10.4). First however, one extra elementary processing step is explained.
Chapter 10

Figure 10.6. The Red channel minus the Green channel (R-G), the Blue channel subtracted from the Green channel (G-B), and the Blue channel added to the Green one (G+B); the peak response of each band is normalized to unity (data and illustration by P. Smet – Ghent University).

Table 10.2. All workable spectral bands generated by the D50

<table>
<thead>
<tr>
<th>Channels</th>
<th>Red</th>
<th>Green</th>
<th>Blue</th>
<th>R-B</th>
<th>R-G</th>
<th>G-B</th>
<th>G+B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectral range (nm)</td>
<td>660-1010</td>
<td>680-1010</td>
<td>780-1010</td>
<td>660-840</td>
<td>660-870</td>
<td>680-810</td>
<td>680-1020</td>
</tr>
<tr>
<td>Spectral range at half max. (nm)</td>
<td>690-815</td>
<td>710-875</td>
<td>795-875</td>
<td>690-790</td>
<td>690-775</td>
<td>710-795</td>
<td>780-875</td>
</tr>
<tr>
<td>Maximum (nm)</td>
<td>730</td>
<td>815</td>
<td>815</td>
<td>730</td>
<td>730</td>
<td>775</td>
<td>815</td>
</tr>
<tr>
<td>Similarities</td>
<td>RBV3</td>
<td>MSS7;</td>
<td>MSS6</td>
<td>~MSS7;</td>
<td>~SPOT3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

10.3.3 Demosaicing

Apart from the few DSCs that have a Foveon X3 sensor, single-shot DSCs usually feature one CCD, Complementary Metal Oxide Semiconductor (CMOS), N-channel Metal Oxide Semiconductor (NMOS or MOS) or Junction Field Effect Transistor (JFET) sensor with an additional CFA to allow one particular spectral band to be captured by each photodiode. Consequently, a mathematical operation must be executed to fill in the DNs for the other two bands, a process commonly referred to as demosaic(k)ing, colour reconstruction, CFA-interpolation or de-Bayering (in case a Bayer array is used). Given the widespread use of CFA’s, a large range of linear and nonlinear algorithms have been created to reconstruct the final RGB image as accurately as possible (e.g. [12, 111, 166, 221, 356, 496]).

However, these methods were designed to demosaic information from the visible domain and the assumptions underlying most of them may not hold for NIR wavelengths, making them sometimes unsuited to interpolate missing information in NIR imagery. Previous research by
Verhoeven (see 7.5.2), however, indicated that the **Adaptive Homogeneity-Directed demosaicing algorithm (AHD)** [390]) performed very well in this invisible domain. As this algorithm is implemented in the programme dcrw, this software has been used to demosaic all NEF images. Moreover, this free RAW decoder works on any operating system and is capable of writing reconstructed 16-bit TIFF files [184] without applying any tonal/gamma curve or WB (omitting the latter two is often of the utmost importance in scientific applications – see Chapter 7). As in-camera generated TIFF and JPEG files do not allow this approach, the following analysis assumes a complete RAW workflow, yielding completely linear developed files in which the DNs are still equal to the ones initially generated by the sensor, but with all three channels completely reconstructed.

### 10.4 Archaeological Results

Do the three dissimilar spectral responses of the D50\textsubscript{NIR} allow the researcher to gain more archaeological information out of a ‘straight-from-the-camera’ NIR frame? The answer to this question is illustrated in Figure 10.7. In the upper part (A + B), two 16-bit versions of the same aerial photograph are shown, taken with the D50\textsubscript{NIR} on July 20, 2007 at 13.30 h above the central Adriatic Roman town of **Septempeda** (43° 14’ 10” N, 13° 11’ 52” E – WGS84). Figure 10.7A was created by opening the original RAW file in Capture NX (Nikon Corporation), a dedicated RAW converter for NEF files. As with all RAW converters, this programme automatically applies a tonal correction to the data [i.e. a gamma-like curve to rectify the mismatch between the approximately logarithmical **Human Visual System (HVS)** and the linear sensor] and white balances the scene by multiplying every spectral channel with a preset weight, thereby correcting for the differential spectral response of the DSC and compensating for the varying spectral output of the light source (see 7.5.1). Figure 10.7B on the other hand was converted and demosaiced using dcrw. The corresponding histogram shows the channels not to be equal (unlike Figure 10.7A), while the maximum DNs are also smaller than the Capture NX version, indicating the file to be completely linearly processed. Histogram stretching of Figure 10.7A, often necessary to tackle the non-maximized tonal range in NIR aerial photographs, yields the greater contrast seen in Figure 10.7C. Although some features start to become apparent faintly, this result is largely inferior to Figure 10.7D, which clearly indicates lighter and darker patches in the colza field, indicating the presence of underground structures such as roads, buildings, and ditches. The approach that yielded this result was a simple arithmetic operation on Figure 10.7B:

\[
F(i,j) = \frac{[R(i,j) - G(i,j)]}{[G(i,j) + B(i,j)]}
\]

in which \(F(i,j)\) is the final pixel and R, G, and B indicate the value of this pixel in the Red, Green, and Blue channels respectively (a computation that is valid, as demosaicing attributed each pixel with three complete spectral channels).
This operation clearly enhances the contrast between the soil and the vegetation as well as biomass differences in the canopy, revealing subtle dissimilarities that are largely masked in the structure of the original image. The result is no coincidence. Although the bands used are rather broad (85 nm FWHM and 95 nm FWHM), dividing them yields a so-called Simple Ratio (SR), a result that is also known as Ratio Vegetation Index (RVI) or Vegetation Index Number (VIN). As the first true Vegetation Index (VI) developed by Birth and McVey [86], Jordan [431], and Pearson and Miller [606], this ratio is known to indicate the amount of green biomass or Leaf Area Index (LAI) better than either band alone [431, 623, 687]. In all these pioneering cases, an NIR waveband was divided by a part of the Red spectrum (740 nm/675 nm; 800 nm/675 nm and 780 nm/680 nm respectively).

Although Gitelson et al. [335] also suggested a $R_{\text{NIR}}/R_{700}$ ratio, it was opted to divide the Red by the NIR band, just out of convenience rather than following other scholars (e.g. [848]). In this way the resulting vegetation marks have a greater resemblance to crop marks as they appear in the visible spectrum. Because the maxima of the Red and NIR bands are situated near 730 nm and 815 nm respectively, the operation has also close resemblance to the $R_{850}/R_{710}$ ratio, the latter being proven by Datt [219, 218] to exhibit a very strong correlation with chlorophyll content.

Besides these comparisons, Figure 10.7D demonstrates that this simple VI is effective, exploiting the fact that when dealing with healthy green vegetation, absorption is high in the
Red band while the plant’s mesophyll tissue allows for a strong NIR reflection. Correspondingly, these areas are displayed dark in the output. In the case of the Roman road in the centre of the picture, the bare soil and/or decreased LAI markedly increase the magnitude of the Red/NIR ratio, creating lighter areas or negative crop marks. Although the Spot 3-similar Blue band \[68\] has the advantage over the Green or Green+Blue channel through not including any visible radiation, the incorporation into the SR did not yield better results (as all pictures were taken before the DSC’s spectral characterisation, and the signal of the Blue channel was not optimised to counter the noise levels). Longer exposures with a higher SNR should yield equally, if not better, results.

In a second example, the same SR was tested on remotely sensed data from a totally different situation. Figure 10.8A shows the greyscale and histogram stretched version of a Canon EOS 300D digital colour photograph of the western grassland part of the Italian Adriatic Roman coastal colony Potentia \(43^\circ\ 24'\ 53"\ N, 13^\circ\ 40'\ 14"\ E – WGS84\), taken on July 17, 2007 at 15.00 h. It shows an excavation area (1), traces of the Roman street pattern (2), and a plot of cut grass (3), needed to perform geophysical research. Additionally, two paths to the excavation area are depicted, one created by mowing (4) and a second, smaller path of trampled vegetation (5) as a result of passage to and from the excavation area. Just as the traces of the wheelbarrow traffic (6), the latter is characterised by a yellowish-brown appearance, a very strong visual indication of plant stress \[13\]. Figure 10.8B and C respectively show a demosaiced, linearly converted and histogram stretched 16-bit aerial D50\(_{\text{NIR}}\) photograph and its extracted Blue layer, taken the same day at 12.45 h.

Figure 10.8. Comparison between a conventional photograph (A; photograph by F. Vermeulen – Ghent University) and three versions of an NIR photograph: (B) is the complete NIR frame, (C) displays the Blue NIR channel, while (D) visualizes the result of the SR.
Due to the extreme and long-term drought-induced stress the plants suffered from in the Italian summer of 2007, Figure 10.8B – and certainly the pure NIR image (Figure 10.8C) – clearly shows the traces of the Roman street pattern much better than Figure 10.8A. Although the stressed plants reflect greater Green and Red radiation (due to the substantial loss of chlorophyll), the street traces stay faint in the visible domain as the surrounding vegetation is also wilted to a certain extent and the lower canopy closure causes an increased reflectance due to a lower density of photosynthetic pigments per unit soil surface area. Consequently, the differences between both vegetation stages in the visible domain are small when compared to the NIR reflectance dissimilarity. The fact that these NIR crop marks are even visible in grassland indicates the very high soil moisture deficits this vegetation is suffering from [430]. Moreover, Colour InfraRed (CIR) imaging was also reported earlier to have a clear advantage over colour photography in detecting archaeological crop marks in pasture during summer [366], while pure NIR should better reveal crop marks in dry vegetation [465].

On the other hand, all other features mentioned are easier to distinguish in the visible domain than in any of the three D50_{nir}'s layers, as the decrease in total chlorophyll content is much larger than the change in internal cellular structure of the vegetation.

However, the aforementioned ratio again clearly reveals (Figure 10.8D) these biomass related traces – the square, both the paths and the wheelbarrow area. As the street pattern almost completely disappears in Figure 10.8D, this feature is less related to large differences in chlorophyll content and LAI.

Although both pictures were not taken simultaneously from the same spot (at 15.00 h from the aeroplane and circa two hours before with the use of the Helikite – marked by its shadow in the middle of the frame), the angles of view of the DSCs and the position of the sun did not change to such an extent that observed differences could be attributed to them. Indeed, the parameter that changed most was the solar geometry, whose effects are known to be of a limited importance [485], certainly when the sun has a very small zenith angle [809].

In addition to negative crop marks, positive crop marks might also be distinguished by the SR. From the contrast-enhanced RAW image in Figure 10.9B, two zones with higher and denser vegetation are obviously registered brighter when compared to the surrounding plant canopy, due to the fact that the larger biomass of both features effects a higher reflection of incident NIR radiation. The visible frame from this scene (Figure 10.9A), simultaneously captured with the NIR image above the centre of the Roman town of Ricina (N 43° 19' 41", E 13° 25' 26" – WGS84) on May 15, 2008 at 11.27 h, gives only a small hint of the presence of these non-archaeological positive grass marks (1 and 2 in Figure 10.9A). Moreover, the hydrographical features visible in Figure 10.9B are largely indiscernible in Figure 10.9A, showing the importance of NIR acquisition in this situation (see also Chapter 9). Notwithstanding, the NIR record fails to clearly distinguish between the stone walls of the Roman theatre (upper part of the frame) and the grass growing in between. Calculating the SR yields Figure 10.9C. When comparing all three frames, the magnitude of reflectance dissimilarity in the grass field seems
largest in the SR output. This mathematical operation also highlighted the lack of contrast between the theatre walls and the vegetation, although it was not able to visualise the old hydrographical features.

Figure 10.9. Visible image of the central part of the Roman town of Ricina; (B) NIR image of the same scene with some contrast enhancement; (C) the output when applying the SR with the channels form (B). Images acquired with a Nikon D200 (A) and a Nikon D50_{nir} (B & C).
10.5 Discussion

From the results presented, it is clear that the archaeological potential of a modified, NIR-enabled DSC cannot be underestimated. Both the use of individual spectral channels (e.g. the pure NIR image generated by the Blue diodes) and arithmetic operations performed on a combination of channels (e.g. the calculation of a SR) offer many opportunities to visually enhance archaeologically related anomalies and/or even reveal completely new, previously hidden archaeological information (as shown in [776] and Chapter 9). Although the application of NIR aerial imaging is by no means novel in archaeological reconnaissance, the additional advantages modified DSCs can offer in the generation and interpretation of NIR photographs is substantial. Not only do they significantly simplify the complete workflow, they also expand the possibilities known from the film-based NIR approach (pure NIR or CIR), without the costs of the latter.

However, the real-world examples also point to some important issues. First of all, both visible and NIR information (pure NIR and calculated SR) clearly need to be used together to get a relevant archaeological picture (e.g. [366]) and in other non-archaeological disciplines [188], certainly at times when stress has developed sufficiently, causing lower NIR reflectance of the canopy. From an interpretational point of view, the visible information remains very important since the HVS is trained to spot and interpret vegetation marks (but also soil, shadow, and other patterns) in this part of the spectrum. Moreover, when dealing with chlorotic vegetation, reflectance data in the visible domain are also of the utmost importance as these very common negative crop marks are extremely hard to distinguish in a pure NIR image (as also witnessed by [282]), even though the SR can tackle this issue to a large extent. Therefore, building a simple camera rig to hold two DSCs is advised, in order to simultaneously acquire NIR and visible wavelengths (while offering the possibility to mathematically combine particular spectral channels).

Secondly, all photographs (except those in Figure 10.9) were acquired in less than optimal circumstances, because long hot dry periods present the least discriminating conditions to fly in [315]. It can be expected that flying directly after rainfall could significantly improve the results yielded by the D50_{NIR} and the calculated SR.

Thirdly, the values of the SR sometimes exhibit very little variation, a phenomenon that can largely be attributed to two causes. On the one hand, the photographs under consideration show grassland and semi-arid zones, regions where the SR is known to be less effective in discriminating biomass/LAI variations [401]. To counteract this, other mathematical operations were tried (particular normalizations and VIs as DVI, NDVI). Generally, it was this SR that yielded the best – and certainly most consistent – results in these low-cover areas, which confirms to a degree the results of Baugh and Groeneveld [60].
Spectral Characterisation to Enhance Observation

On the other hand and more importantly, the applied SR does not really involve the mean Red reflected radiant flux to mean NIR radiant flux. While the Blue+Green channel (spectral range at half maximum of 780 nm to 875 nm and a sensitivity peaking at 815 nm) is well-suited as a reference band, being very little affected by either chlorophyll or water vapour absorption [762], the Red–Green channel is still spectrally too broad to be effectively used as a band that shows maximum sensitivity to pure chlorophyll absorption. Although the Green subtraction proved very useful in removing much NIR radiation from the Red channel, the resulting response curve – which has a spectral range of 690 nm to 775 nm at half maximum – completely overlaps the stress sensitive Red edge region (i.e. the very steep increase in a healthy green plant’s reflectance curve at the edge of the visible light and the beginning of the NIR spectrum [398]), something that should be omitted as it reduces the accuracy of vegetation investigation [141, 311, 763].

A solution to tackle these problems of the D50 NIR and the resulting SR is being worked on, involving flying with another, simultaneously operated, DSC that acquires radiation from the Red edge spectral region (690 nm to 710 nm). This zone has several times proven to give the most consistent leaf (and even canopy) reflectance response to plant physiological stress [140, 141, 143, 144, 145, 148, 338] and therefore of extreme importance in several narrowband VIs for chlorophyll estimation, even at the canopy level [195, 289, 480]. As this range is severely compromised in unmodified DSCs, a similar modified DSC equipped with a narrowband interference filter attached to the lens would be needed to generate aerial frames using only the reflected radiation from this stress-sensitive side of the chlorophyll absorption band. This would increase the correlation of the proposed reflectance ratio to plant senescence and stress, allowing the spectral characteristics of the D50\textsubscript{NIR} to be exploited more fully. Such an approach (which is completely outlined in the next Chapter) offers archaeologists an affordable and easily managed multispectral tool that can provide useful information on the vegetation’s physiological and morphological conditions to aid in the survey of the archaeological subsurface. If flying with a second (visible) or third (visible and 700 nm) DSC is impossible to achieve, the spectral characteristics of the D50\textsubscript{NIR} and the resulting SR will still most likely allow more relevant vegetation information to be gathered by comparison with only a pure NIR band.

However, no matter how efficient and accurate this new ‘tool’ can be, an increase in site discovery rate using multispectral imaging with DSCs is unlikely as long as the predominant flying strategy of ‘observer-directed’ survey and photography is in practice (e.g. [596]). This approach generates extremely selective (i.e. biased) data that are totally dependant on an airborne observer recognising archaeological phenomena. Thus subsurface soil disturbances that are visually imperceptible at the time of flying will not make it into an NIR photograph (even if the spectral response in this domain is distinct). The large-scale use of the techniques advocated in this chapter require a new (or call it additional) approach to archaeological aerial reconnaissance, that is flying to collect geographically unbiased photographs of large areas (a point that was already raised by other scholars concerning aerial imaging in the visible domain [186, 543, 596, 597]. Otherwise, non-visible and narrowband imaging will only enhance the
record of known features and – in the best case – reveal new, previously undetected archaeological details within a site that can be seen from above (which, however, should still not be underestimated, as new evidence may always alter the archaeological appraisal [367]).

10.6 Conclusions

Archaeological aerial reconnaissance has long been, and to a certain extent still largely is, equated to flying around in a small aircraft, using still cameras to record archaeological anomalies recognised by the airborne observer. Although satellite, multi-, and hyperspectral airborne data have been used in a variety of archaeological surveys, most users often lack both the financial and staff resources to acquire and handle the majority of these data (let alone the fact that the image acquisition is executed without taking the specific archaeological requirements and constraints into account). This does not, however, imply that technical enhancements have to be ignored, and certainly not if they can be achieved cheaply. It is therefore encouraging to see that the products of the current digital photography industry can have a great contribution in the low-cost technological improvements needed to better understand the buried landscape record.

In 1936, Reeves wrote about aerial archaeology, pointing out that as “its methods and technique are improved, aerial photography will increase in scientific value” ([654]: 107). Seen from this perspective, the ability of modified DSCs to acquire non-visible data in wide and/or narrow wavebands can be just the tool archaeologists need to increase the scientific value of every single flight. However, testing these tools on their spectral capabilities is an absolute prerequisite for the optimal use of the generated aerial (archaeological) imagery, given the fact that no two imaging matrices are alike. Once all essential characteristics are known, such highly NIR-sensitive devices provide a cheap, compact, robust, and easy to handle means for a 'spectroscopic' aerial approach.

Allowing that the presented imagery was acquired in an unfavourable period and the Red–Green channel seems significantly broader than the ideal 690 nm to 710 nm band, the individual channels of a modified Nikon D50 proved very useful in the calculation of a simple VI to indicate chlorophyll-related issues, whereas the pure, broad-band NIR channels are more suited to reveal severe drought and nutrient stress in the canopy reflectance [248]. Besides using the three channels generated by one single modified DSC, their combination with discrete, specifically chosen spectral bands (that are generated by a tandem of photo cameras) looks promising.

Just as their use is not solely restricted to crop mark archaeology (see Chapter 9), NIR-enabled DSCs could also be applied in several non-archaeological domains, including agriculture, forest management, and the mapping of water bodies. Rather than making the other methods of data acquisition obsolete, modified DSCs thus offer convenient, low-cost possibilities to yield
essential beyond-visible information for the benefit of various aerial and ground-based disciplines.
Is Smaller also Better?
Red Edge and Near-UltraViolet Sensing

« Anyone who has never made a mistake has never tried anything new. »
Albert Einstein (1879-1955), physicist
Crop Marks Should Be Red

Imaging the Blue Shift of the Red Edge to Detect Vegetation Stress
Abstract

Scientists from different research disciplines have provided essential information that relates the biophysical characteristics of plants to their spectral reflectance. This fundamental understanding has facilitated the development of various non-destructive sensing methods for detecting vegetation stresses, monitoring plant growth, and calculating crop yield. To date, this knowledge has been used too little by aerial archaeologists flying around in small aeroplanes. Instead of founding the archaeological interpretation only on direct visual inspection of the conventionally acquired colour photographs, this contribution reviews the reflectance properties of plants and uses them to present a new imaging technique beneficial for the detection of (faint) archaeologically induced crop marks. The new approach consists of three simultaneously operated Digital Still Cameras (DSCs), each of them capturing information in a different spectral waveband. Besides a theoretical underpinning, real-world examples will assess the role of this new approach in crop mark detection and prove this low-cost, multispectral method to be beneficial in isolating and enhancing weak crop stresses that are lost when taking only the broad visible spectrum into account. In the final discussion, some thoughts on future archaeological aerial research are given as well.
11.1 Introduction

Accurate and cost-efficient monitoring and mapping of morphological and/or physiological crop changes is essential to study our hidden past. Supported by this belief, aerial archaeologists often acquire colour photographs from the cabin of a low-flying aeroplane using (digital) photographic cameras to record archaeological anomalies recognised by the airborne observer in the visible spectrum [596, 834]. Due to this ‘observer-directed’ approach, this type of aerial survey might be extremely biased as visually imperceptible soil and crop disturbances will not make it into a photograph, hence yielding inaccurate settlement patterns of previous periods [37, 198, 371, 372, 584, 697, 833]. Secondly, current imaging techniques do not often allow the detection of crop marks in less-optimal circumstances, as the slight differentials of height and colour in crop exhibit too low a contrast with the surrounding matrix [65] to be noticed through normal colour photography. Conversely, spaceborne data is consistently acquired over large, extended areas while these sensors often record beyond visible information. Although satellite imagery has been used in a variety of archaeological surveys (e.g. [34, 298, 301, 302, 345, 464, 463, 465, 524]), it is often less suited for the discovery and detailed recording of small archaeological features as image acquisition is not tailored to the needs of archaeologists (e.g. the resolving power of the imagery is in all but a few cases more than one meter). Moreover, the spectral bands are generally too broad or misplaced spectrally to truly detect plant stress [141]. Airborne multi- and hyperspectral sensors do acquire data in narrow wavebands, but cost, moderate temporal resolution, and low resolving power also significantly hamper its frequent use in archaeological research [37, 372]. An ‘ideal’ system would therefore join the best of these three approaches, offering the operating flexibility and low-cost equipment that are characteristic for conventional archaeological aerial reconnaissance, while also allowing a total coverage in narrow visible and invisible stress-sensitive wavebands.

This chapter wants to outline a new multispectral approach to archaeological aerial reconnaissance that combines these characteristics. By exploiting the preferential spectral absorption of leaf chlorophyll, this technique will prove useful to detect crop stresses that often remain undetected in the approaches sketched above. In order to test the capacity of the system (which is based on a three-camera rig), a conventional aerial reconnaissance flight was performed and areas with known sub-surface archaeological remains were photographed. The results of this flight are presented further on. In a second phase which is planned in the summer of 2009, this technique will be brought into action during vertical block coverage. Because the proposed approach to archaeological reconnaissance is based on basic knowledge of a leaf’s reflectance properties, it is instructive to first develop a basic understanding of the spectral responses (and their information content) of individual leaves in the visible and NIR domain and subsequently transferring this knowledge to the canopy.
11.2 Plants and their Spectral Characteristics

Remote sensing primarily uses visible, Near-InfraRed (NIR), and thermal bands to quantitatively assess plant parameters and detect stress. As Digital Still Cameras (DSCs or digital photographic cameras) will be used in this analysis and their imaging sensors are responsive to radiant energy from about 330 nm to 1100 nm [373] (see also Chapter 8), only the spectral characteristics of vegetation in the visible (400 nm to 700/750 nm) and NIR (700/750 nm to 1400 nm) domain are of concern. This chapter can only provide a summary of basic leaf and plant canopy reflectance principles that are essential to understand the subsequently outlined approach. For a more thorough reading with abundant referencing, the reader is referred to [379] who present an overview of the current understanding of pigment-related visible and NIR plant reflectance, while Verhoeven [778, 779] and Verhoeven and Vermeulen [784] further explore these insights by providing the necessary links to archaeological reconnaissance in the visible and NIR domain respectively.

11.2.1 Individual Leaves

When trying to detect stress in individual plant leaves, it is of the utmost importance to understand Figure 11.1A. It illustrates the most fundamental – and widely accepted – spectral reflectance characteristics of green plants. More in particular, this illustration displays the reflectance curve of a healthy green radiata pine leaf together with the spectral reflection of the same leaf when suffering from nitrogen (N) deficiency. Besides its crucial role in a plant’s oxygen production, it is a well established fact that the presence of leaf chlorophyll, a green pigment which comes in both a chlorophyll \(a\) (Chl \(a\)) and chlorophyll \(b\) (Chl \(b\)) variant, largely dictates how plants appear spectrally. Moreover, chlorophyll content is strongly related to plant senescence and stress [146, 383, 532]. Consequently, both spectral curves do not only give major clues about the amount of this photosynthetic pigment in a plant’s leaf, but also indicate the plant’s health status.

Current knowledge allows us to distinguish the following commonly accepted spectral leaf reflectance and absorption features for all green plants in the visible and NIR domain [129, 140, 141, 145, 169, 330, 335, 337, 336, 338, 379]:

- visible leaf reflectance decreases with increasing chlorophyll content. When chlorophyll content decreases due to plant stress or the principle of seasonal senescence, the total leaf reflectance increases. This behaviour is the most consistent leaf response to plant stress within the 400 nm to 2500 nm spectrum;
- the Blue band (400 nm to 500 nm) is more or less insensitive to variations in the pigment content due to the fact that Chl \(a\) and Chl \(b\) pigments absorb strongly in this range, while also other important plant pigments as carotenoids have a high absorption coefficient in the Blue region;
- near 670 nm (i.e. in the Red band of 600 nm to 700/750 nm), the reflectance and absorption only change in relation to chlorophyll content when its amount is very low to moderate. Once a leaf exceeds 100 mg/m² chlorophyll, the total absorption remains unaltered;
- the highest sensitivity of the reflectance curve to chlorophyll pigment variation occurs in the Green spectrum around 550 nm and in the Red edge (i.e. the steep slope of the reflectance curve in the far-Red to NIR transition spectrum – see Figure 11.1A) around 700 nm, because the absorption of Chl a and Chl b is very low in these regions. As a result, even very small changes in pigment content can alter the reflectance. This observation is confirmed by the grey curve in Figure 11.1A, which represents the Standard Deviation (STD) of reflectance calculated from several spectra of maize leaves containing different chlorophyll concentrations. Besides confirming the statement about the Blue waveband, this curve clearly illustrates the dispersion of reflectance data to be largest in the 530 nm to 590 nm zone and around 700 nm;

![Figure 11.1. (A) Typical reflectance responses of a radiata pine leaf. The discontinuous (green) curve of A represents the mean reflectance for healthy, chlorophyll-rich leaves, while the upper (orange) curve indicates the mean reflectance for a stress situation (N or nitrogen deficiency). The grey curve represents the STD of reflectance calculated from reflectance spectra of maize leaves with varying chlorophyll content (adapted from [145]: Fig. 2B, and [379]: Fig. 2 respectively); (B) structure of a plant leaf and its interaction with incident visible and NIR radiation (adapted from [727]: Fig. 32-3 and [748]: Fig. 2).](image)

- this sensitivity results in two very generic stress responses: the Green reflectance peak around 550 nm is broadened towards longer wavelengths and causes the leaf to appear yellowish (a phenomenon called **chlorosis**). At the same time, the **Blue shift** of the reflectance curve Red edge occurs in the very small region around 700 nm: due to the stress related increase of reflectance at wavelengths near 700 nm, the Red edge tends to shift toward shorter wavelengths (Figure 11.1A). This Blue shift is currently accepted as the most consistent stress response to different stressors;
- in the NIR region (700/750 nm to 1400 nm), the spectral properties of leaves are no longer governed by pigments. As a matter of fact, heavy scattering takes place inside the leaf’s internal cellular structure. More in particular the structure of the spongy
parenchyma mesophyll effects a very high and diffuse reflectance in the NIR spectrum (Figure 11.1B). The very steep increase in reflectance of EM energy between 700 nm and 750 nm is the aforementioned Red edge: the most prominent characteristic in the reflectance spectrum of healthy vegetation. This transition zone of very abrupt reflectance change results from the fact that the absorption by chlorophyll pigments is low and NIR reflection by the spongy mesophyll increases, giving birth to one of the most extreme slopes to be found in reflectance spectra of natural materials. Figure 11.1A shows that healthy green leaves often respond to short-term, acute stressors with a slight increase in NIR reflectance. This certainly holds when plants suffer from dehydration. A significant decrease in the NIR reflectance spectra of a plant will only occur in the case of chronic water shortage and disease, end-of-season senescence, or heavily nutrient-deficient vegetation, corresponding to the effect of major internal tissue degradation. At this stage, also the Red edge disappears completely.

11.2.2 Canopy

From a quantitative point of view, reflectance from the canopy might be seriously modified from that of an individual leaf, because canopy reflectance is an integration of several contributions: soil background reflectance, reflectance from non-green canopy components, the anisotropic behaviour of the canopy and its varying architecture, the amount of shadow, reflectance from weeds, and plant litter [114, 187, 358, 567, 580, 774]. As a result, the information transfer from a leaf’s reflectance to the reflectance of the canopy is often nonlinear, while leaf-based reflectance spectra might not directly and as a whole be transferable to the canopy level.

Canopies with a high LAI (i.e. Leaf Area Index or leaf area per unit of ground) can be characterised by a significant increase in NIR reflectance due to the process of leaf additive reflectance, while the reflectance from the ground might also seriously increase total reflectance in the Red band when imaging early in the growing season. In general, the spectral contribution of the soil is negligible if the canopy consists of a large number of healthy, mature leaf layers [211, 214, 358, 425, 566]. Gitelson et al. [331] have shown that the spectral features of wheat canopy reflectance closely resemble those of individual leaves when the Vegetation Fraction (VF) reaches 60 %.

Even though several techniques exist to counteract particular of these canopy issues in specific situations, many approaches currently used for canopy remote sensing were specifically developed on the basis of individual leaf reflectance and most of them proved to yield decent results. Moreover, archaeological reconnaissance does not care about accurately estimating particular crop parameters as the exact amount of chlorophyll or the extent of phosphorus (P) deficiency. What matters are reliable, generally applicable methods that can be extended across entire landscapes, at different times of the day, and among various growing seasons, rather than very specific approaches that might vary from field to field, crop to crop, and time
of the year. These remote sensing methods should allow to image the contrast exhibited (directly or indirectly) by an archaeological residue in its surrounding matrix [65], without the need to optimize the approach on the basis of crop species. Therefore, a very simple and generally applicable Vegetation/Vegetative Index (VI) is presented.

11.3 Vegetation Indices

The basic leaf and canopy reflectance principles have been used to built VIs, commonplace tools for vegetation assessment today and beneficial for extracting the green plant quantity signal from complete (and even complex) canopies [379, 627]. These VIs are mathematical waveband operations (ratioing, differencing, summing, ratioing differences or linear combinations) trying to relate particular spectral reflectances in two or more wavebands (mostly visible and NIR) from leaves or complete canopies to specific vegetation characteristics (LAI, chlorophyll content, biomass, crop water stress, N deficiency, etc.).

11.3.1 SR and NDVI: Broadband Indices

Although they are largely applied for canopies, many VIs are developed from an understanding of leaf reflectance properties [379]. The least complicated VIs are simple band ratios. As an example, the first true VI was developed by Birth and McVey [86], Jordan [431], and Pearson and Miller [606]: $R_{\text{NIR}}/R_{\text{Red}}$ (where $R_{\text{NIR}}$ and $R_{\text{Red}}$ represent the reflectances in the NIR and Red waveband respectively). The reasoning behind this ratio, which became known as the Simple Ratio (SR), Ratio Vegetation Index (RVI) or Vegetation Index Number (VIN), was very straightforward and results form the sharp dissimilarities in visible and NIR vegetation reflectance: because healthy, chlorophyll-rich vegetation features a high absorption in the Red band and allows for strong reflection of the NIR radiation, the ratio between these two distinct wavebands is very high. In periods of severe stress, this ratio will decrease in magnitude, as the Red and NIR reflectance will increase and decrease respectively. Hence, the SR can be used to detect uneven patterns of growth (crop marks) or indicate zones with a high amount of green biomass, only by taking advantage of an existing inverse relationship in leaf reflectance [431, 623, 687].

Ever since, a very large range of VIs have been proposed, all varying in sophistication, spectral information needed, and effectiveness (for a summary, consider [380, 425, 470]). Most of these VIs have shown they can go far beyond the detection possibilities of the human eye, and are able to retrieve specific plant-related information if the right combination of spectral bands is applied.
A second simple VI that, throughout the years, became a kind of benchmark when developing other VIs is the **NDVI (Normalised Difference Vegetation Index)**. Proposed by Deering [223], this VI equals

\[
\frac{R_{\text{NIR}} - R_{\text{Red}}}{R_{\text{NIR}} + R_{\text{Red}}}
\]

It was built upon the Difference Vegetation Index (DVI) proposed by Tucker [763], which equalled to \((R_{\text{NIR}} - R_{\text{Red}})\) and enhanced the vegetation signal by exploiting the Red edge properties. As this simple difference is very sensitive to unequal solar irradiance [415], a denominator \((R_{\text{NIR}} + R_{\text{Red}})\) was added to reduce this characteristic canopy effect, yielding a new normalised digital image with values ranging from -1 to 1: snow, clouds, and water have negative values, rocks and bare soils fluctuate around zero, while high positive index values indicate large amounts of green biomass (Figure 11.2A). Although it is the most used VI worldwide, NDVI is far from perfect. Generally, NDVI is only a successful measure at the beginning of the growing season or when dealing with sparse vegetation cover [328], because it is insensitive to all but low chlorophyll contents (Figure 11.2B clearly shows that reflectance in the Red waveband around 670 nm only changes till 100 mg/m\(^2\)). This implies that the NDVI rapidly saturates under moderate-to-high biomass conditions (e.g. [50, 129, 328, 331, 337, 339, 701]).

![Figure 11.2](image-url)

**Figure 11.2.** (A) NDVI map of the world, generated from data acquired in April 2002 by the MODIS sensor [462]; (B) reflectance in two visible bands, the Red edge, and the NIR range plotted versus varying total chlorophyll content in maize leaves (adapted from [379]: Fig. 3).

More than thirty years ago, NDVI was developed specifically for the broad wavebands acquired by the sensors onboard of the **Landsat satellite** series [88, 223, 415]. However, the large majority of remotely acquired imagery is still broadband spectral reflectance data obtained by various earth-orbiting satellites carrying particular multispectral sensors: e.g. MODIS (Moderate Resolution Imaging Spectroradiometer), ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer), AVHRR-14 (Advanced Very High Resolution Radiometer), ETM+ (Enhanced Thematic Mapper Plus).
Using spectrally broad wavebands is, however, far from optimal, because particular diagnostic spectral absorption features are often only a few nm wide, which makes acquisition of data with a very high spectral resolution of the utmost importance to assess small variations in the plants’ physiology [102, 761, 764]. Consequently, many imaging systems significantly reduce the diagnostic accuracy of vegetation investigation (and archaeological crop mark detection [366]) as they undersample the reflected information, masking spectral characteristics that are too narrow to be distinguished by coarse spectral resolution imagers [39, 87, 88, 114, 141, 398, 738, 751].

Moreover, broad spectral bands do not enable techniques as derivative analysis to be performed (to reduce background effects for instance [227, 276]). Hence, to really apply VIs for determining plant health or vigour, narrow stress-sensitive wavebands have to be used. Small-band VIs are thus often far more successful, as they can exploit very specific spectral information to a much larger extent. More specifically, Zhao et al. [848] determined the ideal FWHM bandwidths of NIR and Red channels applied in several VIs to be below 50 nm.

11.3.2 Narrowband Index: $R_{700}/R_{800}$

Based on this knowledge and the theoretical spectral reflectance facts outlined before, it seemed archaeologically very interesting to detect the generic stress related Blue shift of the Red edge by performing a ratio analysis of reflectance spectra, more in particular $R_{700}/R_{800}$. This index can be denoted a ‘spectral pigment index’ [88] because it employs ratios of narrow spectral bands that are respectively sensitive and insensitive to pigment content.

First of all, various research has proven that the long-wavelength side of the Chl $a$ absorption band (690 nm to 710 nm) gives the most consistent leaf (and even canopy) reflectance response to various plant physiological stressors and senescence [140, 141, 142, 143, 144, 145, 146, 148, 149, 151, 174, 291, 334, 335, 337, 336, 338, 847, 849], with an increase of reflectance at these wavelengths and an accompanying shift of the Red edge toward shorter wavelengths (Figure 11.1A) [102, 398, 539, 751]. This information became therefore of extreme importance in several narrow-band VIs to accurately estimate chlorophyll content, even at the canopy level (e.g. [139, 140, 141, 146, 148, 149, 150, 169, 195, 217, 219, 218, 277, 289, 329, 330, 331, 332, 334, 335, 336, 337, 338, 339, 340, 383, 480, 531, 623, 632, 685, 848, 849]).

As it is known that Red reflectance is highly correlated to chlorophyll content when standardized by reflectance in a non-absorbing waveband [214, 685, 753], $R_{800}$ is used as denominator. In cases of moderate stress related chlorophyll loss, NIR reflection remains almost equally high (see 780 nm data in Figure 11.2B) due to the unaltered internal scattering of NIR wavelengths [315, 616, 724, 840]. Moreover, Zhao et al. [849] determined the 750 nm to 850 nm waveband to be the most effective range for acquiring NDVI-related NIR reflectance data, with a central wavelength positioned near 800 nm yielding the best results.
Reflectance ratios are also known to correct for geometrical and background effects at the canopy scale [50, 141]. Carter and Miller [148] could therefore concluded that their $R_{694}/R_{760}$ ratio was insensitive to decreased LAI, while its value consistently increased with increasing stress. Moreover, the same $R_{694}/R_{760}$ index could detect chlorosis in pine canopies more than two weeks prior to visual symptoms [144], whereas an index identical to the one proposed here ($R_{695}/R_{805}$) could also visualise stress symptoms prior to visual clues [142]. Additionally, Zhao et al. [849] found that ratio-based VIs gave significantly higher overall success rates than orthogonal VIs such as TSAVI (Transformed Soil-Adjusted Vegetation Index) to indicate stress in cotton canopies.

Supported by these results, the proposed narrowband index should render faint and early archaeologically induced crop stresses better than using NIR, Red or visible reflectance information alone. In case the stress is more severe and/or chronic, NIR reflectance seriously drops [145, 213, 622] and the proposed reflectance ratio will generate an ever bigger contrast between the healthy and stressed plants. However, as archaeologists do need low-cost and generally applicable tools, a system had to be developed that could allow the two necessary small spectral bands to be acquired from the air in a straightforward way.

11.4 Material and Method

In order to test these findings and see whether or not this VI could record very slight stresses in crops, several distinct spectral bands had to be recorded, impossible with straight-out-of-the-box DSCs. Besides their broad spectral bands (see [779]), the sensitivity of these instruments to 700 nm and 800 nm radiation is low to non-existing (see Chapter 8). In an attempt to counteract this inherent characteristic behaviour, three DSCs were fitted to a single bar, allowing for simultaneous data acquisition.

11.4.1 Modified DSCs and Interference Filters

Two imaging devices that could acquire the required wavebands were necessary. A third DSC was also deemed important to allow the comparison of the VI output with conventional colour frames. The latter were generated by a Nikon D200 plus AF-S Nikkor 18-70 mm f/3.5-4.5G ED lens (fixed on the infinity setting at 50 mm), while a modified Nikon D80 and Nikon D50 allowed to capture beyond visible information. On the one hand, the Nikon D50 (hereafter called Nikon D50NIR) registered pure NIR reflectance, as its internal filter package was replaced by an NIR transmission filter which blocks all visible and Near-UltraViolet (NUV) radiation ([776] and Part IV). On the other hand, the internal filter package of the D80 was replaced with a clear glass window to take the Full Spectrum (FS) into account to which the digital imaging sensor is sensitive ([777] or next Chapter). This DSC is now indicated as Nikon D80FS.
To enable the acquisition of the small spectral bands, the D80FS was fitted with a Carl Zeiss interference filter dating from the 1960s. An interference filter is a multilayer thin-film device whose spectral properties result from destructive/constructive wavelength interference rather than absorption [726]. Due to its working principles, these devices easily allow the transmission of well-defined wavebands and reject all other unwanted wavelengths. Although they are seldom used with photographic camera lenses [651], these optical filters allow instantaneous recording of spectrally selective images.

The interference filter used here has its central wavelength $\lambda_c$ at 702 nm. This is illustrated by Figure 11.3A, which shows the filter’s transmission/response curve $T(\lambda)$: i.e. the percentage of incident radiation transmitted ($T$) as a function of wavelength ($\lambda$). As it permits the transmission of only a very narrow band around 702 nm, such an interference filter is denoted as a narrowband interference filter or optical bandpass filter [292].

The wavelength range transmitted is known as bandpass (or passband) and often quoted in terms of the Full-Width at Half-Maximum (FWHM) of $T(\lambda)$: i.e. the width of the spectral band, in nanometers, at one-half of the maximum transmission (see Figure 11.3A). In this case, the FWHM equals 12 nm, which boils down to a spectral transmission from 696 nm to 708 nm.

Even though the maximum transmission of this interference filter is rather low (25 %), it was preferred over modern filters in this testing stage for reasons of cost and convenience (the available range of modern filter diameters is seriously limited due to difficulties in manufacturing, which makes customisation necessary and this accessory thus too costly in case the technique would prove unsatisfying). Using a macro coupler ring, the older filter could be fit perfectly to photographic lenses with a filter thread of 52 mm in diameter.

Equipped with an AF Nikkor 50 mm f/1.8D lens (filter thread 52 mm), this DSC-lens-filter combination makes it possible to acquire reflectance information mainly in the long-wavelength side of the chlorophyll absorption band (690 nm to 710 nm).

![Figure 11.3. (A) Spectral transmission curve of the Carl Zeiss 700 nm interference filter (data by courtesy of P. Smet – Ghent University); (B) response curves of the same filter with varying angle of the incident radiation (data by courtesy of P. Smet – Ghent University).](image-url)
The focal length applied is not arbitrary, as standard interference filters are very sensitive to the angle of the incident beam of radiation [651, 726]. If used with non-collimated radiant energy, it is known that the central wavelength $\lambda_c$ of the filter shifts to shorter wavelengths (Figure 11.3B), a shift that becomes larger with an increase in the angle of incidence [726]. Additionally, the maximum transmission decreases and the transmission band widens (all visualised in Figure 11.3B). This effect is, however, significantly diminished by the 50 mm lens because its limited diagonal Field Of View (FOV) (i.e. 31° 44′ when mounted on a Nikon DX sensor – see 6.6) forces a relatively small acceptance angle for objects/areas imaged furthest from the optical axis [148].

The 800 nm band was provided by extracting the Blue channel from the raw D50$_{\text{NIR}}$ information. This channel only takes pure NIR information into account: it peaks at 815 nm, and is characterised by a spectral range at half maximum of 795 nm to 875 nm (see 10.3.1). This way, it satisfies the guidelines formulated by Zhao et al. [849], who stated that the FWHM of this NIR region should not surpass 100 nm. To focus the incoming NIR radiation, the D50$_{\text{NIR}}$ was coupled with an identical AF Nikkor 50 mm f/1.8D lens.

To ensure a straightforward image processing workflow, all three DSCs should ideally cover exactly the same area, which is only possible if their shutters operate simultaneously and the optical axes of the equi-focal length lenses are perfectly parallel. Hence, all three DSCs were mounted onto an aluminium bar (type Jasper Engineering 10” Twin Camera Bar) normally designed for stereo-photography (Figure 11.4A). Because the interfaces for remotely triggering the DSCs are dissimilar, a Phottix remote control (type WRC-N6 and WRC-N8) was used to trigger both the Nikon D80$_{\text{FS}}$ and D200 with the left hand, while the right hand simultaneously pressed the shutter button of the D50$_{\text{NIR}}$. Although being very basic, this low-cost solution proved to be reliable, while over 90 % of the imagery acquired was perfectly synchronous. A small GPS receiver was simultaneously logging the complete flight path to enable easy geo-locating the acquired sets of images.

Figure 11.4. (A) Stereo bar with three DSCs mounted; (B) inside a two-seater aeroplane holding an older two-DSC rig (photograph by F. Vermeulen – Ghent University).
11.4.2 Image Acquisition

Before mounting the three DSCs and accompanying accessories onto the photo bar and getting airborne, a whole list of settings has to be checked. First of all, each DSC has to be equipped with the largest memory card possible, as changing these tiny items inside a wobbling aeroplane is far from convenient while holding the DSC rig. Afterwards, the lenses are fit onto the bodies, their focus fixed on infinity with tape, and the interference filter attached. Because they acquire ‘unconventional’ radiation, the automatic White Balance (WB) of both the Nikon D50\textsubscript{NIR} and D80\textsubscript{FS} ‘fails’. Therefore, this setting has to be predefined to allow highly reflective targets to be rendered approximately white. Afterwards, the aperture, shutter speed, and ISO value of these two devices is manually determined and fixed in the DSC. This is of the utmost importance because the internal light meters are not designed to cope with these wavelengths. The last DSC setting to be checked is the image format: as scientific aerial photography largely benefits from RAW imagery (i.e. the most ‘pristine’ data generated by the image sensor – see Chapter 7), and this image format is the only one that allows the originally captured Blue channel to be extracted, all DSCs are set to store NEFs (i.e. Nikon Electronic Format, the proprietary RAW format of the Nikon Corporation).

At this stage, the DSCs can be equipped with the Radio Controlled (R/C) units, after which they have to be mounted onto the photo bar. Now, a few test images can be shot to check all previously entered settings and verify the correct alignment of the optical axes. Clearing the memory cards is performed in a final formatting round. Before getting airborne, the photographer puts on fingerless gloves (which are beneficial to constantly retain a good grip on this construction) and synchronises the clocks of all three DSCs with the GPS unit (the latter positioned in the front of the aeroplane – Figure 11.4B). Once in the air, photographs are taken by pressing the two shutter buttons simultaneously, while looking with one eye through the viewfinder of the Nikon D50\textsubscript{NIR}. In essence, this largely boils down to performing the same operations an aerial archaeologist normally does when flying with one DSC, apart from the fact that zooming is made impossible.

11.4.3 Image Processing

After data acquisition, a necessary pre-processing phase is required. This starts with transferring and renaming the photographs, while the D200 frames are rotated 180° (to counteract this DSC’s orientation). Subsequently, files that not belong to any set of three images covering exactly the same scene are deleted. This processing step ends with the implementation of the correct WGS-84 coordinates into the Exif (Exchangeable image file format) metadata header of the NEF files (see also 7.3.3). All mentioned steps can be run in batch operations. This speedy processing is, however, not possible yet for the last, but crucial actions.
First of all, the Red channel is extracted from the fully demosaiced and gamma corrected 700 nm NEF file, as this channel is most sensitive to this radiation and contains least noise (for a detailed explanation of this RAW-related terminology and its implications, consider Chapter 7). Afterwards, this channel is saved as an uncompressed greyscale TIFF (Tagged Image File Format) file. Secondly, the Blue channel is extracted from the NIR frame, the latter developed the same way as the Red edge photographs.

Due to the fact that a digital image is only a matrix of Digital Numbers (DNs) ordered in a fixed amount of rows and columns, mathematical operations between different channels can readily be performed. However, in order to divide a Red edge pixel by the corresponding NIR pixel, the images have to be **coregistered**: i.e. geometrically transformed to allow them to be perfectly superimposed/aligned. In spite of all necessary precautions, small inequalities in scene coverage are nearly inevitable. In addition, the amount of pixels in both frames differs: 3008 pixels * 2000 pixels (i.e. six million pixels) and 3872 pixels * 2592 pixels (i.e. ten million pixels) for the Nikon D50\textsubscript{nir} and Nikon D80\textsubscript{FS} respectively. Currently, this coregistration is performed in AirPhoto (The Unkelbach Valley Software Works) on a one-by-one basis. Afterwards, the division of both data layers is performed in ENVI 4.3 (ITT Industries).

### 11.5 Archaeological Results

As previously mentioned, a flight has been performed to assess the possibilities of this low-cost multi-spectral tool in aerial reconnaissance. More specifically, imagery was acquired above the Potenza valley in central Adriatic Italy (\textit{Regione Marche}) on the 15th of May 2008 between 11.00 h and 12.00 h, yielding solar azimuth and elevation angles of approximately 160° and 65° (as calculated from [197]). Figure 11.5 shows the generation of this simple VI on the central part of the Roman city of \textit{Trea}. Tiles A, B, and C are JPEG-compressed (Joint Photographic Experts Group) frames as they would have been created by the Nikon D200, Nikon D80\textsubscript{FS}, and Nikon D50\textsubscript{nir} respectively.

The second row illustrates the results of the first two processing steps: all three frames are coregistered and the Red (E) channel extracted from the 700 nm frame, while the Blue channel (F) is extracted from the NIR frame (C). As explained, these operations are not performed on the JPEG or TIFF data that might be generated by the DSC, but on the demosaiced and gamma corrected NEF files (two latter two operations executed using dcraw, a free RAW decoder [184]).
Figure 11.5. Creation of the reflectance ratio image (tile I and H) and comparison with a simultaneously generated conventional colour photograph (G). By revealing previously invisible building structures in the central part of the Roman city of Treia (N 43° 19’ 05”, E 13° 17’ 34” – WGS84), tile H indicates the revealing power of this $R_{700}/R_{800}$ index.

At this stage, image E represents the 700 nm data (i.e. the numerator of the VI), while F shows the 800 nm record (i.e. the denominator of the VI). Dividing both images yields tile I. This output, which was created using ENVI and visualised with a linear 2 % stretch of the data, represents the $R_{700}/R_{800}$ VI, which clearly indicates chlorophyll-rich zones to be darker than non-vegetated zones (i.e. the ratio increases as leaf chlorophyll content decreases).

To compare this reflectance ratio image with the conventional colour image, the local contrast of both frames A and I was seriously enhanced and their central portion represented in tiles G and H respectively. This confrontation clearly indicates the VI to yield more archaeologically relevant data: whereas the large structures such as the Roman temple (1) and roads (2) are apparent in both frames, much smaller features – most likely representing walls of tabernae (3, 4, 5) – are only displayed in the VI output. This indicates that the stress induced response in the plants overgrowing them was still too weak at the time of acquisition to be detected by common aerial imagery. The VI, however, maximally exploited the dissimilar reflectance properties in the two narrow spectral bands.

In a second example, a combination of hydrographical marks and Roman building features gives again distinct signatures in the visible record (Figure 11.6A). The VI generated from this area (the southwestern part of the Roman town of Ricina) makes clear that one can not always expect this mathematical operation to yield spectacular new information, although the continuation of the hydrographical feature (1) is only obviously discernable in the reflectance ratio image (Figure 11.6B). When looking at the circles in the two insets, it also becomes clear
that the VI image shows some features to be more distinct and/or better delineated. Besides this slightly contrast-enhanced rendering of dissimilar crop characteristics, the drop in chlorophyll content is large enough in the central portion of the scene to generate a reflectance increase in both the Green and Red visual bands. Consequently, large parts of this field’s vegetation are chlorotic and thus appear yellow.

![Figure 11.6. Visible (A) and R\textsubscript{700}/R\textsubscript{800} image (B) of a southwestern part of the Roman town of Ricina (N 43° 19' 33", E 13° 25' 15" – WGS84).](image)

Apart from the $R_{700}/R_{800}$ ratio, the generated information can be used in several other VIs:

- $\left( \frac{1}{R_{700}} \right)$ [329];
- $\left[ \left( \frac{1}{R_{690-720}} \right) - \left( \frac{1}{R_{760-800}} \right) \right] R_{760-800}$ [330, 333].

The latter index results from the fact that the first part of this index is closely related to chlorophyll, but still much affected by leaf scattering. To get rid of this, the multiplication was added. Figure 11.7 illustrates the differences between these VIs and compares them with a conventional frame (D) and the proposed $R_{700}/R_{800}$ image (A). The large difference in visual appearance of all three VIs is striking, indicating they all show particular phenomena better than the others. Ideally, the final image interpretation is based upon both the visual frame and one (or more) of these mathematically generated products.

On the one hand, the reciprocal of reflectance at 700 nm (tile B) is not able to distinguish between the vegetated zones (1) and the bare soils (2 and 3), which indicates that all three have a very similar reflectance. On the other hand, Figure 11.7A (i.e. the proposed index) easily allows to discriminate the soil from the vegetation, but it is incapable to render the soil differently from the stone slabs (4). All those problems are solved in the output of the $\left[ \left( \frac{1}{R_{700}} \right) - \left( \frac{1}{R_{800}} \right) \right] R_{800}$ index (tile C). This image also looks ‘smoothest’ of all three VIs, but seems to have least discriminative power in relation to the archaeological crops marks (1).
Figure 11.7. (A) The reflectance image generated by $R_{700}/R_{800}$; (B) the result from $(R_{700})^{-1}$; (C) the image generated by $[(R_{700})^{-1} - R_{800})^{-1}] * R_{800}$; (D) the same scene imaged by an unmodified DSC. The imagery depicts the northwestern part of the Roman town of Ricina (N 43° 19' 37", E 13° 25' 07" – WGS84).

Strangely, it seems that none of the three indices could really reveal the tiny building features as good as the visible frame was able to. The clue to this is, however, provided by the $(R_{700})^{-1}$ image (Figure 11.7B), as it seems that the stress response was not entirely picked up in this band. Although other researchers already showed that (digital) monitoring through a narrow bandpass interference filter centred at 700 nm is very well possible and even enables ‘previsual’ stress indications [144, 148, 174], the last example reveals a few drawbacks of the current approach:

- first of all, the interference filter is dated. Consequently, its peak transmission is low, which compels longer shutter speeds. However, too long a shutter speed results in blurred aerial imagery. To counteract this, underexposure was needed, resulting in considerable amounts of image noise. In some cases, the noise masks the faint spectral signals of the vegetation;
- although manufacturers currently produce quality interference filters especially for imaging devices, this was not the case fifty years ago. Even though the filter currently applied transmits a suitable waveband, the optical characteristics do not allow to generate a very clear, pin-sharp photograph. It goes without saying that this effect adds up to the image noise;
- finally, it might not be coincidental that the best results are provided by the nearly perfect nadir photograph in Figure 11.5H (i.e. optical axis vertical to the earth’s surface) because the reflectance of the vegetation canopy is dependent upon the sun and sensor zenith/off-nadir and azimuth angle [439]. Although the view angle is less critical
for NIR imaging compared to visible imaging [358, 680] due to the multiple canopy scattering of NIR photons [485], the effects of view angle and solar angle on archaeological narrowband 700 nm imaging still have to be seriously assessed. However, it is known that the NDVI values are angular dependent, with a decrease in NDVI with increasing view angle [224, 225, 647]. These results already give major hints as to which flying approach will be most beneficial for this kind of multispectral imaging.

11.6 Drawbacks, Improvements, and Considerations

Although real-world examples and spectral reflectance data clearly underpin the ‘crop mark revealing power’ of this multispectral approach, some issues have to be dealt with to be more successful in its application. First of all, a modern, image quality bandpass filter is of the utmost importance. Even today, interference filters come in several qualities and their properties – often expressed in terms of scratch-dig ratio (S/D) and wavefront error – dramatically effect the overall cost and performance of the imaging system. Fitted with a superior optical piece, the image quality obtained with the Nikon D80 will be much better, while the increased transmission can generate less noisy images. Furthermore, the influence of focal length and ‘obliqueness’ of the aerial image on the wavelength shift should be assessed carefully to determine which photographic conditions must be met for the most successful image acquisition. Ideally, the filter should be fitted in front of the sensor instead of the conventional UV-NIR block filter, as this would largely eliminate the variation in the angle of incidence of the radiation hitting the sensor (while the photographer can still look through the lens).

Although the camera rig was suited for handholding operation, it remained difficult to perfectly align the three optical axes and attach all accessories needed. Therefore, a more convenient mount is currently created. This construction will maximally hold four DSCs, but mounting the DSCs will become much easier. Ideally, this camera rig should only hold identical DSCs with identical lenses, allowing for a straightforward and perfectly simultaneous triggering of the shutter (due to the equal R/C interfaces, maximum frame rate, burst rate, and buffer capacity – see 7.8.1 for an explanation of this terminology), while the image-processing chain, the button layout and memory card format are identical as well. Moreover, (post-) processing the final digital frames will be far more convenient, as all photographs contain an equal amount of pixels in width and height.

It is, however, also extremely important to improve the current post-acquisition workflow. To date, this laborious process of file development, channel extraction, coregistration and VI calculation seriously hampers the processing of large amounts of photographs. To be ever successful and widely implemented in large-scale archaeological surveys, this automation is of the utmost importance. It would be nice if this process could also include a stage at which the lens distortions are corrected, as the necessary coregistration would surely benefit from this.
Finally, much more imagery is needed to really reveal the potential of the proposed technique. This imagery should be acquired in various seasons, during different times of the day, and at varying geographical locations. The previous examples also seem to point in the direction of vertical coverage. Therefore, a second phase of testing (to be executed from May to September 2009) will compare the results from vertical block coverage to conventional oblique photography. Once the best approach to acquire 700 nm reflectance data is determined, this information can be used in several other VIs in an attempt to exploit this stress sensitive region even better across a wide variety of landscapes. Examples of possible indices are, amongst the ones presented in Figure 11.7:

- a Red edge NDVI \( \frac{R_{\text{NIR}} - R_{\text{Red edge}}}{R_{\text{NIR}} + R_{\text{Red edge}}} \) [336];

- the Visible Atmospherically Resistant Index (VARI) Red edge \( \frac{R_{\text{Red edge}} - R_{\text{Red}}}{R_{\text{Red edge}} + R_{\text{Red}}} \) [331];

- the Chlorophyll index \( \frac{R_{\text{NIR}}}{R_{\text{Red edge}}} - 1 \) [339, 340].

However, also the simultaneously generated NIR frames and visible information offers various possibilities to detect, compare and mathematically enhance archaeological anomalies that might otherwise go unnoticed. Consequently, the archaeological potential of modified and unmodified DSCs can be exploited more fully when their spectral properties are combined during reconnaissance flights. This way, systems featuring two, three or more DSCs offer archaeologists very affordable and manageable multispectral tools that can provide useful information on the vegetation’s physiological and morphological conditions to aid in the survey of the archaeological subsurface.

### 11.7 Discussion

A basic understanding of DSC principles and vegetation reflectance has fostered the development of a low-cost, three-DSC aerial multispectral system to aid archaeological reconnaissance and reveal moderate reflectance changes. Such systems are rather simple to construct and perfect for hand-held operation [254]. They can provide useful information on the vegetation’s physiological and morphological conditions and allow the comparison of multiple data sources to identify archaeological features and finally draw the archaeological picture as complete as possible. However, to increase the site discovery rate and tackle some of the biases in crop mark discovery, this approach needs major testing on archaeologically ‘bad’ soils (e.g. poorly drained soils which get high levels of rainfall) and get rid of the predominant flying strategy of ‘observer-directed’ survey to generate oblique photographs [596, 834], a point that was already raised by other scholars concerning aerial imaging in the visible domain [186, 198, 371, 372, 543, 596, 597]. In its current practice, observer-directed
narrowband and non-visible imaging will mainly enhance the record of known sites as visually imperceptible features will not make it into a photograph. Provided that the mentally, logistically, and practically challenging step of unbiased block cover is taken, the various multispectral approaches that were recently developed (see Part IV and Chapter 12) might finally show their real potential and prove to be low-cost but powerful alternatives for conventionally generated multispectral aerial information, but with a spatial resolving power that largely surpasses that of most data captured by various airborne or satellite-mounted sensors.

However, the future development of the proposed techniques can only occur if aerial archaeologists expand their knowledge base of the stress related information content in remote sensing data generated by a broad range of imaging sensors. As a consequence, the selection, acquisition, and (re)combination of different wavebands must always be guided by common sense and a solid understanding of spectral leaf and canopy properties, instead of trial and error attempts (as described by [37]). This way, the impact of aerial archaeology can only become more important.

The acquisition of imagery is, however, worthless without the archaeological interpretation. Even though a multispectral system allows the application of various processing techniques for the enhancement of particular spectral properties, the most sophisticated method to recognise, extract, and classify archaeological evidence will mainly remain visual detection and interpretation [55, 836], due to the particular indistinct and varying spectral nature of archaeological marks [65, 372].

11.8 Conclusion

From an archaeological point of view, there are several overarching challenges that must be dealt with in order to advance active aerial photography beyond today’s practice. After about hundred years of relative moderate progress and few technological improvements, coupling several modified and unmodified DSCs with a new flying strategy has the potential to dramatically change archaeological aerial reconnaissance, certainly when the imagery is combined with the remarkable advances in computer hard- and software, Global Positioning System (GPS) receivers, and the power of Geographic Information Systems (GIS).

The approach developed in this chapter makes use of two modified Digital Still Cameras (DSCs) that capture reflectance information in stress-sensitive spectral bands centred at 700 nm and around 800 nm. Division of both information sources yields a ratio which has shown its ability to detect the generic stress related Blue shift of the Red edge. As a result, this new data source allows to extend the archaeological information that can be extracted from conventional photographs that only take the complete visible waveband into account. In the current approach, the visible reflectance data is acquired by a third simultaneously operated DSC
because this information will in most cases remain essential for the reliable interpretation of the archaeological anomalies. Notwithstanding the infancy of this technique and the awareness that several hardware and workflow improvements have to be implemented, it is already clear that such a three-DSC rig offers archaeologists a very affordable and manageable multispectral tool that can capture whole parts of the landscape in several spectral bands with the ease of one single exposure, thus providing useful information on the vegetation’s physiological and morphological conditions to aid in the survey of the archaeological subsurface. In the near future, several oblique and vertical flights will further assess the potential of the system in various landscapes, and dissimilar operating conditions. This way, it is hoped that the results will allow to develop imaging techniques that are less dependent on the time of the year or day, and the angle at which the features are viewed. Ideally, they would make the observation of faint stress patterns better and reliable across space and time.

By exploiting the inherent spectral properties of current DSCs in multi-camera rigs, it is also hoped that some of the characteristic reconnaissance biases might (partly) be tackled, although a new flying practice will be inevitable. In case one or more issues militate against the effective use of blanket vertical photography, the proposed DSC combination can still be used to extract more meaningful information about individual sites that were recorded by the observer directed approach. So whether they are used for oblique or vertical photography, DSCs can be converted to convenient, low-cost alternatives for conventionally generated multispectral aerial data, but with a resolving power that even surpasses that of most panchromatic data generated from various airborne or satellite platforms. This way, a well-considered use of photographic equipment should be able to partly fill in the gap described by Hanson: “the current resolution of multispectral and hyperspectral data, whether satellite or airborne based, does not yet approach that of panchromatic imagery, making identification of more ephemeral remains, such as cropmarks, more difficult” ([372]: 49).
CHAPTER 12

The Attraction of the Unknown
Digital Near-UltraViolet Aerial Archaeology

The content of this chapter is submitted to Journal of Archaeological Science [ERIH (A); SCIE, SSCI, AHCI – IF 2007: 1.439]

Abstract

Even though the sun has its peak emission of ElectroMagnetic (EM) radiation in the visible waveband, it still produces a substantial amount of UltraViolet (UV) wavelengths (10 nm to 400 nm). However, the largest portion of this emitted energy is blocked by the ozone layer of the atmosphere, only allowing the Long Wave UV (also called NUV or UV-A, comprising the spectral band between 315 nm and 400 nm) to reach the earth’s surface. Additionally, sensors acquiring this part of the UV wavelengths must be operated from low altitudes to minimize the effects of strong Rayleigh scattering to which the UV radiation is subjected. Consequently, this waveband of the EM spectrum is rarely employed in aerial photography and its reflected portion only acquired in very specific applications.

Archaeological aerial NUV imaging can thus truly be seen as a completely unexplored research field. Although it is far from certain that this technique could enhance, let alone reveal new archaeologically related anomalies, this chapter discusses the practicalities of digital NUV imaging: from the modification of Digital Still Cameras (DSCs) and the choice of appropriate optics over the extremely important NUV interference filters to the focus and exposure compensation needed. The use of a remotely controlled platform will shown to be indispensable as the inevitable long shutter speed compels the use of such a very stable, unmanned aerial device. By presenting the very first aerial and archaeological digital NUV frames, the weaknesses and/or advantages over conventional visible imaging may be evident (notwithstanding the infancy of this approach).
12.1 UltraViolet Radiation – An Introduction

The principle that allows the Human Visual System (HVS) to observe objects and persons is based on those subjects’ reflection of visible light. Dark objects do not reflect much incoming light, whereas a shiny green apple principally reflects green light. However, this light is only one small portion out of the complete, so-called ElectroMagnetic (EM) spectrum radiated by the sun or other sources (like stars, lamps, etc.). The spectrum can be considered a continuum of varying EM waves, the latter described by the Scottish mathematician and physicist James Clerk Maxwell (1831-1879) as two oscillating magnetic and electric fields perpendicular to each other as well as perpendicular to the direction of propagation (Figure 12.1), whilst travelling at 299 792 458 m/s (denoted c) in vacuum [723, 804, 806].

![Figure 12.1. An ElectroMagnetic wave (adapted from [806]: Fig. 25-3).](image)

Being a wave phenomenon, EM radiation is distinguished by the length of its waves, called the **wavelength** (\(\lambda\)), as well as its **frequency** (\(\nu\)): a figure – expressed in Hertz (Hz) – that indicates the number of complete waves passing a certain point in one second and thus inversely proportional to \(\lambda\). No matter what portion of this broad spectrum is considered, they all obey the same physical laws and the relation \(c = \lambda \nu\) holds for each. This way, the wavelengths of the narrow visible waveband that range from approximately 400 nm (400 * 10^{-9} m) to 750 nm (750 * 10^{-9} m) correspond with frequencies comprised between 7.49 * 10^{14} Hz (749 THz) to 4.11 * 10^{14} Hz (411 THz).

Nonetheless, the complete EM spectrum consists of far more particular wavebands with characteristic frequencies and related wavelengths. To both sides of the visible band there is EM radiation which does not produce a visual sensation, more specifically gamma rays, X-rays, and UltraViolet (UV) rays with shorter-than-visible wavelengths (and higher frequencies), while InfraRed (IR) rays, microwaves, and radiowaves can be found in the long-wavelength, low-frequency region (Figure 12.2).
Although the sun generates abundant visible radiation with a peak emission around 480 nm (perceived by the HVS as green), it still produces a substantial amount of both IR and UV rays, respectively circa 40% and 10% of its total energy distribution when measured outside the atmosphere [450]. Whereas the former spectral range easily reaches the earth’s surface, the shortwave UV radiant energy is largely blocked by the ozone (O₃) layer of the atmosphere, making this waveband – comprised between 10 nm and 400 nm – far less obvious as a natural illumination source. From a photographic point of view, it becomes even more problematic as the complete transmitted UV range can not be used. To clarify this, it is best to look at the four particular UV sub-regions (see Figure 12.3) that are commonly defined by biologists within this wide waveband:

- **UV-A**: also called Long Wave UV or Near-UltraViolet (NUV) by physicists and ranging from 315 nm to 400 nm. This region, discovered by the Munich physician Johann Wilhelm Ritter (1776-1810) in 1801 [450, 494], is known as the glass transmission region and generally the only region of interest to conventional or digital (aerial) photographers;

- **UV-B**: also denoted Medium Wave or Middle-UV (MUV) and comprised between 280 nm and 315 nm. In biology, this portion is called the erythemal or sunburn region;

- **UV-C**: extending from 100 nm to 280 nm, this segment can also be subdivided into Short Wave or Far-UV (FUV: 200 nm to 280 nm), while the region’s higher frequency portion is part of the Vacuum-UV (VUV: 10 nm to 200 nm). The zone from 185 nm to 280 nm...
The Attraction of the Unknown

nm (thus largely equal to the FUV range) is also known as the bacterial or germicidal region;

- UV-D: encompasses wavelengths longer than 10 nm and shorter than 100 nm, thus taking the lower-wavelength portion of the VUV into account [450, 650, 771, 831, 832].

From Figure 12.4, it can be seen that the ozone layer of the earth’s atmosphere only transmits the incident UV-A region and additionally some UV-B radiation to about 295 nm – at which point the absorbance exceeds 99.9 % [206, 393, 450, 494, 771]. Moreover, clouds also reduce the NUV radiation reaching the earth. Consequently, only about 5 % of the sun’s radiant terrestrial energy is UV [38, 130, 450, 771], although the precise amount of NUV energy reaching the earth largely depends on the solar altitude: i.e. the elevation angle of the sun above the horizon which varies with the time of day, the period of the year as well as the geographical location.

However, the blocking of the sun’s short wave UV radiation is not accidental, because too much UV energy would be detrimental for all living creatures on earth – a fact that can be understood by considering the dual character of EM radiation. In addition to the aforementioned wave properties, EM radiant energy is known to exhibit particle-like behaviour, and can – building upon the statements and research of respectively Isaac Newton (1642–1727), Max Planck (1858-1947), and Albert Einstein (1879-1955) – be seen as a travelling bundle of photons (i.e. discrete energy packets) with energy levels that differ according to the wavelength. To calculate a photon’s quantum energy \( E \), Planck’s constant \( h \) (i.e. 6.626 *10^{-34} \text{ Joule/second or 4.136 *10^{-15} eV/sec}) must be multiplied by the frequency of the radiation: \( E = h \nu = hc / \lambda \) [723, 804, 806].

Due to this quantization, an NUV photon with a wavelength of 400 nm will always have 3.1 eV of energy, while 315 nm waves (the lower limit of the UV-A band) are characterised by photons with quantum energies of 3.9 eV. From these figures, it is obvious that the shorter UV wavelengths have a higher radiative energy. In the UV-B, which still reaches the earth to a moderate extent, photons’ energies are sufficient (from 3.9 eV till 4.4 eV) to cause changes in
the DNA (DeoxyriboNucleic Acid) and damage cells, causing sunburn, but also severe eye damage and forms of skin cancer [393]. Besides being harmful for humans, UV-B irradiance is also considered a significant stressor that can disrupt the growth and development for many plants [351]. Fortunately, the highly energetic and extremely damaging UV-C and UV-D are completely absorbed by the ozone molecules of the atmosphere (see Figure 12.4). Because UV-B is not really useful for aerial photography (see infra), the less energetic UV-A wavelengths shall therefore be the radiation under consideration in this chapter, even though the physical term NUV will be constantly used from now onwards.

12.2 NUV Imaging – Archaeological Applications

All natural and synthetic objects transmit, reflect, and absorb EM radiation in varying ratios. Besides the specific chemical and physical structure of these objects, this interaction is also dependent upon the kind of incoming wavelengths. By using invisible radiation, (digital) NUV photography tries to make certain features of the objects under consideration (more) distinct, because the UV reflection will most likely differ from the characteristic visible reflectance properties. Even if they exhibit a similar reflectance in the visible spectrum, objects thus often benefit from this uncommon means of imaging (which is also the reason for animals – notably honey bees – to have UV vision, as NUV reflectance can reveal complex patterns that function as nectar guides like the one shown in Figure 12.5A [250, 659]).

NUV photography can be applied in two distinct forms: reflected/transmitted or direct UV photography and UV induced fluorescence photography, a specific form of (photo)luminescence based on the spontaneous emission of new, visible radiation by an object during the time it is illuminated with pure, mostly artificially generated UV wavelengths [683]. Airborne aerial photography mainly applies the first form by monitoring the reflected NUV portion that is produced by the sun and to a certain extent transmitted by the atmosphere (or an artificial source located at the remote-sensor platform, although this approach is also used at night to stimulate luminescence of ground features). Imaging emitted NUV is almost impossible, as objects must attain at least a temperature of circa 1000 °K (i.e. 727 °C) before emission in this waveband starts [396]. Reflected NUV airborne imaging is, however, not as straightforward a task as visible light or Near-InfraRed (NIR) aerial photography (for a discussion of the latter, consider Chapter 8).

Besides the very modest portions of terrestrial NUV – certainly when compared to the other forms of optical EM wavelengths – UV radiation is seriously subjected to strong Rayleigh or molecular scattering ($\kappa$): i.e. the deflection of EM radiation by molecules and particles up to about a tenth of the wavelength under consideration (e.g. the air molecules or very small suspended air particles – called aerosols – as salt crystals, viruses, and smog parts). This process, which does not transform the energy of the photons but only alters their trajectory by absorbing the photon and then quickly emitting a new one in another direction, more
The attraction of the unknown seriously affects the shorter wavelengths as it is inversely proportional to the fourth power of the wavelength \(K = \lambda^{-4}\). Consequently, the short-wavelength blue and violet light is more likely to be dispersed in all directions than other portions of the visible spectrum, giving the clear sky its dominantly blue appearance [279, 393, 652]. Logically, this strong wavelength dependence of Rayleigh scattering – named after the English physicist John William Strutt (1842-1919) who was known as the third Lord Rayleigh – causes a lot of diffuse terrestrial UV radiation. Hence, the amount of the NUV energy coming directly from the sun generally fails to dominate the diffuse NUV sky radiation received by the earth’s surface [450]. From a remote sensing point of view, such a situation is detrimental, because this large sky component (perceived as haze in the visible domain) causes a reduction in visibility, a loss of detail, and a diminution of contrast [652]. Therefore, any sensor acquiring the UV part of the EM spectrum must be operated from low altitudes (preferably below 300 m [206]) to minimize this masking effect, which increments with increasing flying height [41, 235].

Thirdly, the NUV reflectance of most ground surfaces is extremely low (5 % to 10 %), certain sands and snow being a few exceptions [235, 236, 353, 450]. Thus, it is logical that the NUV waveband is rarely employed in remote sensing – as also remarked by Estes and Holz [278] and Holter et al. [396], but even more recently by Gruzdev et al. [353] and Lillesand et al. [482] – leaving the imaging of its reflected portion only to be used in a few, very specific applications: e.g. detecting oil spills on water and dry land using passive [325, 353, 709, 800] or active...
sensors [53, 583, 799, 803], finding white animals on ice and snow [466, 467, 468, 469], determining the roughness of the sea state in the open ocean [692] or camouflage detection in the arctic region [41].

In archaeology, NUV photography is rarely used but for the occasional examination of subjects not (or less) responsive to visible or NIR photographic methods. In those cases, UV fluorescence photography and to a lesser extent direct NUV photography might reveal faded inks [90, 459, 520, 605], altered documents or engravings [252], can aid in the study of ceramics ([241] – Figure 12.5B), and excavation profiles [115, 242], while it can also be rewarding to detect specific pigments [505, 506, 676] as well as traces of restoration on tapestries [481], textiles [49, 768], paintings [505, 506, 508], and sculptures [241, 252].

Due to the aforementioned reasons, direct aerial NUV photography has never been used in archaeological research, although Gibson [325] states it might be useful in revealing soil marks, but no further reference or illustrative material proves this point. Therefore, rather than being additional to the existing literature or offering an exhaustive overview of the archaeological potential and all related drawbacks of aerial NUV photography, the data presented here must be seen as an account of the first attempts to visualize archaeologically related anomalies and structures using these shorter-than-visible wavelengths, without having the presumption to be completely conclusive.

12.3 NUV Image Acquisition

The application of film-based UV photography always was a tedious task, but the advent of digital photography simplified the imaging process, with far more control over the final outcome. To use the digital reflected NUV method from the air, several key components were needed: a Digital Still Camera (DSC) – i.e. a camera equipped with both a digital image sensor for capturing photographs and a storage device for saving the obtained image signals in a digital way [760] –, an NUV lens, and an NUV pass filter.

12.3.1 Modified DSC

To capture the reflected NUV from the earth’s surface by digital means, an off-the-shelf Digital Single-Lens Reflex (D-SLR) was modified. When buying such a DSC, the imaging sensor – which is usually of the CCD (Charge-Coupled Device), CMOS (Complementary Metal Oxide Semiconductor) or Foveon X3 type – is covered with an UV-IR cutoff filter that allows visible light to pass and which matches the sensitivity of the HVS. By implementing this filter, the camera manufacturer ensures that the sensor’s array of photosensitive detectors – called photodiodes – will generate a photographic signal by mainly taking the visible EM radiation into account. This approach is of the utmost importance, as common silicon sensors are
inherently sensitive to NUV (Figure 12.6) and especially NIR, whose imaging contribution would be detrimental for image sharpness (all wavelengths have a different focus), exposure accuracy (DSC light meters are not calibrated for invisible radiation), and true colour reproduction.

Although such block filters come in various qualities (i.e. a number of DSCs still have some considerable UV and/or NIR response straight-out-of-the-box), their aim is equal: excluding these invisible wavelengths from reaching the sensor as much as possible. By removing this filter, the DSC consequently becomes sensitive to a range of wavelengths exceeding the initial small visible band (Figure 12.7A). Because the spectral response of common silicon detectors found in the current DSCs decreases quite rapidly towards shorter wavelengths, the effect of such a modification is most noticeable when imaging the NIR waveband. Notwithstanding, the DSC’s responsivity to NUV radiation increases as well, although its sensitivity is far more modest in absolute terms (for an in-depth overview of sensor sensitivity and DSC modification, see [373] or Chapter 8).

Selecting an appropriate D-SLR to convert for NUV imaging is as important as the conversion itself, because not all sensors are equally sensitive to this radiation and particular lenses will only fit onto bodies of specific brands. The DSC used for the presented NUV data acquisition is a converted Nikon D80 (hereafter denoted as D80_{FS}), known to have a very poor NUV response as delivered from the factory (Figure 12.7B). The choice of this D-SLR was driven by the fact that it is built around a CCD sensor, which is known to be far more sensitive to NUV than a CMOS.
imaging sensor [373, 689] (see also Figure 12.6). Moreover, D-SLRs of this brand are also used for archaeological NIR imaging (e.g. [776, 781] and Part IV), which makes buying and changing lenses easier as all hardware is centred around only one bayonet system. Finally, most of the UV dedicated lenses currently made all have a Nikon F-mount (see Table 12.1), since Nikon cameras are used very widespread in science and industrial applications.

![Figure 12.7. (A) Wavelength versus Quantum Efficiency for the Kodak KAF-8300 (adapted from [261]: Fig. 5); (B) spectral response of the unmodified Nikon D80 (adapted from [407]).](image)

By replacing the internal filter package with a specific clear glass or fused silica window that largely transmits NUV (a modification executed by a specialised company such as LifePixel and LDP LLC in the USA or Optic Makario in Germany), the D80FS was enabled to take wavelengths from the NIR to about 320 nm into account [275, 373]. Even though the DSC’s individual channels have not yet been thoroughly spectrally characterised, its final NUV response was tested by photographing a neutral test target several times. Compared with the unmodified Nikon D70s – a D-SLR that is renowned for its decent out-of-the-box NUV response [677] – the Red, Green, and Blue channels of the converted D80FS were on average 1.6 photographic stops (i.e. 3.0 times), 0.9 stop (i.e. 1.9 times) and 0.4 stop (i.e. 1.3 times) more NUV sensitive, indicating the usefulness of the modification. While special, far more responsive digital back-thinned UV sensors exist, they are not implemented in current off-the-shelf D-SLRs, hence making them unsuited in terms of low-cost and flexible airborne NUV imaging.

### 12.3.2 UV Transmitting Optics

Equipping a DSC with an UV capable lens is less straightforward than might be expected. Because optical glass does not generally transmit much UV and mostly becomes opaque to EM radiation below 350 nm [38, 206, 252, 368, 369, 505, 506, 650, 651], normal photographic lenses are usually impractical for NUV imaging (see the Zeiss Topar transmission curve in Figure 12.8). Moreover, optical cements can have a very high UV absorption and the coatings on the lens elements might additionally decrease the UV transmittance [506, 650, 651, 733].

However, expert UV lenses with special quartz and fluorite elements for sub-350 nm imaging do exist (Table 12.1), although expensive. Unfortunately, none of these exist in wide-angle variants. Fortunately, some older, rather exotic glass lenses still have a useful UV transmission
to about 350 nm (and sometimes even below – see Table 12.1), often due to their simple optical design with few lens elements and the absence of cemented or multi-coated elements [677]. Such a lens is, for instance, the discontinued Novoflex Noflexar 35 mm f/3.5. This lens is characterized by a good visible transmission, but also known for its excellent UV performance ([678], and Schmitt, 2007: unpublished research), having a cut-on transmitted wavelength of about 330 nm (Figure 12.8; Table 12.1). However, using the Noflexar necessitates a mechanical internal modification to make it fit the Nikon D-SLRs and allow for infinity focus in aerial NUV photography.

Table 12.1. This enumeration lists most expert optics and some glass lenses suitable for direct NUV recording with DSCs (adapted from [690]).

<table>
<thead>
<tr>
<th>Lens</th>
<th>Max. f (mm)</th>
<th>Max. aperture</th>
<th>Mount</th>
<th>Focal shift</th>
<th>Still produced</th>
<th>Cutton (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert lenses with quartz and fluorite elements</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asahi Pentax Ultra-Achromatic-Takumar</td>
<td>85</td>
<td>f/4.5</td>
<td>M42</td>
<td>No</td>
<td>No</td>
<td>220</td>
</tr>
<tr>
<td>Carl Zeiss UV-Sonnar</td>
<td>105</td>
<td>f/4.3</td>
<td>Hasselblad</td>
<td>No</td>
<td>No</td>
<td>215</td>
</tr>
<tr>
<td>Coastal Optical Systems UV-VIS-IR Macro</td>
<td>60</td>
<td>f/4</td>
<td>Nikon F</td>
<td>No</td>
<td>Yes</td>
<td>290</td>
</tr>
<tr>
<td>Coastal Optical Systems UV-Micro-Apo</td>
<td>105</td>
<td>f/4.5</td>
<td>Nikon F</td>
<td>No</td>
<td>Yes</td>
<td>250</td>
</tr>
<tr>
<td>Nikon UV-Nikkor</td>
<td>105</td>
<td>f/4.5</td>
<td>Nikon F</td>
<td>No</td>
<td>No</td>
<td>220</td>
</tr>
<tr>
<td>Nikon Tochigi UV</td>
<td>105</td>
<td>f/4.5</td>
<td>Nikon F</td>
<td>No</td>
<td>Yes</td>
<td>220</td>
</tr>
<tr>
<td>Expert lenses with quartz elements</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asahi Pentax Quartz-Takumar</td>
<td>85</td>
<td>f/3.5</td>
<td>M42</td>
<td>Yes</td>
<td>No</td>
<td>200</td>
</tr>
<tr>
<td>Carl Zeiss UV-Planar</td>
<td>60</td>
<td>f/4</td>
<td>Leica M</td>
<td>No</td>
<td>No</td>
<td>300</td>
</tr>
<tr>
<td>Carl Zeiss Jena UV-Objektiv</td>
<td>60</td>
<td>f/4</td>
<td>Exakta</td>
<td>Yes</td>
<td>No</td>
<td>300</td>
</tr>
<tr>
<td>Wol lensak 4&quot; U.V. Anastigmat</td>
<td>102</td>
<td>f/4.5</td>
<td>M30</td>
<td>Yes</td>
<td>No</td>
<td>180</td>
</tr>
<tr>
<td>Lenses with NUV transmissive glass elements</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nikon EL-Nikkor</td>
<td>63</td>
<td>f/3.5</td>
<td>Leica M39</td>
<td>(Yes)</td>
<td>No</td>
<td>350</td>
</tr>
<tr>
<td>Novoflex Noflexar</td>
<td>35</td>
<td>f/3.5</td>
<td>Nikon F-, Exakta &amp; M42</td>
<td>(Yes)</td>
<td>No</td>
<td>330</td>
</tr>
<tr>
<td>Rodenstock UV-Rodagon</td>
<td>60</td>
<td>f/5.6</td>
<td>Leica M39</td>
<td>No</td>
<td>No</td>
<td>300</td>
</tr>
<tr>
<td>Rodenstock UV-Rodagon</td>
<td>80</td>
<td>f/5.6</td>
<td>Leica M39</td>
<td>No</td>
<td>No</td>
<td>300</td>
</tr>
<tr>
<td>Rodenstock UV-Rodagon</td>
<td>150</td>
<td>f/5.6</td>
<td>M50</td>
<td>No</td>
<td>No</td>
<td>320</td>
</tr>
<tr>
<td>X50</td>
<td>50</td>
<td>f/2.8</td>
<td>Nikon F &amp; M42</td>
<td>(Yes)</td>
<td>(Yes)</td>
<td>310</td>
</tr>
<tr>
<td>X135</td>
<td>135</td>
<td>f/2.8</td>
<td>Nikon F &amp; M42</td>
<td>(Yes)</td>
<td>(Yes)</td>
<td>320</td>
</tr>
</tbody>
</table>

12.3.3 Bandpass Filter

Thirdly, a 2" Baader Planetarium U-Filter (also called Venus Filter and based on Schott UG11 filter glass) is mounted onto the lens using some step rings and a Nikon AF-1 Gelatine Filter Holder which allows the filter to flip up/down for visible focussing. This interference filter – which exhibits a high NUV transmission from about 320 nm to 390 nm (Figure 12.8) – is by far the most important part of the complete imaging system. When removing the internal UV-IR cutoff filter, the NIR response of the D-SLR increased hugely (as can be seen in Figure 12.7A),

275
making this very dense Visual+NIR blocking, UV transmissive filter of the utmost importance in digital UV imaging. Without it, the final digital frame would still be heavily NIR-contaminated, rendering the resulting images useless, as is the case when applying most classic UV band-pass filters: Schott UG1 and UG11, B+W 403, Kodak Wratten 18A, Nikon FF, Corning Glass 5840, 9863 or Hoya U-330, U-340 (Figure 12.8), and U-360. Although these non-interference filters completely block visible light, they substantially leak NIR [38, 275, 506, 677, 733, 831, 832], spoiling the pure NUV picture aimed for. In contrast, the Baader Venus Filter only leaks less than 1 % NIR radiation (Figure 12.8), an amount that is far too low to perceivably contaminate the digital NUV record.

![Figure 12.8. Spectral transmission of several mentioned optical elements. Compare the huge NIR-leakage of the Hoya U-340 filter with the (almost) pure NUV transmission of the Baader Venus Filter. The NIR transmission of the Novoflex Noflexar f/3.5 35 mm can be judged against the transmission of the Zeiss Topar f/2 57 mm, one of the best aerial lenses on the market and exemplary for the NUV transmission of conventional multi-coated lenses. The outcome of both lenses combined with the Venus Filter is also plotted, showing the Noflexar system to be far superior in terms of NUV transmittance (spectral measurements by K. Schmitt – macrolenses.de).](image)

### 12.3.4 Helikite Aerial Photography

Because even a converted D-SLR still is moderately sensitive to NUV (about the same sensitivity as the analogue silver halide emulsions at 380 nm, although the latter generally sensed wavelengths to about 360/350 nm [241, 373] and the optics as well as the interference filter absorb NUV to some extent (see Figure 12.8), imaging the very small quantities of reflected NUV makes long shutters speeds inevitable (some eight to ten stops longer exposures when compared to visible photography). Although this currently makes digital aerial NUV photography practically unusable from a low-flying aeroplane, it can be applied when an appropriate, stable aerial platform is available. The latter was previously constructed (see [781] and Chapter 5) to allow Radio Controlled (R/C) digital NIR and low level aerial archaeology by means of a Helikite: a construction using kite wings attached to a helium balloon (Figure 12.9). Inaugurated HAP or Helikite Aerial Photography (see Chapter 5), this form of remote sensing now proved essential in the acquisition of the aerial NUV imagery.
12.3.5 Image Processing

Once the imagery is captured, the digital frames have to undergo some post-processing. Although NUV imagery should by definition not have any colour, the Colour Filter Array (CFA, see 8.4.2) and processing software inside a DSC will always create colours from the signals generated by the incoming UV photons. In most DSCs, NUV imagery will look pinkish or suffer from a reddish cast (as visible in Figure 12.12B and Figure 12.15C), because the image data generated by the Red channel generally has the highest NUV response. The same phenomenon can also be attested in the D80FS. After averaging several digital frames taken from a neutral test target, the D80FS’s Red channel mean response was calculated to be 4.8 times (i.e. 2.3 stops) and 6.4 times (i.e. 2.7 stops) higher than the Green and Blue channel response respectively.

Because it yields the highest useful response, the digital signal generated in the D80FS’s Red channel is preferentially used as greyscale, pure NUV frame. This again points to the necessity of blocking all NIR radiation as much as possible, since any transmitted NIR wavelengths – to
which the D80<sub>F5</sub> is very responsive – will also be recorded in this Red channel, leading to worthless contaminated NUV frames.

Besides using the Red channel only, it is also possible to take all three Red, Green and Blue channels into account, although one will certainly have to alter the **White Balance (WB)** setting of the digital photograph to get rid of the dominant reddish appearance. No matter which of the two approaches is used, both do largely benefit from a RAW workflow, as does any scientific photography (see Chapter 7). In case only JPEG output is possible, it helps to set the DSC’s internal WB to a very low value (e.g. 2000 °K) to minimize the contribution of the Red channel.

In addition to strong colour casts, NUV imagery coming straight off the DSC is also typically low in contrast due to the aforementioned Rayleigh scattering. After extracting the Red channel, the resulting NUV frame thus benefits from both a histogram stretch as well as some local contrast enhancement. For the visual appearance – and certainly not for interpretative purposes – noise suppression is advisable in the shadow areas of the frame.

### 12.4 Archaeological Results and Explanations

After setting the D80<sub>F5</sub> to manual exposure (1/80 s, f/8, ISO 800) and fixing the focus to just-before-infinity (as for all but a few specialized lenses – see supra –, focusing the lens closer than for visual photography is needed to counteract some focus shift when shooting NUV), it was possible to acquire some archaeological aerial imagery by means of the Helikite platform. The results presented here are the only two study objects photographed from the air so far: a visibly attested crop mark (Figure 12.10) and an excavation area (Figure 12.12), both part of the Roman coastal town **Potentia** (central Adriatic Italy, *Regione Marche*: 43° 24' 53" N, 13° 40' 14" E – WGS84) and serving as test cases for the NUV characteristics of vegetation and soils.

#### 12.4.1 Vegetation

In the first example (Figure 12.10), the crop mark that reveals the presence of **Potentia’s cardo maximus** in the visible domain also proved to be discernable in the NUV records (which are rather limited in ground area covered, due to the focal length of the lens, the format of the sensor, and the Helikite’s altitude). Whether this result means that all crop marks can also be spotted in the NUV, is, however, uncertain. When examining the visible frame, it seems the archaeological feature is mainly visible due to a scarcely vegetated surface, rather than a stress induced discoloration of the grass. This hypothesis is also supported by the very sparse information on NUV reflectance of vegetation.

In the current scientific literature, UV absorption and reflectance studies of natural and artificial surfaces are still largely lacking in the field of earthly remote sensing, notwithstanding the
common application of NUV imaging in several, non-aerial reconnaissance research fields as astronomy, medicine, biology, and atmospheric physics. Some forty years ago, this issue was already raised by Cronin et al., who indicated that UV is most likely the least explored and utilized band among the conventional sub-divisions of the EM spectrum, even stating that “…research in the ultraviolet region has been neglected by terrestrial scientists. The spectral properties of vegetation and soils and of undisturbed weathered natural surfaces of rocks and sediments, in situ, are unknown” ([206]: iii). Nowadays, most soil and vegetation spectroscopy studies still mainly focus on visible and NIR EM wavebands due to reasons as instrument sensitivity and cost. As a result, it remains difficult to present a large sequence of objective figures on the particular UV spectral properties of vegetation species and soils.

Figure 12.10. Comparison between a visible record of a crop mark (left) at the site of Potentia and its rendering by NUV wavelengths (right).

Considering the NUV spectral properties of vegetation, the lack of information can in particular also be attributed to the fact that leaf transmittance and reflectance of Photosynthetically Active Radiation (PAR: 400 nm to 700 nm) had to be completely understood to model the basic photosynthetic activity [351]. Therefore, certain assumptions have to be made to accurately model the interaction of NUV radiation in the vegetation canopy, although some scholars did explore the UV region to some extent. In general, the spectral signatures are less distinct in the NUV when compared to the NIR or visual domain, for both dissimilar types of vegetation and different states of vigour. As an example, Lovelock and Robinson [491] could state that the reflectance of mosses is almost similar (i.e. extremely low) for all different species. In another study, Grant et al. [351] proved a good linear correlation between PAR and NUV reflectance \( R = 0.97 \) for leaves of twenty different deciduous trees. Notwithstanding, the tree leaves absorbed far more radiation in the NUV region compared to the visible or NIR bands (Figure 12.11A), with reflectance values mostly below 10 % (often around 5 %) for both sides of
the leaves [351]. Similar conclusions were drawn by Allen et al. [29], Caldwell [130], Filella and Peñuelas [290], Gates et al. [314], and Gausman et al. [317]. Because most plants transmit very small amounts of NUV, the small reflectance values are due to the large absorption of this short-wave radiation, which first starts in the cuticula or the waxy coverings of the leaves' outermost cell layer (i.e. epidermis – see Figure 11.1B), while organic compounds such as flavonoids – situated directly underneath the cuticula – strongly absorb UV radiation penetrating further into the leaf [130, 317, 350, 351]. For several leaf trees, the small NUV reflectance even seemed to decrease with increasing wavelength [351].

Figure 12.11. (A) Spectral reflectance of several leaf surfaces ([351]; Fig. 2a); (B) absorbance of a vegetation sample in dried and fresh conditions (adapted from [96]; Fig. 2a).

Additionally, Figure 12.11B shows that the differences between dry and fresh vegetation are also far less distinct compared to the visible and NIR bands, with the smallest EM radiation absorption by dry vegetation in the UV region [96]. Consequently, one might already conclude that NUV will not be particularly suited for detecting stress marks in plants. The same conclusion can be drawn when looking at chlorophylls and other plant pigments. Although they dictate the appearance of healthy and stressed green plants in the visible domain, NUV reflectance is virtually insensitive to leaf greenness due to the very low pigment absorption in this range [314]. The initial observation thus seems to be in accordance with these spectral studies: it is most likely that the lighter crop mark traces in the NUV frame of Figure 12.10 were generated by the higher reflective ground visible between the dark, but scarce vegetation.

Figure 12.12. Comparison between a visible frame (A), unprocessed RAW NUV image (B), and enhanced NUV photograph (C) of an excavation area at the Roman town of Potentia. The large dark blob in the upper right corner of (B) and (C) is the shadow cast by the Helikite.
12.4.2 Soil

A comparison of the visual image (Figure 12.12A) with the processed NUV frame (Figure 12.12C) indicates the abundant details visible in the greyscale image when compared to the conventional colour photograph, although the latter was taken with a superior lens and a much higher shutter speed. This reveals the higher spatial resolving power of NUV photography, to be attributed to the shorter wavelengths used for imaging. Due to diffraction of EM waves (i.e. the spreading out of waves when passing small apertures and their bending around tiny objects), the EM waves originating from a single point will never focus to a perfect, infinitesimal small point when passing a lens, even if the latter did not suffer from any aberration [490, 651, 846]. Instead, a so-called Airy pattern, denoted in honour of the English astronomer sir George Biddell Airy (1801-1892), is created in the focal plane. This pattern consists of a bright circular central patch (the Airy disc) and a series of dimmer concentric rings, each ring separated by a circle of zero intensity (Figure 12.13A and Figure 12.13B).

The finite diameter radius \( r \) of this Airy disc equals 1.22 \( \lambda N \), \( \lambda \) being the radiation’s wavelength (e.g. 380 nm) and \( N \) the \( f \)-number or relative aperture of the imaging lens (e.g. 2.8, 4 or 5.6). From the equation, it can be seen that smaller wavelengths – and large lens apertures – theoretically yield smaller Airy disks in the DSC’s focal plane (as an example, a lens with an aperture of \( f/4 \) will yield an Airy disk of 5.4 \( \mu \)m for 555 nm and 3.4 \( \mu \)m for 350 nm). Specifically the size of these central bright areas is crucial for the amount of resolved detail, because the Rayleigh criterion (named after the aforementioned Lord Rayleigh) states that two neighbouring small objects can be spatially resolved as long as their Airy discs do not come closer than half their width (see Figure 12.13C).

Additionally, if the diameter of an Airy disk becomes too large with respect to the size of the photodiodes of the DSC’s sensor (which are also only a few micron in width and length), it begins to have a negative visual impact on the digital image. Consequently, the smaller the Airy discs, the higher the theoretical spatial resolving power of an imaging system [490, 651, 846]. This way, the wave nature of EM radiation and the resulting physical principle of
diffraction set a fundamental resolution limit, but will allow closely spaced points to be spatially resolved better when using short wavelengths as NUV. Moreover, smaller apertures (e.g. f/11) can easily be used in NUV imaging before any diffraction effects reduce the system’s resolving power.

Besides the increased spatial resolving power, Figure 12.12 indicates the contrast difference between the white gravel and the surrounding soil matrix to be more pronounced in the NUV. Again, the interpretation of this result is to some extent limited by the existing reflectance studies. Besides, the reflectance of soil is far more complex than vegetation, because it is not an ideal system and the link between particular soil attributes and measured soil chemistry is difficult [412]. An interesting starting point is the research by Doda and Green [235, 236], who took spectral measurements in the UV waveband (more specifically from 290 nm to 400 nm) from forest canopies, farmland, oceans, and different soils.

In contrast to other approaches, reflectance readings were taken at several altitudes using an airborne platform, subsequently allowing the separation of the atmospheric contribution and the calculation of the theoretical reflection at the ground. From those data, it is obvious that NUV reflectance from different kinds of sands can be largely different, illustrated also by Figure 12.14A. Doda and Green for instance measured desert sand and gypsum sand, with a mean reflectance in the 302 nm to 400 nm of respectively 6.4 % and 42.0 % at an altitude of 300 m (yielding slightly higher values compared to the reflectance at ground level due to several atmospheric effects). Furthermore, an increase in NUV reflection with increasing wavelength could be witnessed for sand, snow, lava, and water [235, 236].

A possible explanation for these differences can be found in their different particle size, which is known to be of main influence on the reflectance of soils [95], just as is the case for minerals [206]. This is not only applicable to the UV: in almost whole of the 225 nm to 2550 nm band, the reflection of sandy soils is highest for fine particles, with the greatest absorption for coarse soils, but the absorption difference between coarse and fine soils was slightly greater in the NUV and Blue compared to the remaining of the visible range [95]. Differences also occur

![Figure 12.14.](image)

Figure 12.14. (A) NUV, visible, and NIR reflectance spectra for six different soil samples (adapted from [412]: Fig. 4); (B) absorbance of a soil sample in wet and dried conditions (adapted from [96]: Fig. 2b).
The Attraction of the Unknown

between wet and dry sandy soil samples (Figure 12.14B): although the lowest absorbance change with regard to moisture content was attested in the NUV at 340 nm [97], the distinction between both wet and dry soils is almost equal to the attested difference in the visual domain, with a decreasing absorption in dry soils [95, 96, 97]. Moreover, the NUV absorption in both wet and dry conditions is mostly very strong (circa 80 % to 85 %), and much higher than the absorption of EM radiation in the visible and NIR bands [95, 96, 97] (Figure 12.14B).

A similar picture is created by rock formations. Even though the divergence between several rocks is commonly lowest in the NUV and their general reflectance is – as for most surfaces – lowest in the NUV compared to NIR and visible radiation (between 15 % to 30 %), Cronin et al. [206] reported that the reflectance difference between rock with rounded and angular grains was greatest in the NUV, while both Dorwin [242], and Glässer et al. [341] proved the usefulness of terrestrial UV photography in a geological context and archaeological excavation respectively. So it seems that soils and rocks might benefit from any possible aerial archaeological NUV approach (as was indicated by Figure 12.12).

Moreover, even if the soil reflectance differences in the NUV range are not very pronounced, the general character of the image often makes the faint dissimilar characteristics more prominent. As an example, consider Figure 12.15. On the left side (A+B), visual records of two types of soil with surrounding grass are displayed, while the pictures on the right (C + D) visualise the same scene in the NUV. The upper row (A + C) shows the photographs as they come straight from the DSC (the Nikon D70s and D80F respectively). Both insets display the values of the Red, Green, and Blue components on the white line in the middle of the frames. Besides the dominance of the Red channel in the NUV image (C), the plot indicates that the pixels of the visual frame span a whole range of intensity values, while the values of the NUV’s Red channel are spread to a lesser extent. After a histogram stretch and subsequent S-shaped tonal redistribution were applied to enhance the contrast, the RGB intensity plots in (B) and (D) show that the channels attain values from very dark to very bright on the intensity scale. However, the values’ spread is far more pronounced in the NUV frame (D), indicating a higher local contrast between the pixels.

The same is indicated by four samples taken in both images: one in the grass, a second in the darker soil, another in the lighter soil and the fourth in a small shadow area of the darker soil. The values that accompany the samples were obtained after converting the frames to the CIE 1976 L*, a*, b* colour model – also known as CIELab or simply Lab. This three-dimensional colour model (i.e. an abstract mathematical system which describes the way colours can be represented) is the most complete model used conventionally to describe all possible colours discernable by the HVS [305, 405, 704]. It uses three primary components, allowing the separation of the chrominance information (which is represented in the independent a* and b* channels) from the lightness component (notated L*). The latter can range from 0 (black) to 100 (white) and is thus a very convenient way to arrange colours along a scale from dark to light – just as luminance, intensity, value, and brightness are used in other colour models [405]. Moreover, all 101 possible values do not correspond with equal luminance steps, because L* is
a perceptually uniform scale. This means that the appearance of every single $L^*$ value is just noticeably different from the neighbouring values. Reading out this perceptual lightness component thus permits to work out the amount of incident light reflected from the object under consideration [741], while reflectance differences can be quantified according to the working principles of the non-linear HVS (see also 7.5.3).

From these figures, one can notice that all surfaces reflect about the same in the NUV (even the shadows become brighter due to the substantial amount of sky radiation that also illuminates portions which are not reached by direct NUV radiation), while the brightness variety in the visual frame is much larger: the samples range from 13 % to 77 % versus 20 % to 40 %. Consequently, the soil difference in the contrast enhanced version (B) of the visual frame is only stressed by an increase of a few percentages in the $L^*$ channel: i.e. from an eight value increase (62 % to 74 %), to a difference of eighteen units (62 % to 80 %). In the contrast enhanced Red channel of the NUV frame (D), the differences in the soil become much greater: i.e. from 55 % to 82 %.

Because these values are based on the human perception of lightness (often used interchangeably with the terms brightness or luminance, although they have sensu stricto not the same meaning [305, 405, 640]) and the HVS is far more sensitive to changes in luminance
information than chrominance data [119, 285, 404, 484, 704], one can distinguish the lighter soil from the darker component only about half as much in the visual frame compared to the NUV frame (18 % versus 27 % or a relative difference of 50 %). This result is solely due to the fact that the initial minimum and maximum values in the frame were very close to each other, allowing a contrast stretch to make the small reflectance differences of the soil far more evident. Notice also the same appearance between the darker soil and the grass, due to the fact they apparently reflect equally in the NUV.

This example does show digital NUV to be often beneficial over the common visual approach, although it might not be observable when using the imagery straight from the DSC. However, the complete process only involves the extraction of the Red channel from the RGB image (an operation that already yields a pure intensity frame), after which a simple contrast stretch can be applied to this greyscale image. Besides, reflectance recording in the NUV could also be fruitful in case one is dealing with ferriferous soils, as Islam et al. [412] could attest the highest correlation between reflectance spectra of different soil samples and their free Iron (Fe) quantity in the NUV at 382.4 nm, while most of the other physical and chemical soil properties (e.g. clay content, pH) showed the highest correlation in the NIR and MIR [412, 628].

Finally, imagery acquired in the 1970s using a scanning system capable of detecting reflectance in the 320 nm to 380 nm range has shown that also other objects potentially benefit from airborne NUV imaging [396]. From Figure 12.16, which depicts the Hayward fault area (California, USA), it is apparent that the roads and other man-made structures contrast a lot more with the surrounding area in the NUV, while even differences in the appearance of the terrain are visible [396]. Moreover, when assembling a false colour image by combining imagery acquired in the same NUV waveband with two simultaneously recorded visible wavebands (Blue and Green), subtle tonal distinctions not apparent on panchromatic, colour, pure and false colour NIR imagery could be presented [397].

Figure 12.16. Hayward Fault (California, USA) imaged in the NUV and Blue waveband from 600 m altitude ([396]: Fig. 56).
These results thus all indicate that aerial NUV photography – although not as straightforward as other forms of optical imaging – might often be beneficial for archaeological soil related research. Therefore, digital NUV aerial imaging (most likely by use of the Helikite) will still continue to be executed, as far more imagery is needed of dissimilar features in different circumstances before the real benefits of archaeological NUV imaging can be accurately outlined and quantified.

### 12.5 Final Thoughts

All terrestrial material such as soils, rocks, sediments, vegetation, and other organic matter mostly have very specific spectral reflection, transmission, absorption, and emission characteristics. In order to make these objects’ spectral signature identifiable, it is often very fruitful to apply imaging in non-visible wavebands. By photographing beyond the limits of the HVS, the digital remote sensing method presented indicates that aerial NUV photography can indeed provide a useful new source of archaeological information, thus warranting the necessary further research needed, even though the initially acquired imagery does not directly reveal completely new information and vegetation exhibits a far less distinct spectral response pattern compared to the longer wavelengths. However, as indicated before, these first examples only want to give the initial impetus to the further examination of NUV aerial reconnaissance, rather than already judging its potential or uselessness.

Looking at the very first digital NUV aerial frames acquired, this new way of imaging seems promising to enhance certain soil related differences. The extent to which this observation is applicable (e.g. kind of soil) and its complete understanding still remain open for further exploration, just as the question whether or not this technique can reveal anomalies other spectral bands are unable to display.

However, to really research digital NUV aerial reconnaissance in a more systematic way, a shorter focal length lens is deemed essential – to allow for a larger ground coverage – while optics with a higher transmission (in the 320 nm to 400 nm range) and decent image quality at large apertures could make shorter shutter speeds possible. Although these improvements could finally create acceptable conditions to perform UV aerial reconnaissance by aeroplane, they should primarily be considered as means to enable a lower percentage of blurred imagery from the Helikite (or similar lifting device), create workable photographic conditions on less sunny days and/or allowing lower ISO values (something absolutely essential if one wants to tackle the noise in the darker image areas). Employing the current outlined approach thus means that both the specific photographic hardware needed and necessary long shutter speeds do not allow digital NUV photography to be systematically applied in aerial reconnaissance, limiting the general aerial archaeological usefulness to those cases where a stable platform is provided.
It must also be obvious that the biggest improvement could be obtained from the imaging medium itself. At present, the digital sensors found in DSCs are – logically – not optimised for this photographic method, which means that the camera will only respond moderately to incoming UV photons, even after the modification (which of course remains of the utmost importance for aerial NUV imaging, even though the operation will void the manufacturer’s warranty). Consequently, as long as no specific UV-sensitised sensors are available in current DSCs, the outlined method is the most cost-efficient and effective way to execute this form of invisible aerial imaging, with the important additional advantage of owning a DSC also useable for ground-based imaging in the NUV, visible, and NIR domain (if the lens is fitted with the right filters).

Complementary to these possible future improvements might be the (partly) inclusion of the Blue waveband – as was advised by Vizy [800] in his analogue approach of NUV photography –, hence largely increasing the amount of captured EM radiation and allowing the exposure time to decrease. Though this idea is supported by the results of research in which the NUV-band showed a very similar contrast compared to the output from the adjacent blue band [341, 396], it is not directly advisable: the spectral band would become too broad, thus possibly reducing the acquisition accuracy of NUV-specific spectral reflectance signatures; moreover, one even risks completely masking particular diagnostic features, as they might be too narrow to be distinguished by such a coarse spectral approach. It goes without saying that, again, far more fieldwork is needed to formulate meaningful answers. In the end, it seems that the following remark of the 1960s still holds true: “ultraviolet photographic reconnaissance … would be best, however, at 200 to 1000 ft, and at such low altitudes it could indeed be valuable for specific kinds of targets” ([206]: 28).
Conclusion

Digital photography is commonplace in archaeological aerial reconnaissance. By acquiring the reflected visible portion of incident wavelengths radiated by the sun, Digital Still Cameras (DSCs) taken aloft in aeroplanes allow to assess remotely the vegetation status over both small and large areas. Because their output renders archaeologically induced colour and height differences in crops almost identically as the Human Visual System (HVS) perceives them, DSCs have always been relied upon as aerial research instruments, without really considering to alter their common behaviour.

In the Potenza Valley Survey (PVS), such straight-out-of-the-box DSCs have also been used for several years, yielding very interesting archaeological information year after year. When combined with geophysical data and information generated by field and topographical survey, the inherent archaeological benefit of oblique aerial imagery can be exploited to a very large extent. In the first chapters of this thesis, several examples have shown how this non-destructive methodology was applied in the PVS with great results on several Roman cities. However, applying DSCs the same way as analogue film-based cameras were used will not really help in advancing the discipline of archaeological aerial reconnaissance. Even though various aerial and spaceborne remote sensing systems have the capacity to offer new imagery acquired in very small to very broad wavebands (both visible and invisible), few of them (if any) generate images in the same flexible and economic way as active aerial photography from a low-flying aeroplane does. Therefore, means must be found to maximize the information gathered by the common latter approach, both during the image acquisition phase as afterwards during the processing of the generated data.

Given this, it is striking to notice that more than a century after aerial archaeology was initiated, the basic routines of active aerial reconnaissance are still largely identical to those of the true pioneers. More than seventy years ago, this fact was also noted by Reeves, who mentioned that aerial archaeology “will increase in scientific value” once “its methods and technique are improved” ([654]: 107). The methods and techniques of aerial imaging in the Near-UltraViolet (NUV) waveband, the Red edge zone, and the Near-InfraRed (NIR) spectral
region by means of low-cost modified DSCs are the author’s first attempts to tackle this issue. Although archaeological NIR aerial imaging is by no means novel, the advantages modified DSCs can offer in the generation and interpretation of reconnaissance information are substantial, using individual pure NIR spectral channels on the one hand and arithmetic operations performed on a combination of channels on the other. When combined with information from the visible wavebands (in an ideal situation reflectance data from the Red edge spectral region), many new opportunities are provided to visually enhance archaeologically related anomalies (more specifically crop marks) and/or even reveal completely new features. In addition to simplifying the complete workflow, the NIR-modified DSCs thus seriously expand and perfect some of the possibilities known from the film-based NIR approaches, without the significant costs and workflow-related drawbacks of the latter.

Data acquisition in a more energetic part of the ElectroMagnetic (EM) spectrum (i.e. the Near-UltraViolet or NUV) has proven to be more difficult. However, the provided real-world examples have shown that NUV photography can be really beneficial for soil mark archaeology once most of the practical problems are solved.

The thesis under consideration thus proves that (narrowband) multispectral imagery acquired in several visible and beyond-visible domains can be very advantageous for the detection and interpretation of archaeological anomalies. Moreover, it has been proven that this information can be generated using the same tools aerial archaeologists normally use: digital small-format photographic cameras. This way, it seems that the boom of the digital photography industry can – finally – meet the demand for technological improvements Reeves asked for.

However, it’s not all roses. Some existing drawbacks still need to be resolved before this research can be really beneficial for the PVS project on the one hand, and the broader archaeological community on the other.

- As previously mentioned, all aerial data were acquired within the wider framework of the PVS project. Consequently, all conclusions drawn (positive as well as negative) are only based on the results of a limited research area – from a geographical, archaeological, geomorphological, and pedological point of view. Even though all physical and chemical principles of NUV, multispectral, and NIR imaging remain equal all over the world, the true potential of these techniques has until now not been explored on a very large scale;

- A second, scale-related problem concerns the amount of data analysed. In spite of the results obtained on several Roman cities along the Potenza river, many thousands of aerial photographs acquired within the framework of the PVS still need to be processed (both oblique and vertical historical ones). This means that a large comparison between the data from the field survey on the one hand, and all information deduced from the aerial remote sensing survey on the other, is not fully executed yet. To really quantify the added value of the outlined new approaches in the PVS (and all large-scale interdisciplinary archaeological field projects by extension), processing all data is absolutely essential. However, the latter is presently hampered by a third problem;
Conclusion

- Even though the processing of NIR or NUV frames is not a problem, the general workflow is presently not streamlined enough. When every proposed image processing technique needs to be applied (e.g. coregistering a number of aerial frames to apply certain arithmetic operations on particular image channels), several (expensive) software packages must be used, while results generated by one package are not directly compatible with the data formats of another. Moreover, processing still occurs on a one-by-one basis. Hence, the proposed techniques are still too time-consuming. Maybe even more important: these new procedures of data handling are not easily accessible for the broader archaeological community;

- Another point of criticism might be the lack of new procedures to extract information from digital colour frames generated by taking only the reflected portion of the visible light into account. Certainly when one knows that these data make up a very large portion of all archaeological reconnaissance data ever generated, building new methods to analyse them might be THE new approach waited for;

- Finally, there is the major drawback of the reconnaissance method itself. No matter how efficient and accurate these new tools can be, an increase in site discovery rate using multispectral imaging with DSCs is unlikely as long as the predominant flying strategy of ‘observer-directed’ survey (i.e. only features spotted from the aeroplane will be photographed) is in practice. This approach generates extremely biased data that are totally dependent on an airborne observer recognising archaeological phenomena. Thus, subsurface soil disturbances that are visually imperceptible at the time of flying will not make it into a photograph, even if their particular spectral response in a particular waveband might exhibit enough contrast with the surrounding matrix.

In an attempt to address these shortcomings, there is need to formulate some guidelines and future research topics. In essence, this future research needs to improve the validity of all methodological-technological innovations developed during the doctoral research and make them easily accessible for interested parties. In essence, new research should largely be based upon the following three cornerstones.

- First of all, it is of the utmost importance to apply all data acquisition procedures as well as their specific data processing algorithms on as many dissimilar archaeological contexts as possible (i.e. differing in geography, topography, pedology, geomorphology, structure type, etc.) to assess and define all possible advantages and drawbacks. Moreover, this approach should define the conditions that are beneficial for (certain of) these techniques and which are not. The generation of such data should ideally be acquired by different scholars using different flying strategies. This way, the results gained in this first cornerstone must allow to enter the debate on the best flying practice for archaeological reconnaissance: “Is it still valid to fly randomly and take observer-directed obliques, or should archaeologists concentrate on collecting geographically unbiased photographs of large areas by vertical block coverage?”; “Are separate approaches more beneficial, depending upon the phase of the research (i.e.
flying for the first time in an unexplored area or for the fifth year in a well-known region) or the spectral information one wants to capture (i.e. visible versus NIR)?

- Secondly, the author intends to create an archaeological toolbox: an integrated software package to execute all image processing procedures developed. Besides, the programme must be able to work with a large variety of file formats (JPEG, TIFF, and all RAW formats), it must exhibit a large degree of scalability (i.e. allow different procedures and algorithms to be implemented), enable batch processing, and be downloadable free of charge (to guarantee an easy distribution and allow archaeologists to implement it in their aerial image processing workflow). While this might seem too optimistic and very difficult to achieve, it is important to state that OpenSource software can be used for reading and writing most file types (e.g. dcraw for the RAW file formats) and several image rectification algorithms for automatic coregistration exist. Another option might be to extend existing free (e.g. ImageJ) or low-cost software (e.g. AirPhoto) with these new processing techniques. However, no matter which option is chosen, it must be clear all data analyses have to become available at low cost and in an easy-to-handle programme if the results of this thesis will ever be used by the community of aerial archaeologists.

It goes without saying that a strong interdependence exists between the research questions of part one and the toolbox development: improvements in data acquisition and processing have to be incorporated in the software package, while the latter can largely aid in testing new algorithms and efficiently process large volumes of imagery. This way, the toolbox becomes a research instrument which can allow to assess the validity of newly developed approaches (e.g. those currently developed for conventional colour frames), and research their applicability for archaeology revealed by phenomena other than crop and soil marks (i.e. snow marks, shadow marks, water marks, etc.) In the end, the programme will build upon a good scientific basis, while providing an easy and affordable means to intelligently process archaeological aerial frames, yielding a maximum of information out of the aerial data without the need to master several software packages;

- Thirdly, all aerial data ever acquired in the framework of the PVS should finally be confronted with the information generated by the field surveys. As soon as the toolbox is ready, batch processing of all remaining aerial photographs will make information extraction very efficient and more profound than has ever been possible. Afterwards, the final interpretation of this output can provide the necessary input to perform the long-awaited GIS-based comparison with the detailed information extracted from the PVS field survey data and the sites mentioned in the literature. This confrontation should formulate meaningful answers to several questions: “Are certain archaeological particularities only susceptible to one survey method (e.g. field survey, NIR aerial survey, etc.)?"; “Which are the necessary requirements (crop species, soil type, and season) for NUV- and/or NIR-photography to be most effective?"; “To what extent is information generated by field and aerial survey supplementary, similar and/or dissimilar?"; “Are these results quantifiable?".

Conclusion

In the end, these analyses should be conclusive on the added value of aerial reconnaissance in general and these innovative technological approaches more specifically. However, it needs to be stressed that all methodologies presented in this thesis could only be developed by a thorough understanding of real-world materials’ reflectance properties, the basic working principles of DSCs, and the building of a new aerial platform. First of all, all terrestrial material such as soils, rocks, sediments, vegetation, and other organic matter mostly has very specific spectral reflection, transmission, absorption, and emission characteristics. In order to exploit these objects’ spectral signatures, it is, however, of the utmost importance to understand how and why particular objects reflect incoming EM radiation. Only when these characteristics are understood, new targeted imaging approaches and processing techniques can be presented.

To maximise this knowledge, a thorough understanding of the spectral properties and operating principles of the imaging hardware is deemed essential. With the advent of RAW digital photography, aerial archaeologists are now given the chance to get the maximum from their photography, as a RAW workflow enables enormous control over the final output, and its use is mandatory practice for strictly scientific photography such as airborne remote sensing. This does not mean that conventionally shot JPEG and TIFF images are unusable; it only indicates that both the development and application of certain methodological improvements necessitates a certain standard of technical photographic knowledge and imaging hardware. Concerning the latter, cost should in most situations no longer be an issue, as even very basic, low-cost D-SLRs offer straightforward RAW photography.

Finally, building a new aerial platform based on a Helikite also proved to be essential. Besides its useful application in the site-based generation of overviews and detailed excavation photographs, NUV aerial imaging would have been impossible without Helikite Aerial Photography (HAP). Due to its characteristics, HAP now functions as a research tool that allows imaging during less-than optimal weather conditions and in spectral domains that would otherwise be inaccessible. The latter application can be very interesting, as exploring the potential of very unconventional wavebands might reveal whether or not future (technological and methodological) research should be directed in this direction.

It thus seems rather self-evident that proposing or developing innovations in aerial reconnaissance will largely come down to a better understanding of both the reflected radiation and particular phenomena aerial archaeologists try to visualise, as well as the properties of the hardware applied. Although sufficient results might be generated with a normal DSC and a ‘put-it-on-automatic-and-press-the-right-button’ method, this approach is unjustifiable from a scientific, researcher’s point of view and will never allow archaeological aerial reconnaissance to step beyond its conventional boundaries.
Publications

The following codes have been used to indicate the impact of the publication:

ERIH: publication part of the European Reference Index for the Humanities;
(A, B, C): journal code as determined in the ERIH Initial List for Archaeology;
SCIE: publication mentioned in the Science Citation Index Expanded;
SSCI: publication mentioned in the Social Sciences Citation Index;
AHCI: publication mentioned in the Arts & Humanities Citation Index;
CPCI-S: publication mentioned in the Conference Proceedings Citation Index – Science;
IF: Impact Factor of this publication as mentioned on the Web of Science.

Publications in International Journals


Verhoeven, G., 2009. Providing an Archaeological Bird’s-Eye View – An Overall Picture of Ground-Based Means to Execute Low Altitude Aerial Photography (LAAP) in Archaeology. Archaeological Prospection. [ERIH (B); SCIE, AHCI – IF 2007: 0.660] Submitted


Publications in Congress Proceedings


Publications in Books


Publications as Editor

References


References


References


References


References


References

[137] Capper, J.E., 1907. Photographs of Stonehenge as Seen from a War Balloon. Archaeologia: or Miscellaneous Tracts Relating to Antiquity 60, 571.


References


References


References


References


[271] Eisenhauer, S., 1995. and for my Fly Rod & Reel... The Aerial Eye 1, 6-7.


References


References


References


References


References


References


References


References


References


References


References


Owen, B., s.d. KAP: Kite Aerial Photography. www.svcs.net/tutorials/imaging/KAP/ (accessed 17.05.2007).


References


References


References


References

For Cultural Heritage. In: S.n., Proceedings of the CIPA 2005 XX International Symposium,
Torino, Italy. CIPA.

Enzelschriften 32)

and Biological Illustration 27, 35-36.

[677] Rørslett, B., 2004. All You Ever Wanted to Know About Digital UV and IR Photography, But


Stoddart, D.R., Johannes, R.E. (Eds.), Coral Reefs: Research Methods. UNESCO, Paris, pp. 45-
53. (= Monographs on Oceanographic Methodology)


[682] Scheer, L., Racko, J., 1987. Assessment of Tree Variables from Large-Scale Aerial

Phosphorescence Emission. An Introduction with Experiments. Allyn and Bacon, Boston. (=
Allyn and Bacon Chemistry Series)


Reflectance Measurements of Corn Leaves from Plants with Different Nitrogen and Water


Australian Archaeology 59, 51-58.


[691] Schnetz, J., Zumsschlinge, M., 1940. Ravennatis Anonymi Cosmographia et Guidonis
geographica. Teubner, Stuttgart.

John Wiley & Sons, New York, pp. 53-64.
References


References


References


References


References


References


References


« Soon I’ll have the courage to leave my thoughts behind. I’ll give back all the knowledge and keep the wisdom precious in my mind. »

Massive Attack (Three), trip hop group