A non-rigid registration method for multispectral imaging of plants

Jonas De Vylder¹, Koen Douterloigne¹, Filip Vandenbussche², Dominique Van Der Straeten² and Wilfried Philips¹

¹Ghent University - IBBT - Dept. of Telecommunications and Information Processing, St.-Pietersnieuwstraat 41,9000 Ghent, Belgium
²Ghent University - Laboratory of Functional Plant Biology, Dept. of Physiology, K.L. Ledeganckstraat 35,9000 Ghent, Belgium

ABSTRACT

Registration of multispectral images remains a challenging task due to the lack of stable feature points. Methods based on intensities are generally more robust for multi-modal image registration, but are computationally demanding or are restrictive to the transformation model allowed in the registration. This paper proposes a new registration framework which overcomes these drawbacks. The proposed method optimizes the location of a set of virtual landmarks in order to get robust and accurate registration.

Keywords: image registration, multispectral imaging, non-affine registration

1. INTRODUCTION

Analyzing images is an interesting non-destructive method to quantify plant phenotypes, e.g. measuring plant growth.¹⁻³ Although several high-throughput monitoring systems for plant phenotyping are available,⁴⁻⁶ only growscreen/fluoro is able to measure plant features outside the visual light spectrum,³ albeit that it only allows analysis of chlorophyll fluorescence images. Recently information coming from multispectral images has shown to be a useful tool for analysing plant physiological parameters.⁶⁻⁷ Lacking industrial high-throughput systems, this multispectral monitoring is often done using a home-based system, where multiple cameras are moved over the plants using a robotized arm. This approach requires that images are first registered, i.e. aligned to each other, in order to cope with difference in scaling, resolution, lens distortion, motion between the capturing of two images, etc.

Two main registration approaches are extensively described in literature: feature based and area based approaches.⁸ In the first approach a set of feature points are detected in both images, the corresponding feature points are matched and based on their location in both images a transformation model between both images is calculated. There exist many different feature points, e.g. SIFT, SURF, RIFT, GLOH, etc., each with their own advantages such as invariance to rotation, scaling and affine distortions. These methods have the benefit of being fast while allowing a wide range of transformation models. The downside is that they require stable feature points which are clearly visual and detectable in both images. This hampers their use for many multi-modal images. The area based registration approach does not require any feature points, instead they calculate optimal parameters of a transformation model in order to minimize a specific dissimilarity measurement. These methods have been proven robust and are useful for multi-modal image registration.⁷ These methods either require a good initialization for the transformation model or require a brute force search for the optimal parameters, which is only feasible for a simple transformation model with few parameters. In this paper we propose a combination of a feature point based method with an area based method. Our method does not need accurate detection of feature points, while allowing for complex transformation models, i.e. non-rigid and non-affine transformation models.

This paper is arranged as follows: the next section will elaborate in detail about the proposed method. The section discusses the registration framework, several pre-processing methods and the use of different optimization criteria. Section 3 shows the results of the proposed method for the registration between chlorophyll fluorescence images and near-infrared images. Section 4 recapitulates and concludes.

Corresponding author: Jonas De Vylder
Jonas De Vylder: E-mail: jonas.devylder@telin.ugent.be, Telephone: +32 9 264 34 16
2. REGISTRATION

2.1 Non-rigid registration

The proposed registration method starts from a set of virtual landmarks. These virtual landmarks have the same function as feature points, however they do not have to correspond to an accurate location of a visual feature since the location of these virtual landmarks are iteratively optimized in order to get a more accurate registration. Based on the location of these virtual landmarks in both images, i.e. \( c_2(t) = (x(t), y(t)) \) and \( c_2(t) = (x(t), y(t)) \) with \( t \in 0, 1, \ldots, N \) and \( N \) the number of control points. A non-rigid transformation model can be calculated using polyharmonic spline interpolation:

\[
p_2 = a + a_x x - a_y y + \sum_{j=1}^{N} w_j \phi(\|p_1 - c_1(j)\|) \tag{1}
\]

where \( p_2 \) represents the coordinates in the second image, corresponding to location \( p_1 = (x, y) \) in the first image. The parameters \( a, a_x, a_y \) and \( w_j \) are a set of weighting coefficients and \( \phi(\cdot) \) is a radial basis function, i.e.

\[
\phi(r) = \begin{cases} r^k, & \text{if } k \text{ is odd} \\ r^k \ln r, & \text{if } k \text{ is even} \end{cases}
\tag{2}
\]

where \( k \) is the degree of the basis functions. Note that \( k = 2 \) is a special case since eq. (3) then corresponds to Thin Plate Spline interpolation.\(^{10}\) The weighting coefficients in eq. (3) are calculated in such a way that:

- landmarks are mapped to the corresponding points in the other image, i.e.

\[
c_2(i) = a + a_x x + a_y y + \sum_{j=1}^{N} w_j \phi(\|c_1(i) - c_1(j)\|) \tag{3}
\]

- the weighting vector \( w \) is orthogonal to \( x \) and \( y \), i.e.

\[
\sum_{i=1}^{N} w_{i,x} x_i = \sum_{i=1}^{N} w_{i,y} y_i = 0 \tag{4}
\]

- the weighting coefficients sum to zero

These constraints result in \( N + 3 \) linear equations for an equal number of variables. Thus the weighting coefficients can be found by solving a linear system, which can be done in a fast way as proposed by Beaton et al.\(^{11}\) For a more detailed discussion on polyharmonic splines and the influence of the different radial basis functions we refer to.\(^{12}\)

The overall registration result depends on how well the landmark locations correspond in both images. Therefore we optimize the location of the landmarks in order to minimize a specific energy function. This energy function expresses the dissimilarity between the registered images. Several dissimilarity measurements have been proposed in literature and we will discuss the most important ones in subsection 2.2.

The optimization of the landmarks location is done using a coarse-to-fine strategy. Around each landmark, a set of potential landmarks is chosen. These potential landmarks are chosen using the following rule: \( c_p = c + (ks, ls) \), with \( k = \{-2, -1, 0, 1, 2\} \), \( l = \{-2, -1, 0, 1, 2\} \) and \( s \) a discrete step size. The potential landmark that minimizes the energy function is detected using an exhaustive search. The resulting landmark is used as initialization for a new iteration where the step size is decreased. In Fig. 1 a visual representation of this optimization method is shown. The yellow square represents the initialization of a landmark, around it on a regular grid, a set of potential landmarks are shown. The green circle represents the potential landmark that minimizes the energy function. This optimal potential landmark, closest to the real optimum is chosen as initialization for the next iteration, where the potential landmarks are located at a finer grid around the initial landmark.
2.2 Dissimilarity measurements

The method proposed in this paper requires the optimization of the location of a set of virtual landmarks in order to minimize a specific energy function. Many different energy functions penalizing dissimilarity between images have been proposed in literature. Due to the abundance of dissimilarity measurements, enumerating all used for image registration would be an unfeasible task, but we attempt to give an overview of the most common approaches: measurements based on $L_1$ and $L_2$ norms, correlation based measurements, statistical measurements and residual based methods. An example of these different dissimilarity measurements is shown in Fig. 2. This example is based on the registration between the green and red channel of a VIS image. The green image was horizontally translated between -20 to +20 pixels.

2.2.1 Sum of squared or absolute differences

One of the easiest dissimilarity measurements are the sum of absolute differences (SAD) and the sum of squared differences (SSD), which respectively correspond to the $l_1$ norm and the squared $l_2$ norm between both images:

$$E_{SAD}[f(\cdot,\cdot),g(\cdot,\cdot)] = \sum_{x=1}^{M} \sum_{y=1}^{N} |f(x,y) - g(x,y)|$$

and

$$E_{SSD}[f(\cdot,\cdot),g(\cdot,\cdot)] = \sum_{x=1}^{M} \sum_{y=1}^{N} [f(x,y) - g(x,y)]^2$$

SAD and SSD are popular measurements used for block matching\(^\text{15}\) and for strain mapping using digital image correlation.\(^ \text{14,15}\) Note that these measurements are sensitive to changes in intensity. Some variants have been proposed, but these mainly cope with global light variation, which does not apply to multispectral images. Although these measurements are mainly used for images captured using the same sensor, they can still perform reasonably well in certain multi-spectral images as can be seen in Fig. 2(a-b). Note that in this example bright pixels in one image typically corresponds with bright pixels in the other image, which is not always the case in multi-modal imaging.

2.2.2 Normalized cross-correlation

A different set of measurements are based on the cross-correlation between two images. Several variations on the cross-correlation measurement have been studied in literature, e.g. the basic cross-correlation, zero-mean cross-correlation, normalized cross-correlation, etc.\(^\text{15}\) One of the most popular optimization criteria is the normalized cross-correlation (NCC), which can handle changes in contrast between both images. An NCC based energy function is defined as follows:

$$E_{NCC}[f(\cdot,\cdot),g(\cdot,\cdot)] = \frac{\sum_{x=1}^{M} \sum_{y=1}^{N} f(x,y)g(x,y)}{\sqrt{\sum_{x=1}^{M} \sum_{y=1}^{N} f(x,y)^2 \sum_{x=1}^{M} \sum_{y=1}^{N} g(x,y)^2}}$$

Note that we define our energy as the negative of the NCC, as defined in.\(^ \text{8,15}\) This is done because cross correlation is a measurement of similarity instead of dissimilarity, so by minimizing eq. (7) we actually maximize the similarity between the registered images. Cross correlation based measurements are closely related to SAD and SSD variants.\(^\text{15}\)
2.2.3 Correlation ratio

Roche et al. propose\textsuperscript{16} to use the correlation ratio for image registration. The correlation ratio can be seen as a generalization of the correlation coefficient, i.e., a statistical measurement expressing the functional dependence between both images. The correlation ratio is defined as follows:

$$E_{CR}[f(\cdot, \cdot), g(\cdot, \cdot)] = \frac{\text{Var}[g(\cdot, \cdot) - E[g(\cdot, \cdot)]|f(\cdot, \cdot)]}{\text{Var}[g(\cdot, \cdot)]}$$

(8)

Where $\text{Var}[]$ and $E[\cdot]$ respectively correspond to the variance and conditional expectation of the variables. In contrast to the previous measurements, the correlation ratio does not impose any linear dependency between both images, thus can handle multi-modal images.\textsuperscript{5,16} In Fig. 2 (d) the correlation ratio results in a minimum when both images are perfectly aligned, the measurement however does not increase monotonically when the translation increases.

2.2.4 Mutual information

Mutual information is considered to be the golden standard for multi-modal image registration. Mutual information, originating from information theory, measures the statistical dependence between two images, where the images are treated as two discrete random variables.

$$E_{MI}[f(\cdot, \cdot), g(\cdot, \cdot)] = -\sum_{x=1}^{M} \sum_{y=1}^{N} p(f(x, y), g(x, y)) \log \frac{p(f(x, y), g(x, y))}{\sum_{i=1}^{L} p(f(x, y), i) \sum_{j=1}^{L} p(j, g(x, y))}$$

(9)

Where $p(i, j)$ represents the probability that a pixel in the first image has an intensity of $i$, while the corresponding pixel in the second image has an intensity equal to $j$. Mutual information expresses the amount of information that can be obtained from $f(\cdot, \cdot)$ by observing $g(\cdot, \cdot)$. It can also be defined in function of the entropy and joint entropy of both images.\textsuperscript{8,17}

2.2.5 Residual complexity

A completely different approach from the statistical and correlation related measurements is proposed by Myronenko et al.\textsuperscript{18} They propose a new measurement of dissimilarity, which quantifies how compact the difference between two images can be stored. The compactness of the residuals is expressed using the discrete cosine transform. This measurement is called Residual Complexity and is defined as follows:

$$E_{RC}[f(\cdot, \cdot), g(\cdot, \cdot)] = \sum_{k=1}^{M} \sum_{l=1}^{N} \log \left( \frac{1}{\sum_{x=1}^{M} \sum_{y=1}^{N} (f(x, y) - g(x, y))^2 \cos \left( \frac{x}{M} \pi \right) \cos \left( \frac{y}{N} \pi \right)} \right)^2 / \alpha + 1$$

(10)

Where $\alpha$ is a weighting term. In Fig. 2(e) the residual complexity is shown for different values for $\alpha$. Note that the minimum is clearly visible, but that the difference between a translation of -20 and a translation of 2 pixels results in a similar residual complexity.

2.3 Preprocessing

Although some energy functions are more robust against non-linear intensity changes between different image modalities, this might not be sufficient. An example of such a situation can be seen Fig. 3, where on the left a regular VIS image, i.e., a regular colour image in the visible spectrum, of a rosette is shown, while on the right a near-infrared image of the same rosette is shown. As can be seen, some leaves are brighter than the background, while others are darker and some are not visible at all. This makes it difficult to model the intensity changes in an energy function. This might be solved by using different image features than intensity.\textsuperscript{18-21} The location of bright intensity might not necessarily correspond to a leaf, but in this example the location of an edge does correspond to a contour in the VIS image. Note that exploiting these features is different from the previously mentioned feature based registration in the sense that we do not attempt to get a binary decision of where a specific feature point is. Instead the image is transformed in such a way that specific features are emphasized. These transformed images are then registered in order to get the optimal registration parameters of eq. (3)
Figure 2. An example of different dissimilarity measurements in function of a horizontal shift between two images.

Figure 3. An example of a multispectral image of an Arabidopsis thaliana rosette. On the left a regular colour image, on the right a temperature based image.
2.3.1 Edges

The benefit of edges was already pointed out by the example given in Fig. 3, where edges typically correspond to either the contour of a rosette, or to some structure in the trays, e.g. the borders of the cups which hold the plants. A widely used edge transformation is the absolute value of the gradient in the image, which is the first step of the popular Canny edge detector, i.e.

\[ F_{edge}(x, y, f(., .)) = |\nabla f(x, y)| = \sqrt{\left( \frac{\partial f(x, y)}{\partial x} \right)^2 + \left( \frac{\partial f(x, y)}{\partial y} \right)^2} \]  

(11)

2.3.2 Corners

Although less present in the Arabidopsis rosette example of Fig. 3, corners can be a very discriminative feature for registration of images. This happens specially in plant species with leaves with high curvature, e.g. glabrum or kelloggii. Also images of flower buds will benefit from a corner map. We propose a corner transform based on the Harris corner detector. For each pixel the Hessian matrix is calculated:

\[ H(x, y, f(., .)) = \begin{bmatrix} \frac{\partial^2 f(x, y)}{\partial x^2} & \frac{\partial^2 f(x, y)}{\partial x \partial y} \\ \frac{\partial^2 f(x, y)}{\partial x \partial y} & \frac{\partial^2 f(x, y)}{\partial y^2} \end{bmatrix} \]  

(12)

If the pixel at location \((x, y)\) corresponds to a corner, then \(H(x, y, f(., .))\) will have large positive eigenvalues, \(\lambda_1\) and \(\lambda_2\). In order to avoid the computationally expensive calculation of the eigenvalues, Harris and Stephens proposed the following feature map:

\[ F_{corner}(x, y, f(., .)) = \lambda_1 \lambda_2 - \kappa(\lambda_1 + \lambda_2)^2 = \text{det}(H(x, y, f(., .))) - \kappa \text{trace}(H(x, y, f(., .))) \]  

(13)

with \(\kappa\) a sensitivity parameter.

2.3.3 Ridges

A feature similar to edges are ridges, but instead of emphasizing contours, they will respond to either small branches or stalks or in the case of high resolution images they respond to the nerves and veins in leaves. For the ridge transform we propose the Lindeberg measure of main principal curvature:

\[ F_{ridge}(x, y, f(., .)) = \frac{\partial^2 f(x, y)}{\partial u^2} + \frac{\partial^2 f(x, y)}{\partial v^2} - \sqrt{\left( \frac{\partial^2 f(x, y)}{\partial u^2} \right)^2 + \left( \frac{\partial^2 f(x, y)}{\partial v^2} \right)^2} \]  

(14)

where \(u\) corresponds to the principal curvature direction in pixel \((x, y)\) and \(v\) perpendicular with \(u\).

3. RESULTS

For the validation of the proposed registration method a dataset of sugar beet seedlings was used. The seedlings were captured using both a near-infrared camera and a chlorophyll fluorescence camera. The near-infrared camera resulted in low contrast images were plant leaves where manually delineated. The chlorophyll fluorescence images showed good contrast where plants appear as bright objects on a black background. These image are accurately segmented using a combination of thresholding and morphological closing.\(^{22}\) Optimal registration would result in a perfect alignment of the manually delineated leaves and the automatically segmented leaves. To measure correspondence between both segmentations we use the Dice coefficient: consider \(S\) the resulting segment from the manual annotation, i.e. the region enclosed by the delineation, and \(F\) the segmentation result of the chlorophyll fluorescence image, then the Dice coefficient between \(S\) and \(F\) is defined as:

\[ d(S, F) = \frac{2\text{Area}(S \cap F)}{\text{Area}(S) + \text{Area}(F)} \]  

(15)

Here \(S \cap F\) consist of all pixels which both belong to the detected segment as well as to the annotated segment. If \(S\) and \(F\) are equal, the Dice coefficient is equal to one. The Dice coefficient will approach zero if the regions hardly overlap.

In total 108 image pairs were analysed using different combinations of energy function and pre-processing steps. The registration was done using 15 virtual landmarks and a biharmonic spline, i.e. \(k = 1\) in eq. (2). The virtual landmarks
<table>
<thead>
<tr>
<th>method</th>
<th>Intensity</th>
<th>Edges</th>
<th>Corner</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mutual information</td>
<td>0.8815</td>
<td>0.2790</td>
<td>0.1193</td>
</tr>
<tr>
<td>Residual complexity</td>
<td>0.8404</td>
<td>0.8601</td>
<td>0.8762</td>
</tr>
<tr>
<td>SSD</td>
<td>0.1027</td>
<td>0.3161</td>
<td>0.3192</td>
</tr>
<tr>
<td>SAD</td>
<td>0.4276</td>
<td>0.3075</td>
<td>0.2263</td>
</tr>
<tr>
<td>Normalized cross-correlation</td>
<td>0.8254</td>
<td>0.8623</td>
<td>0.7527</td>
</tr>
<tr>
<td>Correlation ratio</td>
<td>0.6930</td>
<td>0.3146</td>
<td>0.7115</td>
</tr>
</tbody>
</table>

Table 1. The average Dice coefficient between manually annotated leaves in NIR images and automatically segmented leaves in chlorophyll fluorescence images, using different combinations of energy functions and pre-processing steps.

<table>
<thead>
<tr>
<th>method</th>
<th>time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mutual information</td>
<td>0.1925</td>
</tr>
<tr>
<td>Residual complexity</td>
<td>0.6479</td>
</tr>
<tr>
<td>SSD</td>
<td>0.0186</td>
</tr>
<tr>
<td>SAD</td>
<td>0.0186</td>
</tr>
<tr>
<td>Normalized cross-correlation</td>
<td>0.0317</td>
</tr>
<tr>
<td>Correlation ratio</td>
<td>0.2978</td>
</tr>
</tbody>
</table>

Table 2. The average time to calculate the energy for a typical image.

were manually selected as an initialization. The average Dice coefficient between rosettes in the near infrared images and rosettes in the chlorophyll fluorescence images are shown in Table 1. The use of the ridge detector is not summarized in this table because the near infrared images had too low contrast to show veins or small stalks. Mutual information without any pre-processing performs best, but also residual complexity shows good results. Normalised cross-correlation in combination with a pre-processing step that emphasizes edges gives similar results. SSD, SAD and correlation ratio result in bad registration, even in combination with a pre-processing step.

There is not only a difference in registration quality using different energy functions, but also a difference in the computational cost. The time it takes to calculate the energy is highly important since this has to be done many times during the iteration process. In Table 2 the time to calculate the energy between two images of 1024 × 1024 pixels is shown. For applications where computational time is crucial, it might be interesting to combine edges with the normalized cross-correlation, which results in slightly less accurate registration than using mutual information, but is significantly faster.

4. CONCLUSION

This paper proposes a generic registration framework which can be used for multispectral images. The framework does not require any stable feature points while allowing for a highly flexible transformation model, i.e. non-rigid and non-affine registration. The framework can be tuned in order to get accurate or computationally efficient registration, depending on the requirements of the application.

ACKNOWLEDGMENTS

This research has been made possible by the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT), Ghent University and the Fund for Scientific Research-Flanders. FV is a postdoctoral researcher of the Fund for Scientific Research-Flanders. Finally, the authors wish to thank Laury Chaerle for setup, image acquisition and annotation of the time lapse sequences.

REFERENCES


