Data-driven Summarization and Synchronized Second-screen Enrichment of Cycling Races
Using Live and Historical Sports Data to Reinvent Traditional Reporting

Steven Verstockt
ELIS Department
Ghent University – imec, IDLab
Ghent, Belgium
steven.verstockt@ugent.be

Erik Mannens
ELIS Department
Ghent University – imec, IDLab
Ghent, Belgium
erik.mannens@ugent.be

Jelle De Bock
ELIS Department
Ghent University – imec, IDLab
Ghent, Belgium
jelle.debock@ugent.be

ABSTRACT
Traditional broadcasters of cycling races are experiencing hard times as the numbers of spectators are decreasing each year. Other ways of reporting are needed to keep the viewer interested. In this paper, two possible solutions are proposed that have been evaluated during the Grand Depart of the Tour de France 2019 in Brussels. The first innovation focuses on data-driven summarization and allows end-users to query for personalized stories of a race, tailored to their wishes (such as the length of the clip and the teams and/or riders that they are interested in). The second innovation follows the second screen trend and synchronizes cycling heritage multimedia data with the riders’ live location during the race. Both rich, interactive TV experiences are based on a combination of data mining and computer vision techniques which can also be applied to other sports with similar characteristics. Evaluation by a test audience showed that there is certainly potential in both formats.

KEYWORDS
Cyclist recognition; data-driven summarization; geolocalization; synchronization; enrichment, user experience.

ACM Reference format:

1. INTRODUCTION
“Cobblestones hurt, but the spectator numbers hurt more”. This is one of many recent newspaper headlines focusing on the difficulties/problems that traditional cycling broadcasts are facing. Almost all races face a remarkable decline over the last years, as is shown by the numbers for some of the most popular races in 2017 and 2018 (Figure 1).

![Figure 1: Remarkable decline of number of spectators for 6 of the most popular races in 2017 and 2018.](image)

Races are often too long to be attractive all time and for several of them the end result is rather predictable. The way in which the race is captured also didn’t change much over the last decades. All these factors contribute to a weakening interest during the last years, especially of young generations. Related to this, it is important to mention that not only cycling, but also other sports, see quick rise in average age of TV viewers as younger fans shift to digital platforms or drop out [1]. There is an increased interest in short-term digital interactions such as team/rider stories and highlights that can be shared and viewed on different platforms. Most sports reporting, however, is still old fashioned and does not fully exploit the technology and digital platforms/tools that exists today. We need more personalized, interactive experiences to keep the end user happy and to get back the youngsters. However, since road cycling broadcasts is also used to promote tourism (i.e., it offers places an invaluable exposure to a broad audience), we
must ensure that this economic aspect can continue to play a role in the proposed new formats [2].

The growing availability of data sources from video or IoT devices (such as sensors attached to bikes or athlete wearables) are offering a huge potential to make reporting more interactive, informative and attractive. These data sources, however, are still hardly used in sports events broadcasts. The missing gap is the adequate translation of the sensor data into useful narrative elements ("something happened") and the selection of the correct video fragments that tell this story. A striking example of a recent ‘story’ that was not seen during the live broadcast of the Tour of Flanders 2019 - but reported in post-race summaries afterwards - was the team tactic of team Sky/Ineos of slowing down the entire peloton at the beginning of several hills to save power for the rest of the race. Being able to detect such kind of actions in an automatic way (based on the sensor data) and create highlights/summaries of it is the first challenge we want to tackle within the paper.

In order to make traditional reporting more interactive we also introduce a novel second screen experience that synchronizes cycling heritage multimedia (from the KOERS 1 museum in Flanders) with the location of the riders in the live race. As such, the locations that are seen on television are enriched with stories of historical cycling events that happened in that particular place/region. During boring moments in the race this can be a tool to retain the viewer’s attention. In a similar way to the location-based enrichments, we are currently also performing experiments to do team/cyclist based synchronization and enrichment.

Both innovations (i.e., the summarization and the second screen enrichment) will influence the way in which end-users experience cycling events and will give the viewers at home multiple and more engaging ways to watch the game based on their specific interests. End-user quality-of-experience has been evaluated in real life conditions by a questionnaire and an objective analysis of their interactions on the platform during the Grand Depart of the Tour de France 2019.

The remainder of this paper is organized as follows. Section 2 presents the related work on data-driven analysis and summarization of sport videos. Furthermore, different geolocalization techniques are discussed to find out at which location the cyclists are in the live broadcast. Next, Section 3 explains our data-driven summarization methodology. Subsequently, Section 4 focuses on the location based synchronization of the second screen cycling heritage data. Results of our Tour de France test are shown in Section 5 and finally Section 6 concludes the paper and points out directions for future work.

2. RELATED WORK

Some state-of-the-art examples, such as the RedBull Ski Downhill augmented reality shown in Figure 2, go in the right direction and already (slightly) improve traditional sports reporting. Their sensor-enhanced experiences, however, can only be seen after the game, and still the decision on which content the end-user will see is entirely up to the editor. It is a very good story telling, but it requires a lot of manual work (i.e., the labeling is not done automatic), and it still is a one-fits-all solution (i.e., not personalized to the end user). A similar trend is the real-time data augmentation in basket and soccer, where sensor data and statistics are integrated in the live broadcasts. However, although it is live and interactive, the data is not always shown at the right moment/place and can distract the viewer. Furthermore, both examples do not provide a solution for the most important problem that contemporary cycling races are struggling with, i.e., that watching the race for multiple hours can be boring and too time consuming. A solution for this problem can possibly be found in the area of highlight summarization, an active research topic over the last decade.

![Figure 2: RedBull Ski Downhill – post-race augmented reality: sensor data is projected onto the slopes to provide race insights.](image)

2.1 Highlight summarization

Traditional summarization methods (e.g. HMM based audio-visual highlight detection as proposed by Qian et al. [3] have been improved by recent deep learning based methods that, for example, focus on player action related features to classify video into interesting/uninteresting parts [4]. The accuracy of these techniques is still too low (+- 75% f-score) and the metadata description is too high level to create summaries targeted to specific end users/applications. We aim to achieve higher accuracy, flexibility and better customization by investigating a novel multi-modal summarization method taking into account the actions detected in the sensor data and the engagement scores of the video content. To select the correct video streams corresponding to a particular sensor event, we will first need to adapt/improve player recognition (e.g. 89% accuracy in basket and 83% accuracy in soccer - as reported by Gerke et al. [5]) and combine it with human action recognition for accurate sensor-

---

1 https://koersmuseum.be
video action matching. Within this paper, we propose a methodology to do this.

2.2 Second screen enrichment

Another solution to make cycling races less boring is to keep the end users active and provide/surprise them with interesting multimedia content (on their second screen) related to what they see on television, similarly somehow as is done in the LinkedTV\(^2\) and the NexGen-TV\(^3\) project [6, 7]. To the best of our knowledge, no such synchronized second screen experience exists today for cycling. Studies by Mukherjee et al. [8] and Giglietto & Selva [9] also show a clear tendency of the user to use a device while watching television, as such there is definitely potential for such kind of second screen experience if it is on-topic, related and presented at the right frequency.

The type of synchronization we focus on in this paper is a location-based matching of the content with the live location of the cyclists in the race. In order to perform such kind of matching we need to know the location of the cyclists at each moment in time. Different techniques can be used to perform this task, such as text recognition (when the kilometer indication is shown in the video feed), scene recognition (based on the environmental features) and live sensor data (from third parties such as Velon and Gracenote) which include the kilometer indication or geolocation coordinates.

2.3 Geolocalization of live race cycling video

Text-based geolocalization can be split up in a detection step and a recognition step. For text detection, EAST for example is a state-of-the-art (SOTA) approach that achieves very good results on natural images. EAST significantly surpasses competitors in accuracy (F-score of 0.782), whilst running very fast [10]. In combination with a text recognition tool such as Tesseract [11], for example, it is possible to achieve very high accuracy (90%) on the recognition of the kilometer indications in the textual overlays in cycling videos. With some additional temporal reasoning, our tests showed that it is possible to always estimate the correct kilometer – results of our EAST/Tesseract solution and comparative Azure output are shown in Figure 3. The problem with this kind of technique, however, is that the kilometer indication is not always visible/available in the video stream and the layout of the overlay can also change across different races. Furthermore, the video processing introduces an additional computational cost which can has its impact on the real-time performance of the overall solution.

Scene-based geolocalization only focuses on the environmental features seen in the video images and tries to match them to a location on the map/route. For images which contain characteristic buildings or environments with some unique features this matching (for example to georeferenced StreetView images) can be done rather accurately [12, 13], but for many locations the matching is not so easy due to high similarity in natural landscapes and urban environments. However, by additional temporal reasoning (taking into account the route of the race) the accuracy can again be improved, but our tests revealed that it will still be too low to know the exact location during the entire race. Similar to the previous method these kind of methods also introduce an extra computational cost which is even higher due to the complex matching of building and environmental features.

Sensor-based geolocalization is the last technique we evaluated in this study. Several third parties (such as Velon\(^4\) and Gracenote\(^5\)) provide structured and very detailed live data of sports events at a high frequency. Velon, for example, provided location, heart rate, power, and speed of each cyclist during several stages of Tirreno Adriatico 2019 and Gracenote provided exact location of each group of riders during the Tour of Flanders 2019. If such sensor data of the race is available, this is definitely the most accurate and computationally most interesting solution for geolocalization. When there are multiple groups of riders, however, an additional method (such as team or cyclist recognition) is needed to know which particular group is shown in the live video stream (as is further discussed in Section 3). Figure 4 shows our results for the team recognition method on Strade Bianche 2019 content – team jerseys are extracted using OpenPose skeletons [14] and the team is estimated using our own trained CNN for team recognition that uses transfer learning on a small dataset of team jerseys.

---

\(^2\) http://www.linkedtv.eu/
\(^3\) http://www.nexgentv.fr/
\(^4\) https://www.velon.cc/
\(^5\) http://www.gracenote.com/sports/global-sports-data/
Figure 4: Cyclist recognition - tests performed on Strade Bianche 2019 footage. Skeletons can also be used to detect abnormal behavior/events (~ future work).

3. DATA-DRIVEN SUMMARIZATION OF CYCLING RACES

The proposed data-driven summarization methodology is shown in Figure 5 and consists of 3 steps. First of all, we look for events in the sensor data. In our study we focused on data from Velon, Gracenote and Strava, but other type of data sources can also be used. Based on the Velon and Gracenote data we perform a speed/group-based event detection which compares the speed of each rider with the average speed of the group he is part of. Combined with the location of a rider relative to the median location of the group the algorithm decides if an event is occurring and if it has impact on the race situation. Furthermore, an engagement score is generated based on what happened previously in the race and the gathered knowledge learned from previous events/races. We mainly follow the highlight and detection methodology of O'Donoghue et al. [15] and will extend it to an engagement scoring mechanism. This extension will allow detection of an attack, crash, or move to the front/back of the group and and ultimately also provide it with an appropriate engagement score. Figure 6 shows an example of an attack detected based on Strava data – similar events can be detected in Velon/Gracenote data. In order to know to which group each rider belongs and to know the location of that group, we use the JSON data from the Gracenote API. For the speed data of each rider, we use Velon.

The second step of our summarization algorithm focuses on the video data and analyzes who/what is present in the video images. First, we perform a shot-type detection. We focus on the following 6 cycling-specific shot types: close-up, small and big group, birds eye view, interview and touristic shot. Our algorithm follows a similar workflow as the field sports shot classifier proposed by Minhas et al. [16], which is based on AlexNet CNN. For the close-up and small group shots, we then use the OpenPose pose recognition mechanism and combine it with an object classifier (e.g. YOLO or Mobilenet [17]) to validate that the detected skeleton is positioned on a bike. If not, the skeleton is not further investigated (e.g., the motorcyclist in Figure 4). For the cyclist skeletons, the shirt is extracted by cropping a rectangular region from the upper body joints of the skeleton. A result of such a crop is shown in Figure 7.
Figure 6: Sensor-based detection of events and matching (~validation) with live video during Gent Wevelgem 2019.

Next we feed the crop to our jersey recognition algorithm that predicts the team to which the cyclist belongs to. The algorithm is based on the image retrieval work of Radenovic et al. [10] and Generalized-Mean (GeM) features are used to find the best match in the collection of trained cycling jerseys. One could also think of using live audio commentaries and use audio keyword spotting on rider/team names, however, this has not been investigated yet but will be evaluated in future work. By combining the team info from all detected riders with the Gracenote group info, we can detect which group is shown, which facilitates rider recognition. Once we have matched the rider in the video with the events in the sensor data, we have all necessary metadata to generate interesting and personalized summaries, such as stories of your favorite team.

Figure 7: Result of cycling jersey extraction using OpenPose-based skeleton joints for upper body cropping. Furthermore we show Azure’s text recognition results for the starting number.

The engagement scoring of events can also be further optimized by taking into account stage info and external data, such as Points of Interest (POIs) along the route. In this way, a mix of different environments/topics can be shown within the generated summaries. A digital map of relevant cycling POIs, shown in Figure 8, has been created by KOERS. This type of data can be useful as side-information for our data-driven summarization, but also for the second screen enrichment that is discussed in the next section.

Figure 8: Cycling related Points of Interest in Belgium – a digital map created by KOERS.

4. SYNCHRONIZED SECOND SCREEN ENRICHMENT

As was discussed in the related work section of this paper (Section 2) it is nowadays technologically possible to detect the location of the cyclists in the race using a variety of algorithms. Feasibility tests, of which results are shown in Figure 9, confirm this hypothesis, however more quantitative studies needs to be undertaken to be sure it will always work. Based on the location info, relevant content for that location can be shown on the second screen (such as the cycling related POIs discussed above).

Figure 9: Geolocalization of live cycling video images using Gracenote’s sensor data and the team recognition algorithm. The left image shows the location on the map, the right image shows the live video.

The geotagged content that will be synchronized with the live race can be composed (semi-)automatically using the content
dashboard shown in Figure 10. First of all, the race route (gpx) needs to be loaded. Next, different geotagged datasets (e.g. Wikipedia stories, heritage POIs, and cycling fan clubs) can be fit on the race/stage timeline in an automatic way. If an item is close enough (less than 2km) to a segment of the stage, it is linked to that segment. The content manager can easily evaluate if he has enough content for each segment of the race and (if needed) can move content to other segments. Additionally, manual content can also be added with a content CMS that has been built for this project. Currently, we only support GeoJSON datasets to be automatically fitted on the timeline, but several algorithms to generate them are available on our Github6.

Once all content is ready, the timeline can be exported to the final GeoJSON file which will be used during the live broadcast of the race. This GeoJSON file is used by the end-user application shown in Figure 11. Within the application, we show the current location of the race on a map and a timeline and enrich it with the content that is relevant at that moment. The user can swipe or click through the different content items and can zoom in on pictures and 3D visualizations (e.g. of cycling jerseys). Furthermore, links to other websites (such as Wikipedia, Youtube or the personal websites of the cyclists) can be added for end-users that are interested to read/see more. The second screen application for the Grand Depart of the Tour de France 2019 in Brussels is made available in preview modus (i.e., not live) to be able to showcase/demonstrate it when the race is over. You can test our TDF demonstrator on: https://users.ugent.be/~jcdbock/granddepart/preview/.

5. EVALUATION

The proposed methodology for data-driven summarization and synchronized second screen enrichment has already been tested twice; the first time during E3 Harelbeke (March 2019) and the second time during the Grand Depart of the Tour de France 2019 (July 2019). 22 end-users, with an average age of 32 and a gender distribution of 71% men vs. 29% women participated in our tests. Most of them reported watching live cycling races on a regular base, however 75% only said to be only interested in the last 20km of the race. More than 50% also uses a mobile/second screen to follow the race online and a majority of them (>90%) reads articles about cycling. The traditional broadcasting is considered as too boring by 52% of our test audience and 75% of them is open to get more side-info (such as touristic information of the region, cycling history, etc.).

Personalized summaries, e.g. team stories of their favorite team or a highlight summary of 5 minutes, were found to be very innovative and a worthy alternative to traditional broadcasts. Important to mention is that the length of the summaries can be chosen by the user itself, and in a similar way the enrichment will be made more dynamic in the future (when more content is available to map on the stages). Further improvements, such as an audio-based shot cropping or textual overlay with the story/event descriptions, were also suggested and will be taken into account in the next version of the algorithm. For the second screen application, Hotjar7 and Google analytics results showed that the end-users were rather active and explored the different components of the platform. In the subjective feedback (that was collected with a Google form) they were also mostly positive, but they would like to be able to have more choice in the type of content that is presented to them. However, this is more a content-related problem than a technological shortcoming of the

---

6 https://gitlab.ilabt.imec.be/jcdbock/koers

7 https://www.hotjar.com/
platform. Gamification was reported several times as a method to further improve the second screen experience. Some steps in this direction have already been taken in the Grand Depart test, e.g., end-users can play a Yellow Jersey Quiz and post their results/badges on social media. UI also needs to be improved.

5. CONCLUSIONS

Within this paper we propose two different methodologies to make cycling broadcasts less boring. Firstly, using our data-driven summarization method, automatic/personalized post-game summaries can be generated focusing on one particular rider/team and at different segments of the tracks where engagement is high. Secondly, we propose a methodology to synchronize cycling heritage multimedia data with the riders’ live location during the race. Both rich, interactive TV experiences are evaluated by a test audience and results showed that there is certainly potential in both formats. Future work will mainly focus on further improving the user experience by, for example, automatically generating dynamic visualizations of the events and race situation, and by incorporating on board and on bike camera feeds of teams/riders when broadcasters do not have video footage of an event. In order to better evaluate the impact of the presented tools, the methodology of the questionnaire also needs to be further improved in the future.

ACKNOWLEDGMENTS

The KOERS ‘Museum in de living’-project, of which results/data is discussed in this paper, is funded by the Flemish Government – department of Culture, Youth and Media.

REFERENCES