AdsISee: Advertisement Detection and Tracking for Sponsorship Evaluation in Soccer Matches

Alexander Westermann*
Institute of Applied Information Technology
Zurich University of Applied Sciences
Winterthur, Switzerland
alw.westermann@gmail.com

Philipp Krayer*
Institute of Applied Information Technology
Zurich University of Applied Sciences
Winterthur, Switzerland
philipp.krayer@protonmail.com

Andreas Weiler
Institute of Applied Information Technology
Zurich University of Applied Sciences
Winterthur, Switzerland
andreas.weiler@zhaw.ch

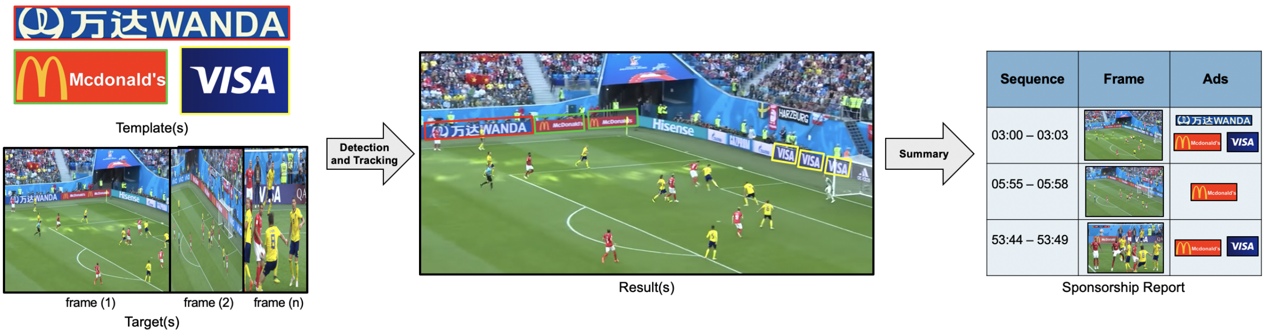

Figure 1: Example of AdsISee with three ads in the soccer match Sweden against Switzerland at the World Cup 2018.

ABSTRACT

In this work, we present *AdsISee*, a real-life application for the detection and tracking of advertisements in broadcasts of soccer matches for supporting business analysts in the task of sponsorship evaluation and reporting. Our approach is based on different combinations of several techniques for object detection and tracking in images. In contrast to other works which use the technology of neural networks, we use alternative solutions to detect advertisements based on provided pre-defined image templates and without any training period. Hereby it was possible to build an application which can be executed on standard hardware by still providing a feasible performance. Furthermore, our evaluations show that we can achieve comparable results against other existing approaches, which use neural networks, for sponsorship evaluation.

1 INTRODUCTION AND MOTIVATION

The market of advertisements in sport events is tremendous. For example, advertisers had to pay US $5.25 million to air a 30-second long commercial during the Super Bowl 2019 broadcast [12]. However, besides the advertisements, which are explicitly shown to the television viewers, there are advertisements, which are directly placed in the sport events itself. These advertisements are shown on the margins of the playing field as perimeter advertising, on the clothes of the players, or somewhere else in the real-world environment of the sports event. For example, the FIFA World Cup 2018 brought an additional of US $2.4 billion to the global advertising market and the sponsors paid up to US $200 million for a sponsorship package [3]. The target audience of these advertisements are primarily the on-site visitors. However, also the television viewers are important receivers of these kind of advertisements. For example, an average of about 191 million viewers watched the soccer matches at the World Cup 2018 as live broadcast [4]. Additionally, millions of viewers are watching full or recapped recordings of soccer matches at any time. In respect to these numbers, every second is important in which the advertisements can be seen on screen by the viewers.

In this work, we present *AdsISee*, a real-life application for the detection and tracking of pre-defined advertisements in the frames of soccer match broadcasts. By using AdsISee, we are able to generate a sponsorship report to support business analysts in their task of sponsorship evaluation and further analysis. We use different combinations of several techniques for object detection and tracking like the FLANN [9] matcher or the MOSSE [7] tracker. In contrast to other works, in which neural networks [1, 8] or the Haar cascade detector [11] are being used, which need to be trained on the detectable object beforehand, we use alternative solutions to detect advertisements based on provided templates and without any training period. One advantage of our implementation[1] is that it can be executed on standard hardware by still providing a feasible performance. Therefore, the main goal of this work is to evaluate various technologies in the field of visual computing and applying them in a domain specific area in order detect objects without the use of machine learning methods. This offers the advantages of less configuration, no manual annotation of advertisements and no learning process in order to successfully detect the advertisements. Our experiments (see section 3) show that it is possible to detect and track advertisements in different quality levels. To evaluate our

*Both authors contributed equally to this research.

[1] https://github.com/AdsISee/AdsISee (November 20, 2019)

work, we created several case studies for different soccer matches. Furthermore, we compare our application against the solution of Orpix ComputerVision Inc. [5], which is the market leader in the area of sponsorship evaluation who uses trained neural networks in order to detect the advertisements in broadcasted soccer matches. The detailed evaluations provide insights into the advantages and disadvantages of the used technologies in order to detect and track advertisements.

2 METHODOLOGY

Our developed solution detects advertisements on the margins of the playing field in broadcasts of soccer matches based on templates in the format of pre-defined images. The broadcast of the soccer match is divided into its individual frames, which are used as target images and are processed in a streaming fashion one after the other. In each of the target images our solution tries to detect the template image for the matching advertisements. In a first step, relevant technologies for object detection in the domain of advertisements were evaluated. By analyzing a common set of advertisements in soccer matches, we were able to discover certain properties, which can be used for the recognition of the ads in the target images. One of the main characteristics are the clearly distinguishable colors with high contrast which supports the viewers and our application to better recognize the exposed brands. By analyzing and extracting the different color components of the template it is possible to filter out large areas of the target image which do not correspond with the colors of the target image. Accordingly, the search area for the advertisements can already be severely reduced.

2.1 Feature Detection

Another unique characteristic of advertisements on banners are the simple geometric properties of the exposed logos. Edges, corners and flat surfaces can be extracted out of the advertisement templates in order to match them in the target images. Therefore, two different methods have been applied to detect those unique features with their positions from the templates and target images. One of the main challenges with this approach is, that in most cases the perspectives and sizes of the advertisements in the target images do not correspond to the advertisements in the templates. The perspectives of the advertisements in the target images depend on the angle and position of the camera recording the sport event. We solved this issue by using the Scale Invariant Feature Transformations (SIFT) in our implementation, which is a technology that was originally designed for panoramic image stitching [2]. By applying the SIFT algorithm on the advertisement templates and target images, unique features (*cf.* Figure 3) such as corners, edges and flat areas can be extracted regardless of their scaling and perspective. To improve the accuracy of the feature detection, each advertisement template is scaled to one of the three main perspectives, in which the banners occur during the broadcasts. These perspectives include the frontal directly visible and the positions on the left or right of the field (*cf.* Figure 2).

Figure 2: The three possible perspectives of an banner ad.

Figure 3: Example of detected features in a template.

2.2 Feature Matching

After extracting the features from the template and target, each feature from the template has to be searched for a match with a feature from the target. An example can be seen in Figure 4. The exact position of the advertisement in the target image can be calculated as soon as the number of matches exceeds a threshold. We use in our implementation the Fast Library for Approximate Nearest Neighbors (*FLANN*) in order to calculate matches between the template and target image. The FLANN matcher calculates the nearest neighbors between the properties of the detected features which are represented by a distance. By applying a threshold for acceptable distances, we are able to distinguish between correct matches, which belong to the advertisement, and incorrect matches.

Figure 4: Example of feature extraction and matching.

The Brute-Force matcher offers a faster alternative to the FLANN matcher, which compares each feature from the template with all the extracted features from the target image and matches the features with the smallest difference [6]. Our evaluations show that the Brute-Force matcher was able to perform faster than the FLANN matcher, however, the accuracy was slightly reduced. Accordingly, we our solution applies the Brute-Force matcher additionally to the FLANN matcher for cases where the performance has to be maximized.

2.3 Matching Multiple Advertisements

During our evaluations we figured out that the matching has major issues by detecting more than one identical advertisement, which is visible in the target image. For example in Figure 1 the application needs to detect the advertisements of McDonalds and Visa more than once. In this case the approach of using the Brute-Force and the FLANN matcher detects identical features of the ads accordingly and therefore it is impossible for feature matching to distinguish the individual features between the same advertisements. To solve this problem, we have developed our own solution which allows to differentiate extracted features between the same advertisements. After an advertisement has been detected, the search for the same advertisement is repeated, excluding the features of the already matched advertisement. This procedure is repeated until no new advertisements are detected. This allows our solution to allocate each feature from the template advertisement to multiple features in the target image according to the amount of the identical visible advertisements.

2.4 Tracking

The perquisite for feature detection and matching is that the searched object has to be sharp. However, due to the movement of the camera, in most cases the advertisements often appear blurry in the target image. As a result, no sharp edges or corners can be extracted out of the target image and the application is not able to detect the advertisement. For these cases we used a combination of tracking technologies, the Median Flow [13] - and the MOSSE [7] tracker, in our solution. In order to detect blurred advertisements nevertheless, we have applied tracking technologies to follow the movement of advertisements in the target image once they have been detected. After evaluating various tracking technologies, we decided to implement the MEDIANFLOW- and the MOSSE tracker which show high performance and accuracy as our experiments have shown. Once an advertisement has been detected through SIFT and feature matching, it will be registered in a tracker. For all following frames, the tracker is updated which determines the exact position of the tracked advertisement. By calculating a matrix for the perspective transformation out of the previously detected advertisement, we were able to reconstruct the exact place and perspective of the tracked advertisement.

As an added benefit, this approach increases the performance of the entire software. The detection of advertisements through feature matching is a relatively time-consuming process. Once an advertisement has been detected in a frame, it can be tracked for all subsequent frames and thus no longer has to be discovered again. Besides of advertisement tracking, this technology can also be used to track advertisement-free areas. As soon as no advertisements have been detected in a frame, the complete target image is marked for tracking, which will be excluded for all the following advertising searches (*cf.* Figure 5). The impact the accuracy and performance of the advertisement and empty area tracking will be tested in a detailed evaluation (*cf.* Section 3).

3 EXPERIMENTS

In a detailed evaluation we tested both the accuracy and the performance of our application in advertisement detection in broadcasts of sport events. In the first series of experiments, the influence of the individual components has been evaluated. All the experiments have been performed on an Ubuntu-Machine with 2.01 GHz, 8 CPU cores and 16 GB RAM.

3.1 Evaluation of Functionalities

In a first experiment we have tested the impact of compressing the target in a pre-processing step. We reduced the resolution of the target image by 50% and compared it with a test run without target compression.

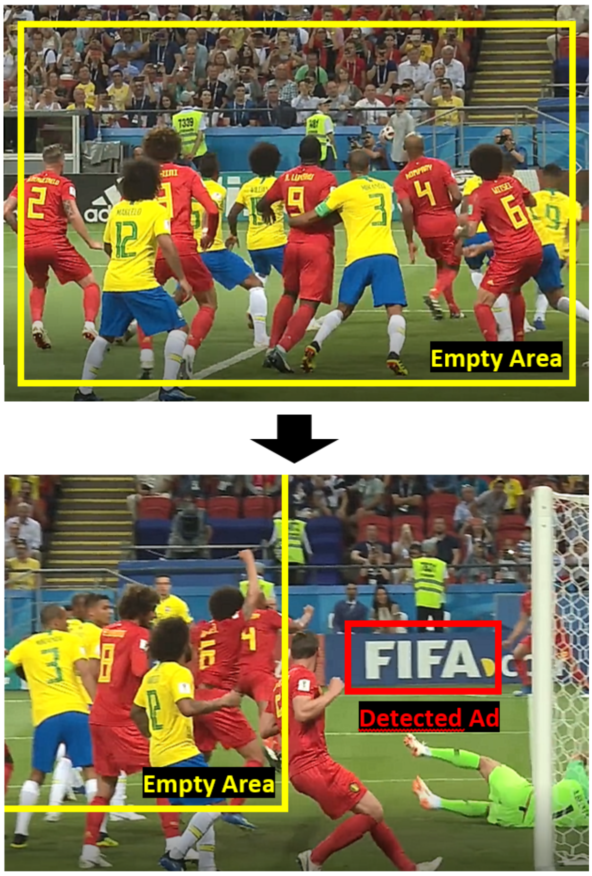

Figure 5: Visualization of empty area tracking.

Target compression:	**Yes**	**No**
Average accuracy:	71%	88%
No. of errors:	1	0
Average time:	1.03 sec.	4.12 sec.

Table 1: Evaluation of target compression

The results in Table 1 show that reducing the resolution of the target image by 50% results in an increase in performance by 400%. However, the accuracy decreases by 20%. Reducing the accuracy erases some detectable features since small edges and corners disappear. Accordingly, fewer features have to be matched in order to detect the advertisement which decreases the search process. If the advertisement is poorly visible, not enough features can be extracted for a successful match, which explains the slight reduction in accuracy.

In another experiment we have tested the impact of each individual feature on the performance and accuracy of the advertisement detection.

Color filtering	Matching algorithm	Tracking algorithm	Average time	Average accuracy
Off	Brute-Force	MF	4.75	84%
On	Brute-Force	MF	4.65	82%
Off	Brute-Force	MOSSE	4.47	82%
On	Brute-Force	MOSSE	4.61	83%
Off	FLANN	MF	5.25	86%
On	FLANN	MF	4.78	85%
Off	FLANN	MOSSE	5.40	85%
On	FLANN	MOSSE	5.06	85%

Table 2: Comparison of various functionalities

This evaluation shows that filtering out the irrelevant colors in a pre-processing step increases the performance slightly, however, resulting in a slight decrease in accuracy (*cf.* Table 2). This can be explained by the color filter interfering with the features of advertisements in the target images. The evaluation of the two different feature-matchers shows that the FLANN matcher performs in terms of accuracy better than the Brute-Force matcher. Though, the FLANN matcher took on average 0.5 seconds longer than the Brute-Force matcher to calculate the matches between the target advertisement and the template image. In addition to the matching algorithms, the two different tracking methods have been compared. The MOSSE tracker showed a slightly better performance without a significant reduction in accuracy compared with the MEDIANFLOW tracker. Accordingly, to these results, all features have been implemented in our solution and the user can decide whether the performance or the accuracy should be enhanced for the advertisement evaluation.

3.2 Ideal Matching Difference

In order to successfully match the features from the advertisement templates with the extracted features from the template, it is necessary to filter out the wrong matches. Each extracted feature from the template advertisement will be matched with the most similar feature from the target image and the difference between the two features is calculated. If there are no advertisements in the target image, a non-existent advertisement will still be detected, but with features whose differences are much higher compared with those who matched a correct advertisement. To prevent this, a threshold is defined for the highest acceptable difference in matches between features. A too high threshold would result in matches which do not belong to an advertisement and accordingly with a too low threshold correct matches would be filtered out. The following experiment has been performed for the purpose of finding the ideal threshold for acceptable matching differences. The results in Figure 6 show that the ideal threshold for acceptable matching differences should be between 0.7 and 0.75 to achieve the best results. This threshold is implemented accordingly in our solution.

3.3 Ground Truth Evaluation

In this evaluation phase we run several tests for optimizing the configuration of AdsISee for maximal accuracy for advertisement detection. The goal was to compare this software with excerpts from live broadcasts of sport events and determine its overall accuracy. In a first run, the software was tested without the use of tracking. This determined the accuracy of the plain detection process of advertisements. 30 frames with clearly visible advertisements and 30 frames without advertisements have

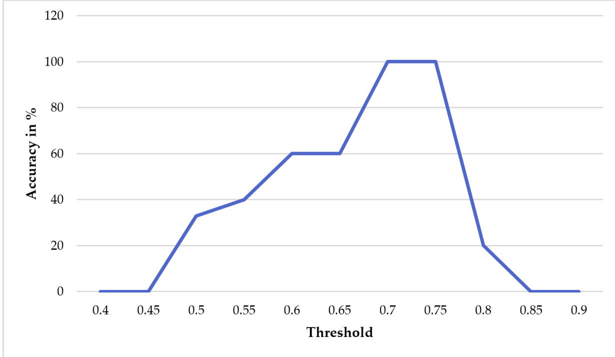

Figure 6: Accuracy with different thresholds

Figure 7: Partially covered ad could be detected

been selected and tested on this software. 71% of all ads have been successfully detected without any incorrectly detected non-existing advertisements. In some cases, the advertisements could be detected, although they were partially covered (*cf.* Figure 7).

In a second run AdsISee is evaluated on various video scenarios of live broadcast soccer matches. In each video clip the advertisements were visible with different properties, which tested the limitations of the software.

Camera movement:	Sudden appearance of ad:	Sudden disappearance of ad:	Accuracy:
Slowly	No	No	98%
Slowly	Yes	No	97%
Slowly	Yes	Yes	98%
Fast	No	No	85%
Fast	No	Yes	91%
Fast	Yes	Yes	70%

Table 3: Comparison of various video scenarios

In Table 3 it is visible, that on average our solution could detect the advertisements with high accuracy of 90%. The tracker was able to track the movements of the advertisements of even fast camera movements. Sudden appearances of the advertisements were always correctly recognized by the feature detection and matching component after a maximum of 3 frames. However, in the last test video, the advertisement was slowly faded away by an animation on the banner-screen. Since the used tracker technologies cannot detect the disappearance of an object slowly fading away, the position of the advertisement was continued being tracked even though the actual advertisement already disappeared. This resulted in a sharp drop of the measured accuracy.

3.4 Comparative Evaluation

Orpix ComputerVision Inc. offers a cloud-based solution for evaluations of advertisement occurrences during live broadcasts of sport events. This solution uses a state-of-the-art convolutional neural networks [10] to detect the advertisements and process the target images in a frame rate of 1 FPS. Neural networks provide excellent results in object recognition, but have the disadvantage that they need to be trained by an elaborate process on the object beforehand. This involves annotating advertisements in hundreds of example templates by hand.

In this test, the accuracy of the product of Orpix is compared to that of our solution. The goal is to show the advantage of AdsISee, that it can perform a sponsorship evaluation by providing only proper advertisement templates, without having to train any algorithm or making any configurations beforehand. Orpix provides one free online example of a sponsorship evaluation of the final game France versus Croatia at the FIFA World CUP 2018, which will be used for the comparison against our solution. The computational performance of their solution is not mentioned by Orpix. Accordingly, no accurate comparison in performance can be made between AdsISee and the solution of Orpix.

Based on 5 different advertising templates, the software was tested on randomly selected video sequences of the World CUP 2018 final game. The tagged advertisements were compared with the individual frames from the example evaluation from Orpix.

Template:	Accuracy Orpix:	Accuracy AdsISee:
Wanda	93.75 %	90.00 %
Hyundai	88.50 %	73.25 %
Qatar Airways	83.25 %	91.00 %
Hisense	99.50 %	92.75 %
Gazprom	96.25 %	40.00 %

Table 4: Comparison between Orpix and AdsISee.

All tests were performed on an Ubuntu system (2.01 GHz, 8 CPU's, 16 GB RAM). Our application needs 500 MB RAM and processes the frames on average with 0.981 seconds per frame. The results (*cf.* Table 4) show that our software performs with an average accuracy of 77.40 % and is therefore slightly lower than the solution from Orpix, which had an average accuracy of 92.25 %. This difference can be explained by the occurrences of animations in the advertisement's banners, which could only be observed in this particular game. Normally most of the times static banners are used in sport events. As a result, the tracker was not able to detect the change of advertisements on the banners and false advertisements were tracked in all subsequent frames. In addition, the advertisement templates were only available in reduced quality which limited the amount of extractable features (as explained in Section 3.5). Also, the solution of Orpix was in contrast to AdsISee able to detect advertisements that were far away from the camera and even hardly recognizable by the human eye.

Nevertheless, in some cases our software was able to detect the advertisements more precisely than Orpix. Often, the solution of Orpix detected the advertisements twice or more, resulting in an inaccurate tagging of the corresponding advertisement (*cf.* Figure 8). In some cases, several adjacent advertisements have been marked as one single advertisement. Additionally, our solution is able to determine the positions of the advertisement with better precision than the solution of Orpix. Unfortunately, the report about the final match at the FIFA World Cup 2018 was the only report, which Orpix provided and therefore our comparable evaluation is just based on this single event and report.

Figure 8: Example of an advertisement that is tagged twice by Orpix.

3.5 Problems Encountered

In order to extract as many different features as possible from the advertising template, the images have to be provided in a good quality. During our search for advertisement templates, we were not able to find any high-quality templates that matched the commercials which appeared on the banners. Accordingly, we extracted the advertisements from broadcasts of sport events manually, acquiring templates with slightly reduced quality. Thus, the tests were not performed with the best prerequisites that could have been possible. Therefore, it can be assumed that AdsISee can achieve even better results by providing high-quality advertisement templates as input.

Our applied tracking technologies are ideal for tracking the movement throughout the screen of detected advertisements. The tracker is able to recognize if the advertisement suddenly disappears and terminates accordingly the tracking phase. However, if the banner changes the advertisement by an animation, the tracker is not able to detect this and the wrong advertisement is continuously tracked. This resulted in some of our experiments in a reduced rate of accuracy.

During the evaluation phase, we observed that in some rare cases non-existent advertisements have been falsely detected (*cf.* Figure 9). We were able to partially solve this issue by implementing a filter which checks the positions and relations of the detected advertisements. This filter prevented most of the false-detected advertisements we encountered during our evaluations. Te filter validates the calculated frame of the advertisement before the detected object is registered for the tracking. We defined the following conditions, which have to be met by the detected object to recognize it as an advertisement:

- The edges of the marked object must not cross each other.
- The aspect ratios of the detected object must match those of the template advertisement.

Figure 9: Example of a falsely detected advertisement.

4 DEMONSTRATION

To demonstrate the functionality and usefulness of AdsISee we created a video compilation[2] with examples of the detection and tracking of several advertisements in soccer and hockey matches. As templates we use a selection of 18 different advertisements (*cf.* Figure 10) of various companies. We tried to cover a wide range of colors and shapes with our template collection. Our demonstration provides examples for the detection and tracking of single or multiple advertisements in static and also fast-moving frames. We also demonstrate in the video that AdsISee occasionally has problems (see Section 3.5) with the detection and tracking of the correct objects. For example, the country name "USA" in combination with the blue color and the banner format (see timeframe 0:39 to 1:30 in the video) is detected and tracked as the Visa advertisement.

Figure 10: Advertisement templates used for the demonstration video.

Additionally, to the video output with the marked advertisements, AdsISee creates a report as a standard text file containing the information about the sequences, frames, and the detected ads. This report can be used to visualize (*cf.* Figure 1) and analyze complete broadcasts of sport events in a compact view. For example, the report can be grouped by time sequences, advertisements, or the number of detected advertisements. This can support analysts in figuring out the sequences with the most advertisements or comparing the on-screen time of the own advertisement with others.

[2]https://youtu.be/KReFUcKiw4E (November 27, 2019)

Figure 11: Detection and Tracking of ads in hockey games with AdsISee.

5 CONCLUSIONS AND FUTURE WORK

In this work, we have demonstrated that it is possible to detect advertisements in broadcasts of sport events, especially soccer matches, for sponsorship evaluation and analysis. We were able to apply alternative technologies for object detection in a specific domain and show certain advantages compared to the state-of-the-art technologies. We have successfully implemented a prototype that can detect advertisements in broadcasts of sport events by only providing templates of advertisements. AdsISee extracts unique features such as color, edges and corners out of the template and matches those with each individual frame from the live broadcast. With the implemented tracking technologies, it allows to track the movement of the advertisements throughout the screen. We tested the prototype for accuracy and performance in an extensive evaluation phase. In addition, AdsISee was compared with a product of Orpix which is the leader in the area of sponsorship evaluation in sport events, where similar results have been measured. Our solution showed some advantages over the product of Orpix, for example, the tagged position of the advertising was calculated which a better precision than the product of Orpix.

For future work we plan to extend our approach to further sport events and also to other detection and tracking areas in the frames, where ads could be placed. For example, Figure 11 shows the result of our approaches to detect and track advertisements in hockey games on the margins of the playing field, as well as on the jerseys of the players. Additionally, many improvements can still be implemented and tested in future. The filter for detecting false positives could be improved by comparing the colors of the matched advertisement with the templates. The advertisement will not be tagged if the colors do not match the template, which would also eliminate the issue with animations. Furthermore, since most advertisements contain a large part of text, adding an additional text recognition feature could improve the accuracy of AdsISee.

REFERENCES

[1] Khaled Almgren, Murali Krishnan, Fatima Aljanobi, and Jeongkyu Lee. 2018. AD or Non-AD: A Deep Learning Approach to Detect Advertisements from Magazines. *Entropy* 20, 12 (2018).

[2] Matthew Brown and David G. Lowe. 2006. Automatic Panoramic Image Stitching using Invariant Features. *International Journal of Computer Vision* 74, 1 (2006), 1–15.

[3] Marketing Charts. 2018. World Cup 2018 Stats. https://www.marketingcharts.com/industries/sports-industries-83790. [Online; accessed 25-October-2019].

[4] Fifa.com. 2019. More than half the world watched record-breaking 2018 World Cup). https://www.fifa.com/worldcup/news/more-than-half-the-world-watched-record-breaking-2018-world-cup. [Online; accessed 25-October-2019].

[5] Orpix ComputerVision Inc. [n.d.]. Sponsorship Analytics and Evaluation-Orpix-Computer Vision. http://www.orpix-inc.com/sponsorship-valuation-analytics/. [Online; accessed 25-October-2019].

[6] Amila Jakubovic and Jasmin Velagic. 2018. Image Feature Matching and Object Detection Using Brute-Force Matchers. *2018 International Symposium ELMAR*

(2018).
[7] Peter Janku, Karel Koplik, Tomas Dulik, and Istvan Szabo. 2016. Comparison of tracking algorithms implemented in OpenCV. *MATEC Web of Conferences* 76 (2016), 1—-6.
[8] Shervin Minaee, Imed Bouazizi, Prakash Kolan, and Hossein Najafzadeh. 2018. Ad-Net: Audio-Visual Convolutional Neural Network for Advertisement Detection In Videos. *CoRR* (2018).
[9] Marius Muja and David G. Lowe. 2009. Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration. In *Proc. of the Fourth International Conference on Computer Vision Theory and Applications (VISAPP 2009)*. 331–340.
[10] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2017. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39, 6 (2017), 1137—-1149.
[11] Sander Soo. 2014. Object detection using Haar-cascade Classifier. Institute of Computer Science, University of Tartu.
[12] Statista.com. 2019. Super Bowl average costs of a 30-second TV advertisement from 2002 to 2019 (in million U.S. dollars). https://www.statista.com/statistics/217134/total-advertisement-revenue-of-super-bowls/. [Online; accessed 25-October-2019].
[13] Kalal Zdenek, Mikolajczyk Krystian, and Matas Jiri. 2010. Forward-Backward Error: Automatic Detection of Tracking Failures. In *20th International Conference on Pattern Recognition (ICPR)*. 2756–2759.