Replay Detection and Multi-stream Synchronization in CS:GO Game Streams Using Content-based Image Retrieval and Image Signature Matching

Van-Tu Ninh[1*], Tu-Khiem Le[1*], Duc-Tien Dang-Nguyen[2], Cathal Gurrin[1]
[1] Dublin City University, Ireland
[2] University of Bergen, Norway
tu.ninhvan@adaptcentre.ie,tukhiem.le4@mail.dcu.ie,ductien.dangnguyen@uib.no,cathal.gurrin@dcu.ie

ABSTRACT

In GameStory: The 2019 Video Game Analytics Challenge, two main tasks are nominated to solve in the challenge, which are replay detection - multi-stream synchronization, and game story summarization. In this paper, we propose a data-driven based approach to solve the first task: replay detection - multi-stream synchronization. Our solution aims to determine the replays which lie between two logo-transitional endpoints and synchronize them with their sources by extracting frames from videos, then applying image processing and retrieval remedies. In detail, we use the Bag of Visual Words approach to detect the logo-transitional endpoints, which contains multiple replays in between, then employ an Image Signature Matching algorithm for multi-stream synchronization and replay boundaries refinement [1]. The best configuration of our proposed solution manages to achieve the second-highest scores in all evaluation metrics of the challenge.

1 INTRODUCTION

Game analytics is a new research area in recent years that has become popular due to the explosion of both modern live streaming technologies (e.g., Twitch, Discord, Youtube) and e-sports developments [2, 4]. Although e-sports has gained significant attention from audiences across a variety of ages, not much research has been conducted to analyse the game play to provide the spectator with either a quick look into a match (e.g., player's performance statistics, highlights, critical moments, summary) or a deep insight into a game (e.g., whole game statistics, outstanding team's playing style analysis, playing strategy trending analysis).

In GameStory: The 2019 Video Game Analytics Challenge, the organisers use the same dataset used in the previous year, which was provided by ZNIPE.tv [8]. More details of the data and task descriptions can be found in [8].

2 RELATED WORKS

A lot of research in sports videos was conducted to analyse the highlights, the replays, and the story of the match. Jinjun Wang, et.al. performed a shot classification followed by a proposed scene transition structure analysis on the labels of the classified shots to detect replay scenes in sports [12]. Later, Bai Liang, Song-Yang Lao, et.al. proposed the Perception Concept Network-Petri Net (PCN-PN) model to search and locate interesting events in sports videos [5, 10], which is also a potential approach to detect replays as they are usually critical events in the match. In 2019, Ali Javed, et.al. proposed a novel approach for key-event detection and summarization using Confined Elliptical Local Ternary Patterns (CE-LTPs) to extract feature of motion history image for each key-event candidate lying between the beginning and end of a gradual transition, then applying Extreme Machine Learning (EML) to learn the pattern to detect the key-events in sports videos [3]. The results of this work can be used for both replay detection and story summarization of different kinds of sports videos.

3 APPROACH

In the spirit of previous efforts in the field of video event detection and segmentation, we propose a straightforward method to detect replays in videos which are bounded by Intel Extreme Masters logo transition scenes. The logo in these frames is shown in a plain blue background and occupies approximately 80% of its image. We propose to use a Bag-of-Visual-Words (BOVW) approach to detect these two endpoints of each replay segment and use image signature matching to synchronize multi-stream players' views.

3.1 Logo-bounded Video Detection

3.1.1 Frame Extraction and Filtering: Firstly we use the *ffmpeg* tool to extract frames from commentator stream. To capture the frames in which the Intel Extreme Masters logo is clearly visible, we perform frame extraction at fps=2. As the video length is long (approximately 12 hours), many redundant frames are generated. We reduce the number of extracted frames by using a two pointers technique to eliminate consecutive frames which have similar ORB features [11] and the degree of similarity of a gray-scaled color histogram. For the ORB features, we only consider the top 500 features in total for comparison. Specifically, at frame i, we iterate through its next consecutive frames until we reach a frame j ($j > i$) such that the number of matched ORB features does not exceed α and the L2 distance between the two corresponding histograms is greater than or equal to β. In our implementation, we choose $\alpha = 200$ and $\beta = 0.8$ ($\beta \leq 1$), which means that 60% of the ORB features are different and the degree of similarity between the two color histograms is small. The ones between the frame i and j are redundant and do not provide any extra information for the next steps.

[1]https://github.com/nvtu/gamestory_mediaeval2019

* These two authors contributed equally.

MediaEval'19, 27-29 October 2019, Sophia Antipolis, France

3.1.2 Logo-bounded Video Detection using a Retrieval Approach: From the result in section 3.1.1, we use this filtered data to construct a dictionary of visual words (codebook). We run K-Means clustering [6] on the rootSIFT-feature data points [1, 7] extracted from images in the same corpus. Before processing through the remaining BOVW's steps, frames are center-cropped to reduce background noise. Another purpose is to force the model to focus on certain parts of the frame, which have a high probability of containing the Intel Extreme Masters logo only. After having retrieved the desired frames, we can then select the proper endpoint pairs by assuming that the replay's length would not be more than 20 seconds.

One problem in this stage is that we can only determine the logo-bounded videos, which might contain many other replays from either the same perspective or multiple perspectives. As the transitions in these videos are smooth, we propose to use the results generated from the section 3.2 to refine the replay detection result.

3.2 Multi-stream Synchronization

With the result obtained from the section 3.1.2, we extract frames of these logo-bounded videos with their original fps, which is 59 for the commentator stream. We apply the same process to the players' videos with their corresponding fps. The problem now is to find in the indexed database the image which is a near duplication of the query one. Therefore, the players frames are then encoded by using [13], added to the database, and then indexed for searching with Elastic Search engine [2]. Instead of searching the whole database for the nearly duplicate images, we determine the match and shape (roundness) of the logo-bounded videos based on provided metadata and their time in the commentator stream. Thereby, we can both narrow down the search space and locate the start frame of each replay.

3.3 Refine Replay Detection Result

As stated in section 3.1.2, for a logo-bounded video with multi-player view, the problem could be solved easily using multi-stream synchronization output. However, it would be harder for the one with single-player view only as we now need to detect abrupt scene changes in the video to split it into proper replays. By using the searched rank list output from the section 3.2, for two consecutive frames, we choose the most similar frame from the indexed database, compute the absolute time difference in the synchronised player stream, and then set a threshold. Thereby, we could reuse the result of the previous step to detect abrupt scene transitions between two replays in the stream with single-player view and refine replay detection output based on the source time of similar frames in player streams set.

4 RESULTS AND ANALYSIS

It can be seen from the graph shown in figure 1 that the best run of our team (DCU-Computing) manages to gain a precision of 73.17% and a recall of 68.18%, yielding F1-score of 70.59%, which is the second-highest score among the submissions for evaluation at Jaccard threshold of 0.5. For evaluation at Jaccard threshold of 0.75 shown in figure 2, all of our scores decrease significantly (approximately 23.53%), except for the average overlapping score

[2]https://github.com/EdjoLabs/image-match

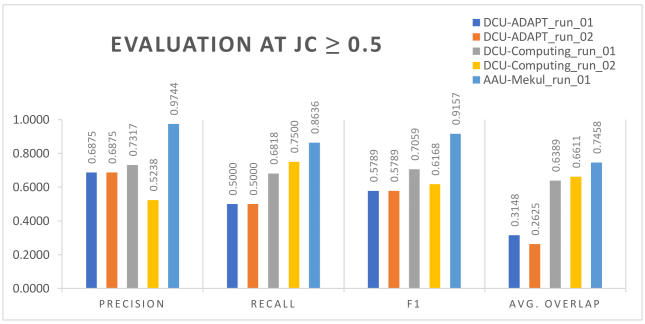

Figure 1: Evaluation of all team's runs at Jaccard-Index threshold = 0.5

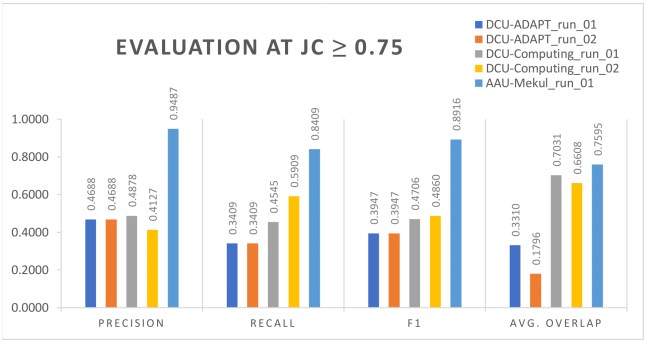

Figure 2: Evaluation of all team's runs at Jaccard-Index threshold = 0.75

of multi-stream synchronization. The evaluation shows that our algorithm works well to retrieve correctly more than 50%, but less than 75% of a replay. Our result is lower than the best approach by the AAU-Mekul team by around 20.98% [9]. As we detect replays based on two logo-transitional endpoint pairs, our algorithm cannot handle the case that misses one endpoint.

There are many situations that we mistakenly split the replays of a player due to wrong image signature matching results. For instance, in case that the player stands still and shoots, then moves a little bit and finally returns to the original point and shoots again, most of the frames will all find its source at one time point, while a sudden player's movement yields a different time point, which causes our algorithm to consider it as a scene change and split replays. Or when the player's view is affected by smoke/blind grenade, the features of the white scene are not distinctive enough to determine its source. These cases all lead to the wrong replay splits that reduces our detected replays' length, which decreases our score. Our approach works perfectly with logo-bounded streams containing only one replay. For other cases, we can generate acceptable results which might have few incorrect splits.

ACKNOWLEDGMENTS

This publication has emanated from research supported by Science Foundation Ireland under grant numbers SFI/12/RC2289 and 13/RC/2106.

REFERENCES

[1] R. Arandjelović and A. Zisserman. 2012. Three things everyone should know to improve object retrieval. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*. 2911–2918. https://doi.org/10.1109/CVPR.2012.6248018

[2] Nathan Edge. 2013. Evolution of the Gaming Experience: Live Video Streaming and the Emergence of a New Web Community. *Elon Journal of Undergraduate Research in Communications* 4 (2013), 2153–5760.

[3] Ali Javed, Aun Irtaza, Yasmeen Khaliq, Hafiz Malik, and Muhammad Tariq Mahmood. 2019. Replay and key-events detection for sports video summarization using confined elliptical local ternary patterns and extreme learning machine. *Applied Intelligence* 49, 8 (01 Aug 2019), 2899–2917. https://doi.org/10.1007/s10489-019-01410-x

[4] Mark Johnson and Jamie Woodcock. 2018. The impacts of live streaming and Twitch.tv on the video game industry. *Media Culture Society* (12 2018). https://doi.org/10.1177/0163443718818363

[5] Bai Liang, Song-Yang Lao, Alan Smeaton, Noel O'Connor, David Sadlier, and David Sinclair. 2009. Semantic Analysis of Field Sports Video using a Petri-Net of Audio-Visual Concepts. *Liang, Bai and Lao, Songyang and Smeaton, Alan F. and O'Connor, Noel E. and Sadlier, David and Sinclair, David (2009) Semantic analysis of field sports video using a petri-net of audio-visual concepts. The Computer Journal, 52 (7). pp. 808-823. ISSN 0010-4620* 52 (10 2009). https://doi.org/10.1093/comjnl/bxn058

[6] S. Lloyd. 1982. Least squares quantization in PCM. *IEEE Transactions on Information Theory* 28, 2 (March 1982), 129–137. https://doi.org/10.1109/TIT.1982.1056489

[7] David G. Lowe. 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* 60, 2 (01 Nov 2004), 91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

[8] Mathias Lux, Michael Riegler, Duc Tien Dang Nguyen, Marcus Larson, Martin Potthast, and Pål Halvorsen. 2019. GameStory Task at MediaEval 2019. In *Proceedings of MediaEval 2019*.

[9] Kevin Mekul. 2019. Automated Replay Detection and Multi-Stream Synchronization. In *Proceedings of MediaEval 2019*.

[10] Wolfgang Reisig. 1985. *Petri Nets: An Introduction*. Springer-Verlag, Berlin, Heidelberg.

[11] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. 2011. ORB: an efficient alternative to SIFT or SURF. *Proceedings of the IEEE International Conference on Computer Vision*, 2564–2571. https://doi.org/10.1109/ICCV.2011.6126544

[12] Jinjun Wang, Eng Chng, and Changsheng Xu. 2005. Soccer replay detection using scene transition structure analysis. *IEEE International Conference on Acoustics, Speech, and Signal Processing* 2, ii/433 – ii/436 Vol. 2. https://doi.org/10.1109/ICASSP.2005.1415434

[13] Chi Wong, Marshall W. Bern, and David Goldberg. 2002. An image signature for any kind of image. *Proceedings. International Conference on Image Processing* 1 (2002), I–I.