GameStory Task at MediaEval 2019

Mathias Lux,[1] Michael Riegler,[2] Duc-Tien Dang-Nguyen,[4] Johanna Pirker,[5]
Martin Potthast,[6] and Pål Halvorsen[2,3]

[1]Alpen-Adria-Universität Klagenfurt, Austria; [2]SimulaMet, Norway; [3]Oslo Metropolitan University, Norway;
[4]University of Bergen, Norway; [5]Graz University of Technology, Austria; [6]Leipzig University, Germany

mlux@itec.aau.at, michael@simula.no, ductien.dangnguyen@uib.no, johanna.pirker@tugraz.at,
martin.potthast@uni-leipzig.de, paalh@simula.no

ABSTRACT

Game video streams are watched by millions, so that, meanwhile, one can make a living from broadcasting and commenting video games, whereas some have become professional e-sports athletes. E-sports leagues and tournaments have emerged worldwide, where players compete in controlled environments, streaming the matches online, and allowing the audience to discuss and criticize the gameplay. In the GameStory task, held for the second time at MediaEval, we foster research into this exciting domain. Our focus is on analyzing and summarizing video game streams. With the help of *ZNIPE.tv*, we compiled a high-quality dataset of a *Counter-Strike: Global Offensive* tournament alongside ground truth labels for two analysis tasks, forming a basis for summarization.

1 INTRODUCTION

The e-sports industry has grown significantly in the past decade. Exemplified with one of the most successful games, League of Legends (LoL), in 2012, the number of concurrent online viewers in an LoL championship for a single event exceeded one million.[1] From 2016 to 2019, the peak viewer counts rose from 28.26 (2016) to 106.27 (2017) to 205.11 million (2018) viewers.[2] The video game streaming industry competes with traditional sports events for top viewership counts.[3]

E-sports has been compared to traditional sports a lot. Like in traditional sports, leisure activities, like playing soccer for fun, may lead to professional training and organized competitions for athletes. In addition, Freeman and Wohn [5] posit that e-sports is defined by the spectatorship and the governing bodies like the ESL Gaming Network.[4] Hamilton et al. [6] found that streaming games focuses on social engagement and community building, which contrasts traditional sports, where the focus lies on the highest levels of play. E-sports appears to be in-between game streaming and traditional sports. Though it depends on a participatory community, especially streams with a large amount of viewers struggle to maintain meaningful social engagement. Seo and Jung [11] see e-sports at the heart of consumer communities with consumers also being players interacting with the game beyond the game interface.

In-game streaming and e-sports, a lot of content is created. Besides videos, platforms like *Twitch.tv* or *YouTube* allow spectators to interact with the players, influencing their gameplay. Altogether, the data streams that can be collected for an individual player include video and audio, commentaries, game data and statistics, interaction traces, viewer-to-viewer communication. The level of detail of the data available, as well as its heterogeneity, render video game streams a challenging subject to multimedia research, allowing for a manifold of research questions.

New research fields like game analytics [2, 3] have emerged, investigating the highly interactive and narrative nature of video games [1]. Relating to summarization, while a lot of work has been done on videos and multi-modal summarization [7, 10, 12, 16], video games have hardly been investigated so far. At MediaEval 2018, GameStory was organized for the first time [8, 9]. Here, participants were given the multi-player plus commentary streams for an e-sports match of *Counter-Strike: Global Offensive* (CS:GO),[5] a game with a solid population of players and viewers, and the task was to generate an entertaining summary.

At GameStory 2019 we built on top of the 2018 dataset by adding a particular subproblem of e-sports match summarization. Since the data are comprised of video streams from multiple perspectives on a given match, as well as a commentators' stream, we ask participants to identify critical moments by finding replays in the commentators' stream and by aligning them with the source video clip in the players' streams. Then, as a second task, following our 2018 summarization task, participants are asked to create a multi-modal summary that provides a captivating story of the match's progress. The latter task is open-ended in the sense that there is no ground truth, but participants can be creative and identify critical moments themselves, whereas the former task may guide participants in solving the latter.

2 BACKGROUND AND DATASET

CS:GO is a first-person shooter (FPS) game and, as an e-sports game, it has very strict rules of- play. Two teams, the terrorists and the counter-terrorists, with five players each, compete in a virtual 3D world, called map. Matches consist of several rounds and players only re-spawn in between rounds. Depending on the success of players and teams, players get awarded virtual money. With that money, players can outfit their avatars with weapons, ammo, tools, and armor in-between rounds.[6]

[1]https://web.archive.org/web/20130608053015/http://www.riotgames.com/articles/20130509/549/league-legends-season-two-championship; all URLs in this paper have been last accessed on July 26, 2019, and been archived at the Internet Archive.
[2]https://escharts.com/tournaments/lol
[3]https://onlinebusiness.syr.edu/blog/esports-to-compete-with-traditional-sports/
[4]https://www.eslgaming.com/
[5]https://csgo-stats.com/
[6]http://www.tobyscs.com/csgo-economy-guide/

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

Our dataset was recorded at the CS:GO Intel Extreme Masters (IEM) ESL tournament in Katowice 2018. According to the ESL rules, a typical CS:GO match is decided in a best-of-30 fashion. After 15 rounds, the teams switch sides, i.e., terrorists become counter-terrorists and vice versa. If two teams end up with a draw after 30 rounds, the teams play overtime to determine the winner. Strategies typically span over multiple rounds, including different sets of constraints, e.g., teams can afford to buy what they need, teams have to save money, etc. In general, a winning condition is to eliminate all opponents. In a DE_Map round, as we have them in our dataset, the goal for the terrorist team is to plant a bomb at one of some specific locations and protect it from being defused (until the end of the round), while the counter terrorist team has to defuse the bomb. In these maps, a bomb going off or the prevention of that event is an additional winning condition.

The data consists of twelve video streams along with metadata. Ten files give the view of the players with the in-game audio streams. One file gives the commentator stream, where a professional cutter selected the parts of the player views to be shown as well as videos from the audience, the teams' players, and the commentator provides the spoken content mixed with recordings from the game and audience cheering. The last one shows the map from above with icons indicating the position of players. A metadata file indicates the start and end of games and the content in the commentators' stream. JSON files, one for each match in the dataset, capture player activity and events in addition to the raw video. Events range from kills, deaths, starts, and ends of rounds to what the players bought at the beginning of rounds and when a bomb was set or a grenade was thrown and went off. The data covers three days of the tournament and is split into training (two days) and test (one day) sets. A ground truth for where to find replays in the commentators' stream is given for the first day of the training dataset. Additionally, for all the video streams, we provide synchronization data as the actual video is off up to 40 seconds from the time stamps given in the metadata.

3 TASKS AND EVALUATION

Compared to sports summaries [4, 13–15], video games are not focused on a small number of attention points like, for instance, the ball and the two goals on a soccer field. Rather, video games comprise multiple views and concurrent events within a well-defined game world (map), which can change in-between games, but stays the same in one single game. Typically, game statistics only convey active and obvious events, but miss those with semantics on a tactical level, including fake tactics, intentional misses, intentional risk-taking (e.g., a player avatar's death or re-spawn), an the like.

With that in mind, we defined two tasks for GameStory 2019:

(1) Find all replays in the commentators' streams and locate the source clips in the respective player streams.
(2) Create a short and captivating summary with a maximum length of five minutes of a single match.

The evaluation of the first task is based on the overlap of the found clips with the ground truth. To determine if a given replay has been successfully found, we employ the Jaccard index in terms of video frames:

$$J(A, B) = \frac{A \cap B}{A \cup B},$$

where A denotes the set of consecutive frames identified as a replay clip, and B the set of consecutive frames of the actual replay from the ground truth.

We consider a replay to be successfully found if $J(A, B) > t$ for threshold t, using two thresholds $t_1 > 0.5$ and $t_2 > 0.75$. Using the Jaccard index, we calculate precision, recall, and the F1 score for the set of replays to be identified. Should a given replay be identified more than once, only the clip with the highest Jaccard index is counted. In a second step, we determine the goodness of the match between true positive replay clips and original player streams. Again for we use the Jaccard index to determine the overlap of the source segment from the ground truth with the segment found in the run. To quantify the degree of overlap for all found replays we average for all found replay segments.

The evaluation of the second task is based on a jury of experts, including CS:GO players and game researchers. The jury members watch the summaries individually and independent of each other with the tasks of summarizing and arguing both strong and weak points of each submission, as well as rating them on a 5-point Likert scale (strongly agree to strongly disagree) concerning the following statements:

(1) The submission gives a summary of the match at hand.
(2) The submission is entertaining.
(3) The submission provides the flow and peak of a good story.
(4) The submission provided an innovative way to present a summary of a CS:GO match.
(5) A summary like this submission can be applied to games other than CS:GO.

4 DISCUSSION AND OUTLOOK

GameStory is at the forefront of research on data-driven analysis of video games and game streaming, a domain that has hardly been addressed within multimedia computer science to date. Here, huge amounts of content and data are generated by millions of players and viewers daily, by both amateurs and professional creators and producers. Their highly interactive nature makes the outcome of games mostly unpredictable. Research in this area is still in its infancy but has the potential for a high social impact. Players and viewers are often young, and games have become an important part of youth culture, sometimes having a strong and long-lasting influence on people's lives. With GameStory, we seek to explore this exciting new direction of research. In the coming years, our goal is to diversify and grow the tasks with a combination of objective, quantitative analysis tasks (e.g., finding kill streaks, synchronizing streams, or identifying relations between consecutive wins and economy in the game streams) and freestyle synthesis tasks, where the analysis technology can be readily employed.

ACKNOWLEDGMENTS

We thank Michael Wutti for providing the scores and the sync points, Natascha Rauscher, Shivi Vats, and Simon Bernard for creating the ground truth, and Sabrina Kletz for the helpful discussion of the evaluation metrics.

REFERENCES

[1] Jason Bryce and Jason Rutter. 2002. Spectacle of the deathmatch: Character and narrative in first-person shooters. *ScreenPlay: Cinema/videogames/interfaces* (2002), 66–80.

[2] Anders Drachen, Magy Seif El-Nasr, and Alessandro Canossa. 2013. Game analytics–the basics. In *Game analytics*. Springer, 13–40.

[3] Magy Seif El-Nasr, Anders Drachen, and Alessandro Canossa. 2013. *Game Analytics: Maximizing the Value of Player Data*. Springer Publishing Company, Incorporated.

[4] Farhad Bayat Farshad Bayat, Mohammad Shahram Moin. 2014. Goal Detection in Soccer Video: Role-Based Events Detection Approach. *International Journal of Electrical and Computer Engineering (IJECE)* 4, 6 (Dec 2014), 979–988.

[5] Guo Freeman and Donghee Yvette Wohn. 2017. eSports as an emerging research context at CHI: Diverse perspectives on definitions. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 1601–1608.

[6] William A Hamilton, Oliver Garretson, and Andruid Kerne. 2014. Streaming on twitch: fostering participatory communities of play within live mixed media. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 1315–1324.

[7] Zhuo Lei, Ke Sun, Qian Zhang, and Guoping Qiu. 2016. User Video Summarization Based on Joint Visual and Semantic Affinity Graph. In *Proceedings of the 2016 ACM workshop on Vision and Language Integration Meets Multimedia Fusion (iV&L-MM)*. 45–52. https://doi.org/10.1145/2983563.2983568

[8] Mathias Lux, Pål Halvorsen, Duc-Tien Dang-Nguyen, Håkon Stensland, Manoj Kesavulu, Martin Potthast, and Michael Riegler. 2019. Summarizing E-sports matches and tournaments: the example of counter-strike: global offensive. In *Proceedings of the 11th ACM Workshop on Immersive Mixed and Virtual Environment Systems*. ACM, 13–18. https://doi.org/10.1145/3304113.3326116

[9] Mathias Lux, Michael Riegler, Duc-Tien Dang-Nguyen, Marcus Larson, Martin Potthast, and Pål Halvorsen. 2018. GameStory Task at MediaEval 2018. In *Working Notes Proceedings of the MediaEval 2018 Workshop*. http://ceur-ws.org/Vol-2283/MediaEval_18_paper_8.pdf

[10] Jia-Yu Pan, Hyungjeong Yang, and C. Faloutsos. 2004. MMSS: multi-modal story-oriented video summarization. In *Proc. of IEEE International Conference on Data Mining (ICDM)*. 491–494. https://doi.org/10.1109/ICDM.2004.10033

[11] Yuri Seo and Sang-Uk Jung. 2016. Beyond solitary play in computer games: The social practices of eSports. *Journal of Consumer Culture* 16, 3 (2016), 635–655.

[12] Rajiv Ratn Shah, Anwar Dilawar Shaikh, Yi Yu, Wenjing Geng, Roger Zimmermann, and Gangshan Wu. 2015. Eventbuilder: Real-time multimedia event summarization by visualizing social media. In *Proceedings of the 23rd ACM international conference on Multimedia (ACM MM)*. ACM, 185–188.

[13] H. C. Shih. 2017. A Survey on Content-aware Video Analysis for Sports. *IEEE Transactions on Circuits and Systems for Video Technology* PP, 99 (2017), 1–1. https://doi.org/10.1109/TCSVT.2017.2655624

[14] Yih-Ming Su and Shu-Jiun Liang. 2008. A novel highlight event decision approach for baseball videos. In *Proceedings of IEEE International Symposium on Consumer Electronics*. 1–2. https://doi.org/10.1109/ISCE.2008.4559549

[15] Min Xu, Ling-Yu Duan, Changsheng Xu, M. Kankanhalli, and Qi Tian. 2003. Event detection in basketball video using multiple modalities. In *Proceedings of the joint International Conference on Information, Communications and Signal Processing and Pacific Rim Conference on Multimedia*, Vol. 3. 1526–1530 vol.3. https://doi.org/10.1109/ICICS.2003.1292722

[16] Y. Yuan, T. Mei, P. Cui, and W. Zhu. 2017. Video Summarization by Learning Deep Side Semantic Embedding. *IEEE Transactions on Circuits and Systems for Video Technology* PP, 99 (2017), 1–1. https://doi.org/10.1109/TCSVT.2017.2771247