TGIF!: Selecting the most healing TNT by optical flow

Changeun Yang[1], Pujana Paliyawan[1], Ruck Thawonmas[2], Tomohiro Harada[2]
[1]Graduate School of Information Science and Engineering
[2]College of Information Science and Engineering
Ritsumeikan University
Kusatsu, Shiga, Japan

Abstract

In this paper, we propose TNT Gained In optical Flows (TGIF), a new algorithm to select the most entertaining TNT explosions in an Angry Birds-like action puzzle game. We assume that spectators like a high amount of object movement distributed equally over both space and time and that such movement entertains spectators. We, hence, first divide a game video into multiple frames and estimate the optical flows of each frame. With these optical flows, we compute a total displacement and two kinds of Shannon entropy: spatial entropy and temporal entropy. Spatial TGIF and Temporal TGIF are computed by multiplying the total displacement and the respective entropy. We then predict the best explosion video using these two methods. Our results show that the proposed Spatial TGIF's correct rate is the highest, i.e., 95% which is much higher than 65% by our previous work.

Introduction

Video game live-streaming is a hot trend these days. There are more than millions of spectators watching video game live-streaming. Live-streaming is increasingly grabbing attention from researchers, and some of them are studying on factors leading to entertaining live-streaming or are investigating effects from live-streaming on human emotions.

For example, one work uses an audience participation platform in an Angry-Birds-like game where humorous messages are extracted from a chat window and used for generating game levels corresponding to the messages (Jiang et al. 2018). One of the main groups of Twitch users focuses on entertaining themselves (Smith et al. 2013), and there is a correlation between spectators' emotions and enjoyment (Downs et al. 2013). We aim at finding a way to generate an enjoyable game video for improving the spectators' emotions.

Angry Birds is the targeted game in this study. It is an action puzzle game whose goal is to destroy pigs by shooting birds to them. As with our previous study (Yang et al. 2018), the purpose of this research is to find the best position to place a TNT in a given Angry Birds level, so that players or spectators can experience the most entertaining content. We employ two hypotheses. The first hypothesis is that spectators' interestingness towards an TNT explosion is proportional to a large amount of movement equally distributed over space. The second one is the same but over time. Based on these hypotheses, we propose two respective methods both using optical flows to select TNT explosions that help promote the spectators' emotions.

Related Work

Angry Birds

Two annual competitions named Angry Birds AI Competition (Angry Birds AI Competition; Stephenson et al. 2018) and AIBIRDS Level Generation Competition (AIBIRDS Level Generationi Competition; Stephenson et al. 2018) inspire many studies on Angry Birds. A number of agents playing Angry Birds were developed using various ways such as qualitative reasoning (Waga et al. 2016), Bayesian regression (Tziortziotis et al. 2016), and deep reinforcement learning (Ma et al. 2018). In case of generating levels, a search-based approach initiated the procedural-content-generation field for Angry Birds (Ferreira and Toledo 2014). Stephenson and Renz (2017) proposed a level generation algorithm that guarantees the generated levels are solvable. Jiang et al. focused on health promotion with Angry Birds. They developed an entertaining level generator using funny quotes (2017) and also proposed an audience participation platform for promotion of social well-being (2013), both of which inspire the present work.

Optical Flow Estimation

Optical flows represent motion patterns between two consecutive frames. FlowNet, a convolutional neural network to estimate the optical flows of a video, had promising results of estimating optical flow (Dosovitskiy et al. 2015). After that, FlowNet2.0, the enhanved version of FlowNet, is proposed (Ilg et al. 2017). With stacked several CNNs and one more parallel CNN for detecting small displacements, FlowNet2.0 decreased the estimation error by more than 50%. FlowNet2.0 shows almost the lowest endpoint error in comparison to other methods at the time when it was proposed. As a result, it is used in this work although more recent methods such as LiteFlowNet (Hui et al. 2018) and PWC-net (Sun et al. 2018) are worth examining.

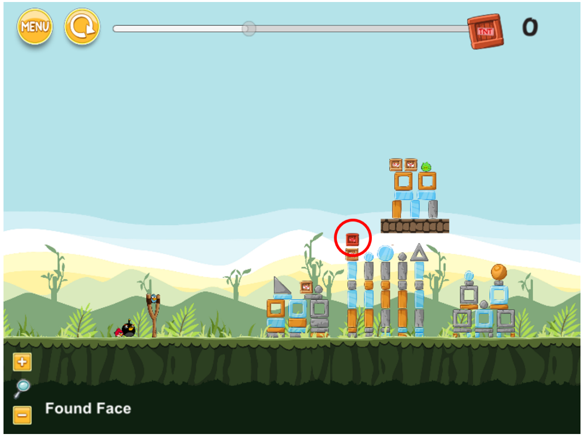


Figure 1: Game Screen of ABT

Event Detection in Game Videos

Our work is related to event detection. Chu and Chou detected events and highlights in broadcast game videos based on a number of features (2015) and built a highlight forecasting model (2017). Optical flow was used as one of the features, but the concept of equal distribution of movement over space or time was not taken into account. Ringer and Nicolaou (2018) used a deep unsupervised model to generate highlight clips using audio, webcam and game screen, but this model can not generate highlight clips only with game screen. Fan-Chiang et al. (2015) proposed a concept of Segments of Interest for live game streaming. With proper feature extracter, it can save bandwidth on live game streaming efficiently.

Approach

Science Birds

In this research, we used Science Birds, a clone game of Angry Birds developed for research purposes (Ferreira and Toledo 2014). Its source code is available in GitHub (Github of lucasnfe). Science Birds contains almost every component of the original Angry Birds. For levels, we used a sample level generator, made available by AIBIRDS Level Generation Competition organizers as a baseline. This generator and its instructions are available in the competition website (AIBIRDS Level Generation Competition).

Dataset

We reused the same dataset as the one in our previous work (Yang et al. 2018). To make this paper self contained, we describe here how the dataset was obtained using a Science Birds implementation with smile interface called Angry Birds with a red TNT (ABT). ABT is a modification of Science Birds embedded with an emotion recognition tool. In this version of game, there is a special TNT that explodes only when the player smiles. With ABT, we aimed at improving spectators' emotions by showing them Angry Birds videos generated from ABT. For data collection, we first took videos of 20 levels with five different TNT explosions in each level. We then conducted a user study with these 100 videos and obtained the average interestingness score of each video. After that, we trained a random forest (Liaw and Wiener 2002) regression algorithm to predict which one, among five videos of a given level, is the most entertaining video. This algorithm will be the baseline in our experiment.

However, here we additionally applied a noninferiority test (Walker and Nowacki 2011) to the five videos of each level to find which explosions are still good, even though they are not the best. Namely, if a video of interest is non-inferior to the best one, that video will be also considered as a correct answer. The boundary of this test was set to 0.8, which is a common value for this kind of test, and four different confidence levels were used: 80%, 85%, 95%, and 97.5%.

TNT Gained In optical Flows

TNT Gained In optical Flows (TGIF) is an algorithm for ranking TNT explosions based on their optical flows. First, frames are extracted from each video at a frame rate of 30 fps. FFmpeg (Homepage of FFmpeg), A free framework for handling video and audio, is used for extraction. From these frames, optical flows are estimated using FlowNet2.0. Spatial TGIF and Temporal TGIF are then computed. Spatial TGIF considers movement displacements over space in the video screen; Temporal TGIF, instead, adds all the movement displacements in each frame and considers the displacements of the video over time.

The Spatial TGIF and Temporal TGIF values were computed for all of the 100 TNT explosions (videos) in the dataset. For each of the 20 levels, its five explosions were ranked according to either the Spatial TGIF value or the Temporal TGIF value, and the ranking results were then compared with the ranking by the average interestingness score obtained in the aforementioned user study. Figure 2 show examples of an aftereffect of an explosion and an optical flow per frame.

Spatial TGIF Once optical flows are computed, each explosion consists of several frames, and each frame consists of two-dimensional vectors, corresponding to respective pixels in a frame of interest. Equation (1) defines an explosion (denoted as S) lasting for n frames. In this equation, s_i represents the ith frame consisting of two-dimensional vectors (x_{ij}, y_{ij}), where j is the index to the jth pixel whose maximum value is m.

$$S = \{s_i | 1 \le i \le n\} = \{(x_{ij}, y_{ij}) | 1 \le i \le n, 1 \le j \le m\} \tag{1}$$

The magnitude of each pixel's vector is first calculated for each frame. For pixel j, the sum of the magnitudes for all frames, w_j, is then taken. These w_j spatially form total displacement field (Fig. 3). Total displacement TD is defined as the sum of all w_j, from which Shannon entropy $H^{spatial}$ is calculated as follows:

$$w_j = \sum_{i=1}^{n} \sqrt{x_{ij}^2 + y_{ij}^2}\ ,\ TD = \sum_{j=1}^{m} w_j$$

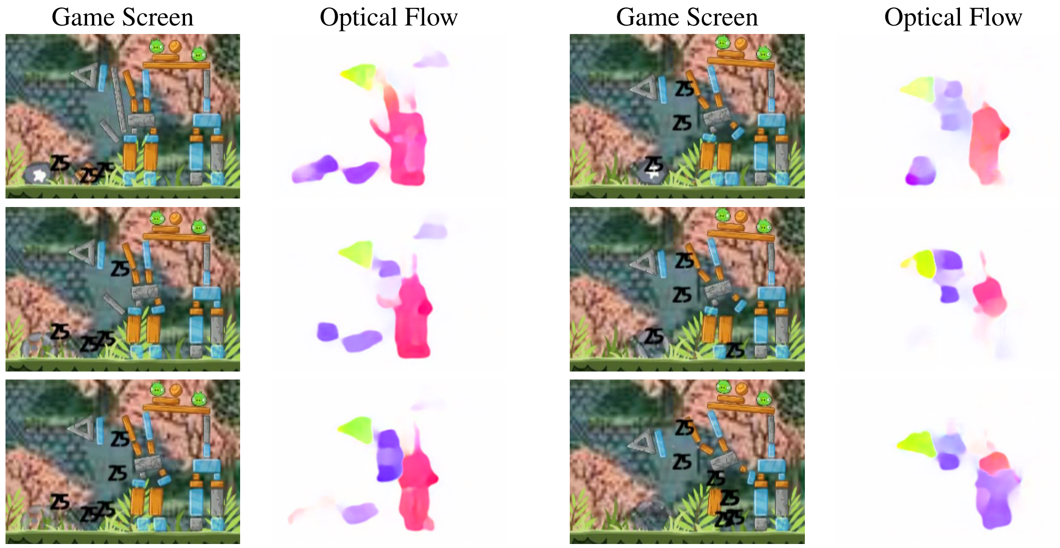
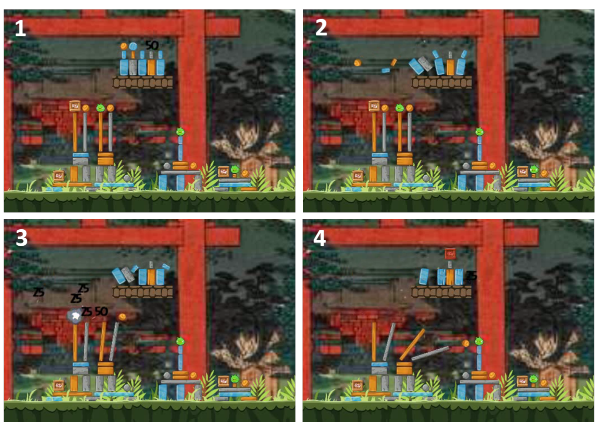


Figure 2: Six frames of cropped game screens and their optical flows from the top left to bottom left and then top right to bottom right.



(a) Four frames of game screen

(b) The total displacement field

Figure 3: An example of total displacement field. (a) is four sampled frames of cropped game screens and (b) is the total displacement field of explosion (a). A brighter pixel means that there were many or fast movements in that pixel.

$$p_j = \frac{w_j}{TD}\ ,\ \ H^{spatial} = -\sum_{j=1}^{m} p_j \log_2 p_j$$

TD indicates the dynamic of aftereffects, i.e., the faster and the more pixels move, the more total displacement is gained. On the other hand, $H^{spatial}$ indicates the spatial size of the aftereffects, i.e., the bigger the collapsing area is, the larger value the entropy is. Using TD per pixel and normalized $H^{spatial}$, Spatial TGIF of S is given as follows:

$$TGIF^{spatial}(S) = \frac{H^{spatial}}{log_2 m} * \frac{TD}{m}$$

Temporal TGIF In Temporal TGIF, the entropy is calculated in a different way. Here, $H^{temporal}$ indicates the consistency of movement over time and is calculated as follows:

$$w'_i = \sum_{j=1}^{m} \sqrt{{x_i}_j^2 + {y_i}_j^2}\ ,\ \ p'_i = \frac{w'_i}{TD}$$

$$H^{temporal} = -\sum_{i=1}^{n} p'_i \log_2 p'_i$$

Temporal TGIF is then calculated in the same fashion as follows:

$$TGIF^{temporal}(S) = \frac{H^{temporal}}{log_2 n} * \frac{TD}{n}$$

Filter

There exists irrelevant motion such as ABT bar rising or birds jumping before they are placed in the slingshot. Since these movements are not relevant to the total displacement, they must be removed from consideration. This is done by cropping all the in-screen user-interfaces and assigning zero vectors to all the pixels in the portion where birds reside.

In addition, even though FlowNet2.0 is good at suppressing noises, we found that noises still exist in flow field as shown in Fig. 3. And this issue might affect video evaluation. To avoid this, we applied a filter that sets the magnitude of any pixel that has the value below a given threshold to zero. We examined two types of filter: small-size filter whose threshold is the average of the magnitudes of all pixels in all frames of the 100 videos and mid-size filter whose threshold is the minimum value of the maximum magnitudes in each frame of the 100 videos. The former and the latter are denoted as *sml* and *mid*, respectively; the one where no filter is applied is denoted as *non*.

Discounted Cumulative Gain

DCG (Discounted Cumulative Gain) is originally used to measure effectiveness of web search engine algorithms. It uses a graded relevance scale of individuals to measures the usefulness of individuals. In order to use DCG, There should be two assumptions:

Assumption 1 *Highly relevant documents are more useful when appearing earlier in a search engine result list.*

Assumption 2 *Highly relevant documents are more useful than marginally relevant documents, which are in turn more useful than non-relevant documents.*

Relevance score corresponds to interestingness in our case. Then, highly relevant documents become videos with high interestingness. And search engine can be changed as an order of estimated rank. Finally, We used nDCG to compare methods with these new two assumptions :

Assumption 3 *Highly ranked videos are more useful when appearing earlier in an order of estimated rank.*

Assumption 4 *Highly ranked videos are more useful than marginally ranked videos, which are in turn more useful than low rank videos.*

We used this method to measure and compare the ranking quality of algorithms for selecting the most entertaining video. We computed the normalized DCG with top 5 (nDCG@5) and top 3 (nDCG@3) in each level and obtained the average of these scores for each method.

Experimental Results

All the methods were evaluated using two metrics: correct rate and discounted cumulative gain. The correct rate indicates how many correct videos (the most entertaining videos) are selected by a method of interest, where the rate is adjusted by the noninferiority test as mentioned above. Alo, ranking quality is measured by normalized DCG. As a baseline, the regression algorithm in our previous work (Yang et al. 2018) was used.

Correct rate

Figure 4 compares the correct rate of each method, where ”No noninferiority” shows the results with no adjustment in selecting the most entertaining video for each level while the others are adjusted by the noninferiority test with different significance levels γ. It can be seen that all the methods show better results as the significance level is relaxed. Both Spatial TGIF and Temporal TGIF outperform the baseline regressor. When the mid-filter is applied, the correct rate drops in both Spatial and Temporal TGIFs. In low significance level ($\gamma \leq 0.95$), Spatial TGIF shows higher correct rates than Temporal TGIF, while the score of Temporal TGIF is slightly equal to or higher than that of Spatial TGIF with the significance level of 0.975 or no noninferiority. Spatial TGIF with the sml filter has the highest result, reaching 0.95 with $\gamma = 0.8$.

nDCG

The average of normalized discounted cumulative gains for all levels is calculated for each method (the baseline and the six combinations of the proposed ones). Figure 5 shows the results. Again both Spatial and Temporal TGIFs outperform the baseline. As with the correct rate, in both Spatial and Temporal TGIFs, sml-filter is better than the ones with non-filter or mid-filter. In both nDCG@5 and nDCG@3, Spatial TGIF with sml-filter is of the highest performance with the values of 0.986 and 0.969, respectively.

Discussions

We adapted optical flows and proposed two methods to find the most entertaining videos of an TNT explosion: Spatial TGIF and Temporal TGIF. Spatial TGIF focuses on the entropy based on the amount of optical flows among all the pixels, and Temporal TGIF focuses on the entropy based on the the amount of optical flows among all the frames. Before each TGIF is calculated, in order not to be disturbed by noises from optical flow estimation, all vector magnitudes less than a given threshold are set to zero.

The two methods were evaluated with two metrics. In the correct-rate comparison, we compared their best-video-selection performances. And in the nDCG-measure comparison, the overall ranking quality of each method was compared. Both proposed methods outperformed the baseline in terms of both metrics. In particular, Spatial TGIF with sml-filter was the best. It reached up to 95% of the correct rate as the range of entertaining becomes wider, i.e., when more videos in a given level are considered interesting.

Future Work

Estimating the most entertaining video is not yet perfect. If the significance level of noninferiority test (Walker and Nowacki 2011) is high, the correct rate stays around 70%. Our future work is to aim for 90%, by combining both TGIFs or additionally introducing other features. We noticed that one video has the highest interestingness rating in its level because of a domino effect with few objects flying and the area of aftereffects not that large. Optical flow estimation

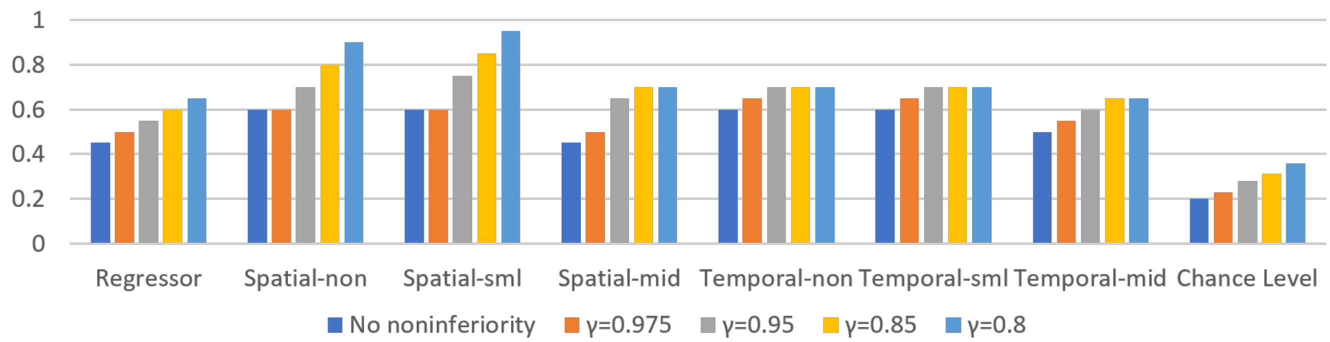


Figure 4: Correct rate comparison between the baseline, Spatial TGIF and Temporal TGIF in different filters and significance levels (γ). Each bar in the most right group shows the chance level for each setting.

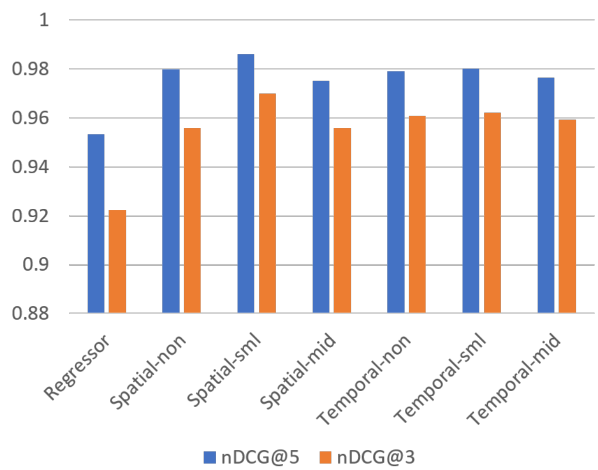


Figure 5: Average of nDCG@5 and nDCG@3 for each of the seven methods

only focuses on the size in space or the length in time of aftereffects, but not creative movement due to an explosion. In this particular case, use of optical flows is not sufficient to predict the interestingness, which requires future work.

The proposed TGIF methods can be applied in various ways such as highlight detection or procedural play generation (Thawonmas and Harada 2017). In highlight detection, TGIF predicts entertaining parts for viewers, from which highlight videos can be made. In procedural play generation, we can, for example, train an Angry Birds playing agent with a reward of TGIF. Since our previous work (Yang et al. 2018) showed that a set of the most entertaining explosion videos improved spectators' emotions, we expect that the trained agent can also improve their emotions trough its gameplay.

Acknowledgment

This work was supported in part by JSPS KAKENHI Grant Numbers JP15H02939.

References

Jiang, Y., Paliyawan, P., Thawonmas R. and Harada T. 2018. An audience participation angry birds platform for social well-being. the 19th annual European GAME-ON Conference on Simulation and AI in Computer Games(GAME-ON'2018). September 18-20, Scotland, United Kingdom.

Smith, T., Obrist, M. and Wright P. 2013. Live-streaming changes the (video) game. ACM Press, 131.

Downs, J., Vetere, F., Howard S. and Loughnan S. 2013. Measuring audience experience in social videogaming. Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration. November 25-29. Adelaide, Australia.

Yang, C., Paliyawan, P., Harada T. and Thawonmas R. 2018. Blow Up Depression with In-Game TNTs. 2018 IEEE 7th Global Conference on Consumer Electronics. October 9-12. Nara, Japan.

Angry Birds AI Competition. `http://aibirds.org/`.

Stephenson, M., Renz, J., Ge X. and Zhang, P. 2018. The 2017 AIBIRDS Competition. arXiv:1803.05156 arXiv:1803.05156v1.

AIBIRDS Level Generatioin Competition. `https://aibirds.org/other-events/level-generation-competition.html`.

Stephenson, M., Renz, J., Ge, X., Ferreira, L-N., Togelius, J. and Zhang, P. 2018. The 2017 AIBIRDS Level Generation Competition. In: IEEE Transactions on Games. doi: 10.1109/TG.2018.2854896.

Waga, P-A., Zawidzki, M. and Lechowski, T. 2016. Qualitative physics in Angry Birds. In: IEEE Transactions on Computational Intelligence and AI in Games, vol. 8, no. 2, 152165.

Tziortziotis, N., Papagiannis, G. and Blekas, K. 2016. A bayesian ensemble regression framework on the Angry Birds game. In: IEEE Transactions on Computational Intelligence and AI in Games, vol. 8, no. 2, 104115.

Ma, Y., Takano, Y., Zhang, E., Harada, T. and Thawonmas, R. 2018. Playing Angry Birds with a Neural Network and

Tree Search. 2018 Angry Birds AI Symposium. Stockholm, Sweden.

Ferreira, L. and Toledo, C. 2014. A search-based approach for generating angry birds levels. In: 2014 IEEE Conference on Computational Intelligence and Games. 18.

Stephenson, M. and Renz, J. 2017. Generating varied, stable and solvable levels for Angry Birds style physics games. In: 2017 IEEE Conference on Computational Intelligence and Games, 288-295.

Jiang, Y., Harada T. and Thawonmas R. 2017. Procedural Generation of Angry Birds Fun Levels using Pattern-Struct and Preset-Model. In: 2017 IEEE Conference on Computational Intelligence and Games, 154-161. New York, USA..

Dosovitskiy, A., Fischer, P., Ilg, E., Husser, P., Hazrbas, C., Golkov, V., van der Smagt, P., Cremers, D. and Brox, T. 2015. Flownet: Learning optical flow with convolutional networks. In: IEEE International Conference on Computer Vision.

Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy A. and Brox T. 2017. FlowNet 2.0: Evolution of optical flow estimation with deep networks. In: IEEE Conference on Computer Vision and Pattern Recognition.

Hui, T-W., Tang, X. and Loy C-C.: Liteflownet 2018. A lightweight convolutional neural network for optical flow estimation. In: IEEE Conference on Computer Vision and Pattern Recognition.

Sun, D., Yang, X. and Liu, M-Y., Kautz J. 2018. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In: IEEE Conference on Computer Vision and Pattern Recognition.

Chu, W-T. and Chou, Y-C. 2015. Event detection and highlight detection of broadcasted game videos. In: Proceedings of ACM Workshop on Computational Models of Social Interactions, Human-Computer-Media Communication. Brisbane, Australia.

Chu, W-T. and Chou, Y-C. 2017. On broadcasted game video analysis: event detection, highlight detection, and highlight forecast. Multimedia Tools and Applications 76, 7, 97359758. DOI:http://dx.doi.org/10.1007/ s11042-016-3577-x.

Ringer, C., and Nicolaou, M-A. 2018. Deep Unsupervised Multi-View Detection of Video Game Stream Highlights. In: Proceedings of the 13th International Conference on the Foundations of Digital Games, 15.

Fan-Chiang, T-Y., Hong H-J. and Hsu, C-H. 2015. Segment-of-interest driven live game streaming: saving bandwidth without degrading experience. In: Proceedings of international workshop on network and systems support for games, 16.

Github of lucasnfe. `http://github.com/lucasnfe/Science-Birds`.

Yang, C., Jiang, Y., Paliyawan, P., Harada, T. and Thawonmas, R. 2018. Smile with Angry Birds: Two Smile-Interface Implementations. In: Proceedings of 2018 NICOGRAPH International, 80. Tainan, Taiwan.

Liaw, A. and Wiener M. 2002. Classification and regression by randomForest. R News 2, 18-22.

Walker, E. and Nowacki, A-S. 2011. Understanding equivalence and noninferiority testing. Journal of General Internal Medicine, 26, 192-196.

FFmpeg Homepage. `http://www.ffmpeg.org/`.

Thawonmas, R. and Harada, T. 2017. AI for Game Spectators: Rise of PPG. AAAI 2017 Workshop on What's next for AI in games. 1032-1033. San Francisco, USA.