PokerFace: Game Players Themselves Are Truly Memorable

Jong-Woon Yoo, Ki-Woong Park, Sung Kyu Park, Hyunchul Seok, and Kyu Ho Park
Computer Engineering Research Laboratory
KAIST
Daejeon 305-701, Republic of Korea
{jwyoo|woongbak|skpark|hcseok|kpark}@core.kaist.ac.kr

***Abstract*—Video games support instant replays so that game players can see previously occurred events (e.g., goals in sports games) again from a different point of view and permanently store such exciting moments and feelings as pleasant memories. However, players are likely to have different feelings and enjoyment depending on the *player context*, such as who they play with, their looks, or conversation during gaming even for the same game. Therefore, instant replays without such player context are not enough for users to remind the enjoyable moments that could not happen again, and thus precious memories will be forgotten. This paper presents *PokerFace*, a new video game console that provides advanced instant replays containing the player context. PokerFace detects memorable situations during gaming and generates advanced instant replays by combining both the game replay and the user's reaction and synchronizes them in time so that the players can see both their own looks and those of other players at the time when the exciting moment take place. Our system provides game players with better memories by storing advanced instant replays into the permanent storage or web blogs. We have implemented a prototype of PokerFace and conducted user studies. From the evaluation results, we conclude that PokerFace can have the competitiveness in game consoles market by giving more enjoyment and better memories for users.**

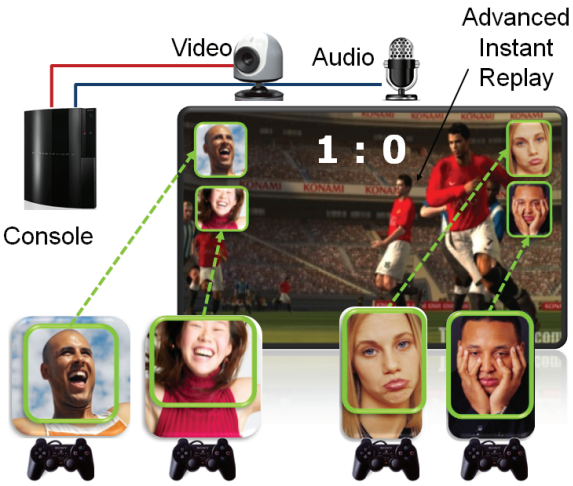

Fig. 1. An illustrative example showing the concept of PokerFace. PokerFace records the player's looks and sounds and combines them with instant replay to provide the game players with better memory.

I. Introduction

Video games have created a huge wave of popular interest because of the breakthroughs in video game consoles, such as Xbox™, PlayStation™, or Wii™, and attractive game content. While the purpose of games is getting more diverse [1], such as education [2], health [3], or training [4], entertainment is still the primary goal of gaming.

Video games, especially sports games, usually support instant replays for users so that they can see previously occurred events (e.g., goals in sports games) again from a different point of view and permanently store such exciting moments and feelings as pleasant memories. Even for the same game, however, players are likely to have different feelings and enjoyment depending mainly on the *player context*, such as who they play with, their looks, conversation, reaction, the interaction in the game, and so on. Instant replays without such player context are not enough for users to remind the enjoyable moments that could not happen again, and thus precious memories will be forgotten.

This paper presents a new video game console, *PokerFace*, which provides advanced instant replays containing the player context. Fig. 1 shows an illustrative example where four players are enjoying a video soccer game and the left-side team just scores a goal. PokerFace automatically detects this memorable situation by using not only game-originated events which are statically defined in the game (e.g., goals), but also user-originated events. Here the user-originated events include the sound of cheering of players who score a goal or looks of groaning players of the opposite team. To recognize the user-originated events, PokerFace hardware includes a camera and a microphone, as shown in Fig. 1. Note that the user-originated events can occur independently of the game-originated events: even though any special event happens in gaming, there could be memorable moments for the players due to their conversation, interaction, or other user-specific situations. When PokerFace detects an event, it generates an advanced instant replay that includes both traditional game replay and the reaction (e.g., looks or sounds) of the players. The traditional game replay shows only the replay of game characters as programmed. Therefore, the traditional game replay lacks user reactions like ceremony or cheering, which are likely to be unique and irreproducible. On the other hand, the advanced instant replay of PokerFace combines both the game replay and the user's reaction and synchronizes them in

978-1-4244-7180-5/10/$26.00 ©2010 IEEE

time so that the players can see both their own looks and those of other players at the time when the exciting moment take place. PokerFace provides game players with better memories by storing advanced instant replays into the permanent storage or web blogs.

The reminder of this paper is organized as follows. Section III describes our PokerFace in detail. Section IV presents how PokerFace detects the game-originated and user-originated events. Section V and IV give the implementation and the evaluation result via user studies. Section II briefly surveys related works. Finally, Section VII gives concluding remarks and directions for future research.

II. Related Work

An important trend in developing video games is the use of multi-sensory (visual, auditory, and haptic) feedback interfaces for the game users as well as the development of various input devices. One popular feedback method is the use of vibration in game controllers like DualShock [5]. It includes two vibration motors that can provide the vibration feedback based on the action of the game. However, all feedbacks are already coded in the games. For the general use of the vibration feedback on pervasive games, Dongju Chi *et al.* [6] developed a sound-specific vibration interface, which vibrates according to the sounds of games. In order to avoid the situation of vibrating too often, they use the sound matching instead of frequency detection by using digital signal processing. In addition, for customizing user's own vibrating patterns, they provide the vibration pattern editor.

In addition to the vibration, sound and visual contents can also be used as sensory feedback to the users. Martin Faust et. al. [7] use the AirBats, which is the input device of the Airkanoid [8] game. This game is a remake of Taito's classic Arkanoid game [9], in which a game's paddle is moved by user's control to prevent a ball from falling to the playing field by bouncing the ball. In the Airkanoid, they used webcams to track the AirBats and the paddle moves following the movement of AirBats to re-bound the ball. When the ball is hit, a sound plays an important role for feeling the realism with the vibration made by the AirBats. By using two sensory feedback of both sound and vibration, it gives the virtual realism to the users.

The researches of feedback interfaces have developed to increase the reality and the immersion in the games. With these multi-sensory feedback, users can feel the virtual reality and have more fun with video games. However, these interfaces only provide the enjoyment in an instant. Although the previous works can improve the player's performance and feeling, they cannot provide the memorial scenes such as players' conversation, interaction, and reactions during the game.

III. PokerFace System

This section describes the system architecture of PokerFace in detail. Fig. 2 shows the block diagram of the architecture of the proposed video game console. It consists of five major components as follows:

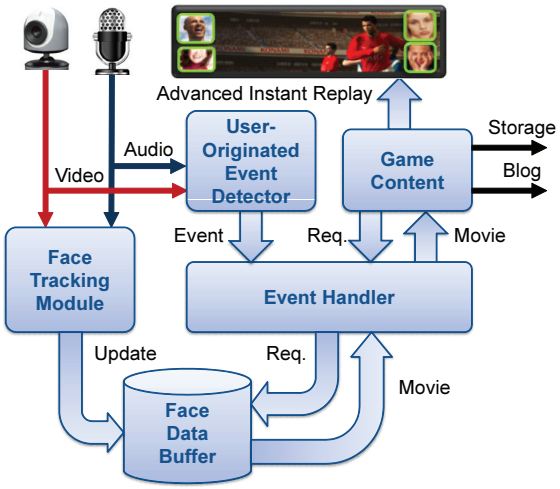

Fig. 2. The architecture of the PokerFace system.

- **Face Tracking Module**: It tracks the faces of players and stores data into the face data buffer. Because the replay movies of the players' reaction should be quickly shown in Picture-in-Picture (PiP) manner, showing whole bodies of the player on a small pop-up screen makes the players hard to see their faces that are likely to contain the best expressions of their feelings. Therefore, our current system records only the faces of the players.
- **Face Data Buffer**: It is a simple First-In First-Out (FIFO) buffer that stores the face movie data of each player.
- **User-Originated Event Detector**: It continuously monitors the audio and video stream from the microphone and camera and detects 'meaningful' reactions of the players. Section IV discusses the event detector in detail.
- **Game Content**: The game content in PokerFace requests user data (i.e., recorded face movies) to the event handler when an event takes place. The request message contains some information, such as controller ID, the length of movies, and so on. Then, it generates an advanced instant replay that consists of both game replay and the users' face movies. The game replay movie and the user's face movie are synchronized in time so that players can see both their own looks and those of other players at the time when the exciting moment take place. The game content will support permanent storing or blogging for the users' memories.
- **Event Handler**: According to the request from the game content software, it retrieves the required amount of movie stored in the face data buffer and provide it to the game content software.

IV. Event Detection

Event detection in PorkerFace is particularly crucial for understanding semantic concepts of players' interest for more advanced user experience and memory-reminder purposes. Detecting meaningful events from playing game is a challenging multimedia understanding problem. Various approaches have

978-1-4244-7180-5/10/$26.00 ©2010 IEEE

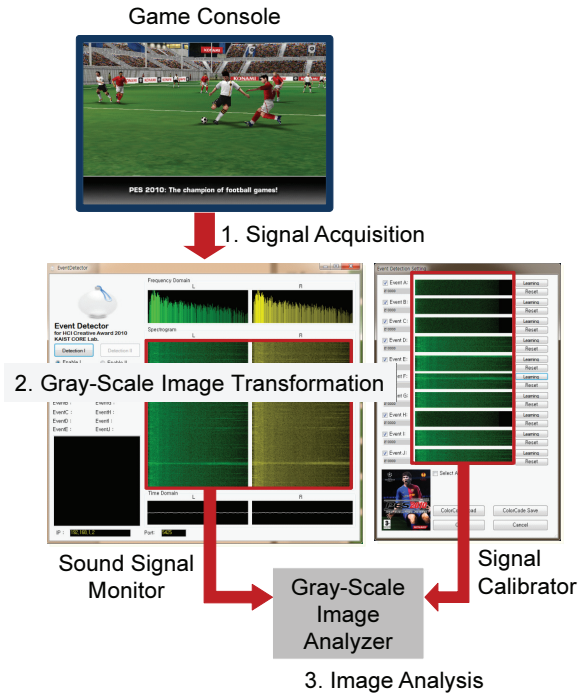

Fig. 3. The event detector module of PokerFace.

been proposed for context detection and semantic annotation of news video [10] and home video [11].

In an attempt to detect meaningful events from playing game, an event detection module that exploits game audio signal has been designed and implemented. The technical concepts used in our event detection procedure is presented in [12]. It involves three main steps: 1) the development of audio signal transformation subroutines, which are aimed at transforming raw audio signal from playing game into vertically expanded image streams, 2) signal fingerprint extraction, the sound signal where the meaningful game-context, e.g., goals scored, near-misses, fouls in soccer game, is characterized by capturing the corresponding image, and 3) the development of image analysis subroutine for investigating images that are transformed for the event detection.

A. Internal Architecture of Event Detector

Fig. 3 shows the overall architecture of the event detection module. The three major components are described as follows:

- **Signal Calibrator**: Characterized signal fingerprints corresponding meaningful game events are determined by the parametric registration. The parameters determine the signal characteristics in terms of the signal strength and the corresponding frequency. These are converted to gray-scale image. Each pixel of the image is mapped to a certain frequency region and brightness of each pixel is set by signal strength. Although this technique can be directly used in end products, this is more useful for game developers to tune the system parameters and events using the visualized audio signals. In this work, we configure the internal parameters by registering the gray-scale images of meaningful game event through pilot evaluation.
- **Sound Signal Monitor**: This component periodically collects sound signals periodically (every 10 ms) from a sound port of game console as shown in Fig. 3. It then transforms the raw sound signal from playing game into vertically expanded image streams on the basis of gray-scale coding and transmits the transformed gray-scale images to gray-scale code analyzer.
- **Gray-Scale Code Analyzer**: It analyzes the received transformed images and triggers a certain signal if a meaningful event which is registered by signal calibrator. Event Detection module then automatically sends a notification signal to camera module to record user's face.

B. Event Detection Procedure

The event detection module can be used in detecting meaningful events from playing game. It detects the events by transforming the captured sound signal into gray-scale image. The transformation involves a series of operations: firstly it extracts the gray-scale image that consists of multiple pixels. Each pixel indicates the signal strength of a certain frequency region of the captured sound signal; it then produces a stream of images, where the horizontal axis is mapped to frequency value and the vertical stream is accumulated over time; and, finally, it facilitates the analysis of the image streams for the game event detection. The event detection procedure involves three main steps as follows:

- **Signal Acquisition**: Sound signal is collected from the audio output port of game console devices with 128K sampling rate. The collected sound signal is simultaneously transformed in the next step.
- **Gray-Scale Image Transformation**: This process transforms the captured sound signal (with a resolution of 128Kbps) into vertically expanded image streams in forms of the gray-scale image. The transformation is based on the correspondence between the vertically expanded image and the current time. The transformed images are then transmitted to the image analyzer.
- **Gray-Scale Image Analysis**: Gray-scale images are analyzed with purpose of investigating the transformed images for detecting meaningful events from playing game. For every received transformed image, the signal characteristic that corresponds to the time of the image acquisition is matched with the registered signal fingerprints. If a meaningful event which is registered by signal calibrator, this module then automatically triggers a notification signal to camera module to record user's face.

V. Implementation

We have implemented a prototype of our PokerFace system (Fig. 4). For the simplicity of implementation, we have exploited four Logitech webcams which are capable of tracking the face of a user. With better implementation techniques, we

978-1-4244-7180-5/10/$26.00 ©2010 IEEE

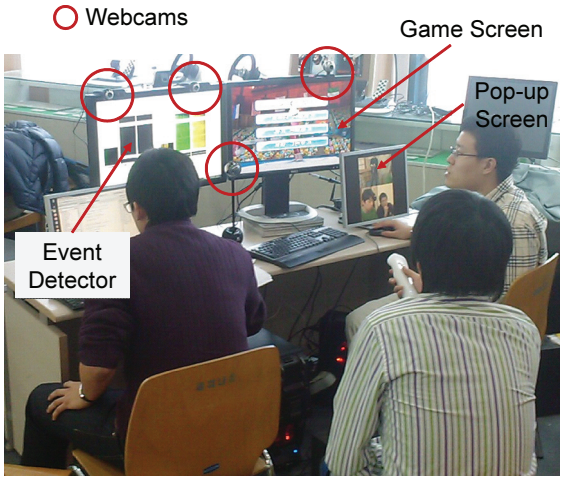

Fig. 4. A prototype implementation of PokerFace

believe that only one webcam would be sufficient for tracking all users as today's image processing technologies support simultaneous tracking of multiple faces [13][14]. Currently, our PokerFace prototype supports up to four players.

To display the reactions of game players, e.g., their faces, recorded when an event is detected, we simply use a secondary monitor, as shown in Fig. 4. However, this secondary monitor would not be needed by the final commercialized version of PokerFace as the pop-up screens can be overlapped on the main screen in a Picture-in-Picture (PiP) [15] manner. The position of each pop-up screen may affect game playing: if the pop-up screen suddenly appears at the center of the main screen and hides some important scenes, the game players will feel angry. Therefore, the location of the pop-up screens should be carefully determined to prevent game players from feeling annoyed and, go further, to improve enjoyment. Even though using a secondary monitor makes the implementation easy, our user study result shows that it negatively affects on game playing. Section VI discusses this issue in detail.

Our current system uses only the sound generated by game software for event detection. We analyze the sound of each game content in off line and register meaningful sound patterns as memorable events. The next steps of our future research include improvement of the event detection mechanism so that it can monitor users' reactions using vision technologies and generate user-originated events during gaming.

VI. Evaluation

This section evaluates the PokerFace system via user study.

A. Evaluation Setup

We have implemented a prototype of PokerFace using two commercial video game consoles (PlayStation™ and Wii™), desktop computers, and four webcams for evaluation. In user study, we choose Winning Eleven 2010 [16] for PlayStation™ for male users and Vancouver Winter Olympic [17] for Wii™ which is easier to control for female users as game titles. In order to operate the PokerFace system, we analyze the sound of the games and extract meaningful contexts like goals scored, near-misses, or fouls, in advance using the event detection system described in Section IV. When a predefined event occurs, PokerFace shows advanced instant replays containing the player context for about five seconds around the event. After a game, we ask users who play the game to evaluate the concept of PokerFace based on their experiences.

We surveyed 24 users (18 males and 6 females) in their 20's and 30's with seven questions listed in Table I. These questions are carefully chosen based on four appraisal standards for game: game ability, technical skill, competitiveness, and profitability except for management. They can show whether the PokerFace system is good or not.

B. Result

This section describes the user evaluation results from the seven questions listed in Table I. The score of each question is from 1 point (very poor) to 5 point (very good). Fig. 5 shows an average score of each question according to the characteristic of users which is divided into 4 groups: game-unfriendly male, game-friendly male, game-unfriendly female, and game-friendly female. A user who plays any video game more than once a week is classified as a game-friendly user.

The scores of question 1 and question 2 are over 4 points. It means that most users agree the PokerFace system adds more enjoyment to existing games by showing interesting player contexts and they believe that the concept of PokerFace can be easily applied to various kinds of games. One interesting finding here is that even though PokerFace increases enjoyment, many users pointed out that it disturbs to concentrate on game playing. Thus, the score of question 3 is about 3 point. The main reason that the users gave low points on the question 3 is that, as previously mentioned in Section V, we used a secondary monitor for a pop-up screen, and it makes users change their point of view to see their reaction. We believe that this problem can be solved by having an effort to consider its replay timing, use of Picture-in-Picture (PiP) technology, and the position of pop-up screens. According to the game characteristic, the timing and location have to be changed. As other comments, one pointed out that it would be better to apply more effects when editing the replay video. In the competitiveness and profitability area, many people expect that this PokerFace system can have marketability in a game console market.

TABLE I
Questions for User Study

Game Ability
Q1. Does PokerFace give more enjoyment to games?
Q2. Is it possible for PokerFace to apply various kinds of games?
Q3. Does PokerFace increase the concentration on games?
Technical Skill
Q4. How do you rate the detection technique?
Q5. How do you rate video editing technique?
Competitiveness/Profitability
Q6. Does PokerFace increase sociality in multi-user game?
Q7. Does PokerFace have marketability in a game-console market?

978-1-4244-7180-5/10/$26.00 ©2010 IEEE

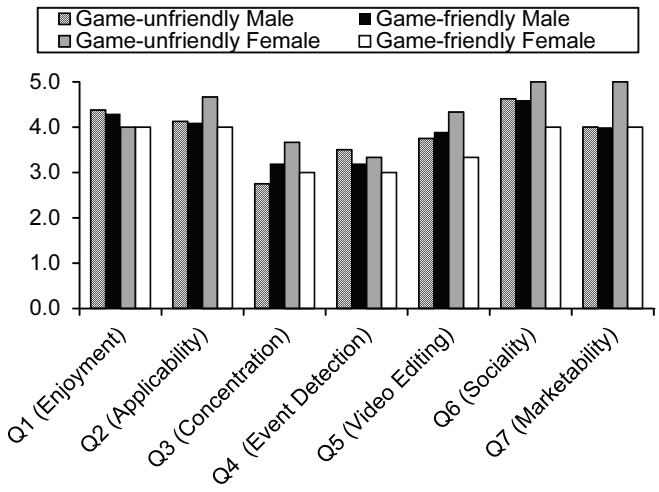

Fig. 5. User study result.

From these evaluation results, we conclude that the PokerFace system can have the competitiveness in game consoles market by giving more enjoyment and better memories for users. However, it still lacks in many things like concentration and techniques. In order to succeed in game console markets, we have an effort to improve the PokerFace system as the further works.

VII. Conclusion

This paper has presented *PokerFace*, a new video game console which provides advanced instant replays containing the player context. PokerFace detects memorable situations during gaming and generates advanced instant replays by combining both the game replay and the user's reaction and synchronizes them in time. PokerFace improves the enjoyment of games by allowing players to see both their own looks and those of other players at the time when the exciting moment take place. By storing advanced instant replays into the permanent storage or web blogs, PokerFace provides game players with better memories. We conducted a user study with 24 participants. From the results, we conclude that PokerFace can have the competitiveness in game consoles market by giving more enjoyment and better memories for game players.

Our future research includes the improvement of the event detection module, employment of PiP technology, and the investigation on the effect of the location of pop-up screens on the concentration of game players. In our current implementation, the event detection module exploits only game audio signals, resulting in the decrease of the detection ratio of the user's memorable events. By using the reaction of the users, e.g. sound and faces, PokerFace can help the users remind memorable moments which can not be represented by the game audio signal.

Acknowledgment

This research is supported by Ministry of Culture, Sports and Tourism (MCST) and Korea Creative Content Agency (KOCCA) in the Culture Technology (CT) Research & Development Program 2010.

References

[1] D. R. Michael and S. L. Chen, *Serious Games: Games That Educate, Train, and Inform*. Muska & Lipman/Premier-Trade, 2005.

[2] S. Alers and E. I. Barakova, "Multi-Agent Platform for Development of Educational Games for Children with Autism," in *Proceedings of the International IEEE Consumer Electronics Society's Games Innovations Conference*, ser. ICE-GIC '09, 2009, pp. 47–53.

[3] D. Thompson, T. Baranowski, R. Buday, J. Baranowski, V. Thompson, R. Jago, and M. J. Griffith, "Serious Video Games for Health: How Behavioral Science Guided the Development of a Serious Video Game," *Simulation and Gaming*, vol. 41, pp. 587–606, 2010.

[4] R. E. Chatham, "Games for Training," *Communications of the ACM – Creating a Science of Games*, vol. 50, pp. 36–43, July 2007.

[5] Sony, "DualShock 3 Wireless Controller," http://en.wikipedia.org/wiki/DualShock, 2007.

[6] D. Chi, D. Cho, S. Oh, K. Jun, Y. You, H. Lee, and M. Sung, "Sound-Specific Vibration Interface using Digital Signal Processing," in *Proceedings of the International Conference on Computer Science and Software Engineering*, vol. 4, 2008, pp. 114–117.

[7] M. Faust and Y.-H. Yoo, "Haptic Feedback in Pervasive Games," in *3rd International Workshop on Pervasive Gaming Applications*, ser. PerGames '06, 2006.

[8] D. Cermark-Sassenrath, M. Faust, and H. Rosch, "AirKanoid – Virtual Presentation vs. Physical Proximity in Mixed Reality Entertainment Applications," in *2nd International Workshop on Pervasive Gaming Applications*, ser. PerGames '05, 2005, pp. 2–10.

[9] Taito, "Arkanoid," http://www.taito.com/csm/title/2008/arkanoid_ds/, 1986.

[10] Y. Ke, R. Sukthankar, and M. Hebert, "Efficient Visual Event Detection using Volumetric Features," in *Proceedings of the 10th IEEE International Conference on Computer Vision*, ser. ICCV '05, vol. 1, 2005, pp. 166–173.

[11] D. Xu and S.-F. Chang, "Visual Event Recognition in News Video using Kernel Methods with Multi-Level Temporal Alignment," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, ser. CVPR '07, 2007, pp. 1–8.

[12] K.-W. Park, W. Hwang, and K. H. Park, "FalconEye: Data Center Status Extraction using Vision-Transformation Techniques," in *IEEE 34th Annual Computer Software and Applications Conference Workshops*, ser. CloudApp '10, 2010, pp. 388–392.

[13] F. J. Huang and T. Chen, "Tracking of Multiple Faces for Human-Computer Interfaces and Virtual Environments," in *Proceedings of the IEEE International Conference on Multimedia and Expo*, ser. ICME '00, vol. 3, 2000, pp. 1563–1566.

[14] Z. Shaik and V. Asari, "A Robust Method for Multiple Face Tracking Using Kalman Filter," in *36th IEEE Applied Imagery Pattern Recognition Workshop*, ser. AIPR '07, 2007, pp. 125–130.

[15] M. Masuda, K. Okamoto, H. Nishijima, K. Ono, H. Otsubo, and S. Otsuka, "Picture in Picture System with a Digital Memory for VCRS," *IEEE Transactions on Consumer Electronics*, vol. CE-33, no. 3, pp. 230–239, 1987.

[16] Konami, "Pro Evolution Soccer (PES) – Winning Eleven 2010," http://www.konami.com/games/pes/, 2009.

[17] Sega, "Mario & Sonic at the Olympic Winter Games," http://sega.jp/corp/release/2009/0213/home.shtml, 2009.

978-1-4244-7180-5/10/$26.00 ©2010 IEEE