Initial Perceptions of a Casual Game to Crowdsource Facial Expressions in the Wild

Chek Tien Tan
chek@gamesstudio.org

Hemanta Sapkota
hemanta.sapkota@student.uts.edu.au

Daniel Rosser
daniel.j.rosser@gmail.com

Yusuf Pisan
yusuf.pisan@gamesstudio.org

Games Studio, Faculty of Engineering and IT, University of Technology, Sydney

ABSTRACT

The performance of affective computing systems often depend on the quality of the image databases they are trained on. However, creating good quality training databases is a laborious activity. In this paper, we evaluate BeFaced, a tile matching casual tablet game that enables massive crowdsourcing of facial expressions for the purpose of advancing facial expression analysis. The core aspect of BeFaced is game quality, as increased enjoyment and engagement translates to an increased quantity of varied facial expressions obtained. Hence a pilot user study was performed on 18 university students whereby observational and interview data were obtained during playtests. We found that most users enjoyed the game and were intrigued by the novelty in interacting with the facial expression gameplay mechanic, but also uncovered problems with feedback provision and the dynamic difficulty adjustment mechanism. These findings hence provide invaluable insights for the other researchers/practitioners working on similar crowdsourcing *games with a purpose*, as well as for the development of BeFaced.

Keywords

Games with a purpose; crowdsourcing; facial expression analysis

Categories and Subject Descriptors

H.5.m [**Information Interfaces and Presentation**]: Miscellaneous; I.2.1 [**Applications and Expert Systems**]: Games.

1. INTRODUCTION AND RELATED WORK

Facial expression analysis is a major topic in the domain of affective computing. It generally involves recognizing facial expressions like happy or sad, and then making inferences about what they mean. Facial expression recognition is still an open problem but most modern algorithms depend on learning from a set of exemplar face images. To obtain good recognition performance, the face database needs to be large with images having high variability. This includes intra-person variability like pose and illumination as well as inter-person variability in terms of gender, age and other facial features. Unfortunately, collecting such databases is costly and time-consuming.

In computer vision research, public datasets are also central to the advancement of the state-of-the-art because they provide common benchmarks to allow researchers to compare different algorithms objectively. Popular datasets in the area of facial expression analysis include the Cohn-Kanade database [8], or CK+, and the MMI database [11], amongst a plethora of others [13]. Many prominent works in facial expressions analysis have used these databases for training and testing purposes (for example [1]). Though widely used, these databases are greatly limited in the number of unique participants, and are mostly confined to laboratory settings. This problem can be attributed to the method of collection, which is often rather manual and time-consuming. Dataset collection is also a rather one-off activity, where extensions to the corpuses depend very much on the creators' plans. For example CK+, an extension to the original Cohn-Kanade (CK) database [6], was released after almost 10 years.

In a recent work [9], crowdsourcing [2] was used to generate what is known as the Forbes dataset. The researchers acquired arbitrary online viewers to watch media over the internet and recorded their facial expressions. They managed to collect 3268 videos in under two months and showed that the variability was higher than traditional datasets. It demonstrated much potential in using crowdsourcing methods for facial videos, but suffer from two issues. Firstly, their approach requires participants to be willing to watch commercial videos, something that many might dread doing unless there is enough incentive or they are convinced by the nobleness of the deed. The researchers might have been able overcome this issue by advertising on the highly visible Forbes website. Secondly, the expressions obtained greatly depends on the videos shown, which in their case were of a funny nature to solicit joyful expressions. To increase the variability of expressions requires conscious effort to either create new video content or evaluate existing video content, which is laborious. When using existing videos, copyright issues might be a hassle as well. Extensibility and reproducibility of their work are hence difficult. BeFaced hence aims to alleviate these shortcomings by utilizing popular gameplay to enable intrinsic motivation, as well as use an extensible design approach to continuously solicit a growing number of different expressions. Moreover, it allows the option of simply sending facial feature locations instead of the actual facial images.

Using games for crowdsourcing has enjoyed tremendous success in other domains. One of the most well-received applications was

the game FoldIt where players solved the problem of deciphering the accurate protein model of an AIDs-causing virus [7]. They demonstrated that a properly crafted crowdsourcing game provided enough intrinsic motivation for gamers to solve a biological problem that stumped scientists for fifteen years, in just ten days. BeFaced hence aims to utilize a similar concept to help advance the state-of-the-art in facial expression analysis.

Based on the above motivations, BeFaced [12] was developed to enable massive crowdsourcing of facial expressions. It is a tablet game with a core gameplay mechanic that is based on a tile matching mechanic common in many popular casual games. For example Bejeweled[1] is an immensely popular puzzle game based on this mechanic, which has been downloaded over 150 million times. In BeFaced, an alternative version of the tile matching gameplay mechanic was created that included facial expressions as the primary player input and feedback interface. The aim is to use a popular gameplay mechanic to obtain a large database of natural and varied facial expressions in the wild. A major advantage is also the ability to "request" the player for any type of expressions depending on the tiles designed in the levels.

In order to enable massive crowdsourcing, the primary focus needs to be on gameplay first. The more enjoyment and engagement the game provides, the more a person plays, and more people would eventually play, which leads to more varied facial expressions. It is thereby important to have several design iterations before BeFaced is released for free to the public where the effectiveness in collecting facial expressions will eventually be evaluated. Hence the focus of this paper is to perform an initial investigation on the core gameplay mechanic with a first playable prototype of BeFaced.

2. THE BEFACED GAME

The full game design and engine implementation details of BeFaced can be found in a separate paper [12]. This section provides some essential background of BeFaced for the pilot study in this paper.

The core gameplay of BeFaced involves matching facial expression tiles as shown in Figure 1. Whenever three or more tiles are aligned, the player has three seconds to make the expression shown on the tiles in order to destroy them, score points and get new tiles to advance in the game. When the player makes an expression, it is automatically captured and transmitted to an online database, labelled according to the expression shown on the matched tiles as well as any demographic information if available. By default, the system only uploads the tracked feature positions as it is hoped to at least collect anonymous face data that protects the privacy of users who do not wish to share actual face images. Ideally, the player would have indicated explicit permission for uploading face images as well, and the system will also upload them as part of the data record. Depending on the affective algorithm employed, both types of data are believed to be useful to researchers. The ability to collect expression data whilst protecting the privacy of the user is a novel contribution of BeFaced as well.

In order to clear the tiles, the game runs the captured facial expression through a dynamic facial expression classifier. If the probability of the classified expression is greater than a certain threshold, the expression matches and the tiles are cleared. Dynamic difficulty adjustment (DDA) [4] of the threshold is employed as well in order keep the players interested in playing more, and hence providing more examples to our database. Note that the goal of the expression recognition system is different from most computer vision applications. It does not aim to provide accurate recognition, and instead would be willing to reduce the accuracy of recognition in order to keep the players engaged, so as to obtain more varied records in the dataset.

[1]www.bejeweled.com

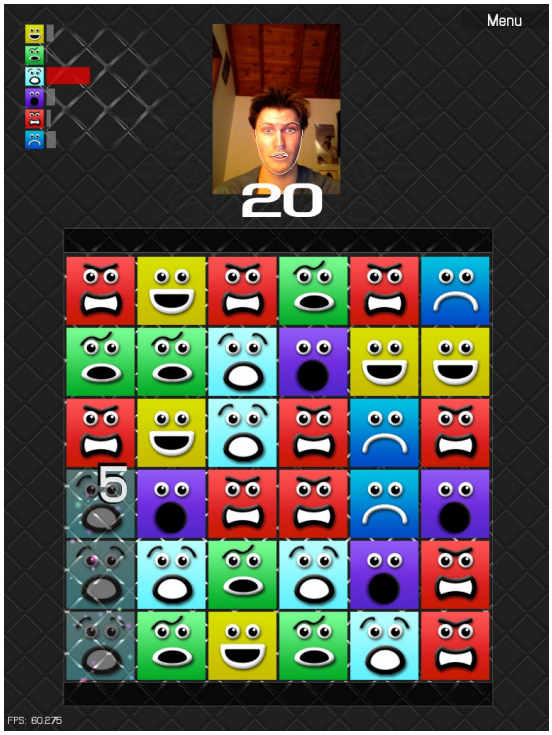

Figure 1: A screenshot of the BeFaced game. The player has aligned three tiles in the bottom left corner and cleared them to score 5 points by successfully making the shocked facial expression on the matched tiles. The facial expression is tracked in real-time and the white splines in the video on top shows the tracked feature points.

In the current first prototype version of BeFaced for the pilot study, there is only one level with six basic expressions on the tiles. This is so that the core gameplay mechanic can be tested first before moving on to further development, which is commonly regarded as good practice in game design [3].

3. METHOD

Individual playtesting was employed as it has been shown to be an effective method to improve game quality in the iterative game design process [10]. Data collection includes recorded gameplay videos of participants, observations made by a researcher during play, and semi-structured post-play interviews common to playtests [5]. The think-aloud protocol, also common in playtests, was not used as it disrupts the facial expressions during play because the facial tracker needs time to re-establish a lock when it loses tracking.

Participants were recruited via undergraduate and postgraduate university mailing lists. 18 male participants (P1-P18) took part in the study and were aged between 19 and 36 (M = 24.2, SD = 4). Most participants were familiar with a tablet device as 12 indicated they

own an iPad, two indicate they own an Android-based tablet and four do not own any tablets. The participants were primarily casual players where two participants indicated they play games for more than 20 hours per week, two participants between 10 to 20 hours, seven participants less than 5 hours per week, and six participants less than 1 hour. One indicated he/she did not play games at all. Also, 13 participants indicated they have played a version of the Bejeweled game whilst five participants indicated they have not.

The experimental procedure consisted of three parts. (1) After signing a consent form, participants filled in a pre-play questionnaire to determine player demographics including age, sex, gaming interests, gaming habits, and the type of tablet they have as well as whether they have played Bejeweled games (resulting in the data as presented above). Participants were briefed on the structure of the experiment before proceeding. (2) They were then given an iPad (third generation) and proceeded to play BeFaced until they felt like stopping. (3) After finishing play, the researcher conducted a semi-structured interview whereby planned question categories included whether they learned the game quickly, whether they felt joy and whether there were anything they disliked. A high definition webcam was used to record video and audio of the play sessions and interviews. Each session took around 15 minutes in total and no compensations were given after the session. Data reduction involved transcribing the post-game interview videos and thereby iteratively recording themes by analyzing each transcript in succession.

4. RESULTS AND DISCUSSION

The average time taken to play the current single-level BeFaced game was 3m22s (SD = 1m18s). The longest play time was 6m02s (P12) whilst the shortest was 25s (P15) which correlates to their gaming habits of 10-20 hours of games per week (P12) and less than 1 hour of games per week (P15) respectively. This provides a brief notion that BeFaced was perceived as a normal game.

The coded transcriptions of the post-game interviews are the primary data used for this analysis and will be augmented with the observational data. Three broad themes emerged from the data, which is described as follows.

4.1 Enjoyment and Novelty

Nearly all participants (17/18) made explicit positive comments on their overall impression on the game. The remaining participant (P1) said that it felt like Bejeweled. Most of them thought the game was interesting and were intrigued by the novelty of the core gameplay mechanic. Descriptive terms like “interesting" (P3, P4, P6, P10, P12, P16, P17 and P18), “awesome" (P5) and “Cool" (P11) were used. For example:

> “I think it’s quite interesting. Like it’s a different approach for gaming because it makes you use your face but it’s also a strategy game. So I think it’s quite interactive." (P10)

> “Um.. it’s pretty new. Something different. When you compared to other ones like bejeweled, candy crush. This one’s like pretty different." (P13)

Some participants also expressed enjoyment during play, using terms like “fun" (P10, P11, P12, P13 and P18) and “interactive" (P7 and P10). For example:

> “I think the game was interesting. It’s fun to play actually." (P12)

The interface also appeared to have contributed to the enjoyment.

> “It’s lovely how you have the feature that show lines come up on your face." (P15)

These responses show that even with a single-level, the core mechanic did manage to establish an obvious psychological draw to users. Firstly, the novelty factor of the facial expression mechanic to clear tiles was attractive enough for players to want to pick it up and play. Secondly, players did express that this core mechanic was fun.

There were also some opinions on optimal play situations. Two participants (P3 and P9) expressed that they would like to play with friends on social networks whilst one (P4) mentioned that it might be embarrassing to play in public.

4.2 Game Challenge

13/18 participants expressed making some of the facial expressions were challenging. For example:

> “Some faces were really hard to make. Especially sad ones." (P5)

> “I was choosing expressions that were easier to make." (P6)

On first look, this might seem like an undesirable experience. However, a well-designed game is not about providing easy gameplay, with many well-known games that are fun but extremely challenging (for example Super Meat Boy[2]). This notion also seemed to be expressed by one participant:

> “Yeah. It’s fun. It’s really fun. Especially when you cannot make those faces if you don’t do the right faces." (P18)

The current single-level BeFaced game immediately presents a typical level with six expressions to be made, but the eventual game would consist of many preceding levels that have a gradually increasing challenge. For example, the first few levels might only involve a single happy expression, and then happy and sad for the next few levels, and so on. Nevertheless, these findings will not be overlooked and the challenge aspects will be re-evaluated when the prior levels are in place, in order to properly balance the challenge progression.

[2] supermeatboy.com

4.3 Perceived Face Tracker Performance

6/18 participants felt that their expressions were not detected properly. For example:

> "Sometimes the facial recognition wouldn't work. I tilted the display to make it work." (P6)

> "Sometimes it doesn't recognise the confused face." (P12)

Via observations from the gameplay videos, it can also be seen for some participants (P12, P13 and P18) that the face tracker often lost track of the face in the midst of play. During this time, they still continued to make expressions possibly resulting in a notion that the face tracker was faulty when there were no feedback from the game. This was due to a combination of large pose and lighting changes on the face, a common issue in many face tracking algorithms. However, this problem was not frequent in other participants. Nevertheless this implies that an obvious feedback mechanism needs to be incorporated that informs and instructs the player how to reposition the iPad in order for the tracker to re-establish a lock on the face.

The other observation was that the DDA system might have reduced the difficulty by too much too quickly. Currently, the dynamic classifier is set to definitely accepting an expression (and adding it as a learning sample to the classifier) after two unsuccessful tries, a design decision that was thought to be able to alleviate overly challenging experiences for some. This might have also attributed to the notion of a faulty face tracker. The DDA system hence needs a much more granular and player-specific adjustment mechanism.

5. CONCLUSION AND FUTURE WORK

From the pilot study performed, we found that the first playable prototype of BeFaced already establishes potential for crowdsourcing facial expressions as participants found it enticing and fun to play. However, several crucial issues have also emerged that will serve as guidelines for other researchers/practitioners working on similar crowdsourcing *games with a purpose* as well as for the next iteration of BeFaced. These guidelines are:

1. providing timely and specific feedback to let the player know how to reposition the iPad.
2. providing a more granular and player-specific DDA system, and
3. designing progressive levels to balance challenge progression.

After the next development iteration, the subsequent evaluation step is to release the BeFaced app onto the Apple App Store coupled with an in-game questionnaire to evaluate player experiences on a larger scale with high external validity. Improving the game quality is of utmost importance as a better game means a higher quality dataset for advancing facial expression analysis research. The release will also serve to investigate its feasibility in collecting a high quality facial expression database and comparing it to current popular datasets.

BeFaced represents a crowdsourcing platform for advancing computer vision research using popular gameplay. As a start we are focusing on facial expressions analysis but there exists vast potential for applications to other areas of computer vision. For example, popular Kinect-based gameplay like those of the Just Dance [3] series can be used in our platform to crowdsource common dance moves to train and test automated dance gesture recognition algorithms.

6. ACKNOWLEDGMENTS

This work was supported by the Centre for Human Centred Technology Design and the School of Software at the Faculty of Engineering & IT at the University of Technology, Sydney.

7. REFERENCES

[1] Bartlett, M. S., Littlewort, G. C., Frank, M. G., Lainscsek, C., Fasel, I. R., and Movellan, J. R. Automatic Recognition of Facial Actions in Spontaneous Expressions. *Journal of Multimedia 1*, 6 (2006), 22–35.

[2] Bernstein, M. Crowdsourcing and Human Computation : Systems , Studies and Platforms. In *Proc. CHI 2011 Ext. Abstracts*, ACM (2011), 53–56.

[3] Fullerton, T. *Game Design Workshop: A Playcentric Approach to Creating Innovative Games*. Elsevier Morgan Kaufmann, 2008.

[4] Hunicke, R. The case for dynamic difficulty adjustment in games. In *Proc. CHI 2005*, ACM Press (2005), 429–433.

[5] Isbister, K., and Schaffer, N. *Game Usability: Advancing the Player Experience*. CRC Press, 2008.

[6] Kanade, T., and Cohn, J. Comprehensive database for facial expression analysis. In *Proc. Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, IEEE Comput. Soc. Press (2000), 46–53.

[7] Khatib, F., DiMaio, F., Cooper, S., Kazmierczyk, M., Gilski, M., Krzywda, S., Zabranska, H., Pichova, I., Thompson, J., Popović, Z., Jaskolski, M., and Baker, D. Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nature Structural Molecular Biology 18* (2011), 8–10.

[8] Lucey, P., Cohn, J., and Kanade, T. The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. In *Proc. IEEE CVPRW*, no. July, IEEE Comput. Soc. Press (2010), 94–101.

[9] McDuff, D., Kaliouby, R. E., and Picard, R. Crowdsourcing facial responses to online videos. *IEEE Trans. on Affective Computing 6*, 1 (2012), 1–14.

[10] Mirza-Babaei, P., and Nacke, L. How does it play better?: exploring user testing and biometric storyboards in games user research. In *Proc. CHI 2013*, no. May, ACM Press (2013).

[11] Pantic, M., Valstar, M., Rademaker, R., and Maat, L. Web-Based Database for Facial Expression Analysis. In *Proc. IEEE ICME 2005*, IEEE Comput. Soc. Press (2005), 317–321.

[12] Tan, C. T., Rosser, D., and Harrold, N. Crowdsourcing facial expressions using popular gameplay. In *Proc. SIGGRAPH Asia 2013 Technical Briefs*, ACM Press (2013).

[13] Zeng, Z., Pantic, M., Roisman, G. I., and Huang, T. S. A survey of affect recognition methods: audio, visual, and spontaneous expressions. *IEEE Trans. PAMI 31*, 1 (2007), 39–58.

[3] http://just-dance-thegame.ubi.com/