
BeFaced: a Casual Game to Crowdsource Facial Expressions in the Wild

Chek Tien Tan

Games Studio, Faculty of
Engineering and IT, University
of Technology, Sydney
chektien.tan@uts.edu.au

Hemanta Sapkota

Games Studio, Faculty of
Engineering and IT, University
of Technology, Sydney
hemanta.sapkota@student.uts.edu.au

Daniel Rosser

Games Studio, Faculty of
Engineering and IT, University
of Technology, Sydney
daniel.j.rosser@gmail.com

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author. Copyright is held by the author/owner(s).

CHI 2014, Apr 26 - May 01 2014, Toronto, ON, Canada
ACM 978-1-4503-2474-8/14/04.
<http://dx.doi.org/10.1145/2559206.2574773>

Abstract

Creating good quality image databases for affective computing systems is key to most computer vision research, but is unfortunately costly and time-consuming. This paper describes BeFaced, a tile matching casual tablet game that enables massive crowdsourcing of facial expressions to advance facial expression analysis. BeFaced uses state-of-the-art facial expression tracking technology with dynamic difficulty adjustment to keep the player engaged and hence obtain a large and varied face dataset. CHI attendees will be able to experience a novel game interface that uses the iPad's front camera to track and capture facial expressions as the primary player input, and also investigate how the game design in general enables massive crowdsourcing in an extensible manner.

Author Keywords

Games with a purpose; crowdsourcing; facial expression analysis; affective computing

ACM Classification Keywords

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

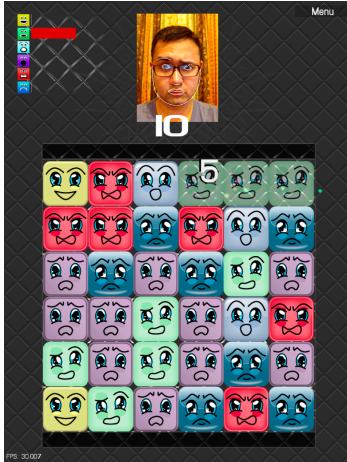


Figure 1: A screenshot of the BeFaced game. The player has aligned three tiles in the top right corner and cleared them to score 5 points by successfully making the disgust facial expression on the matched tiles. His expression is tracked in real-time and the white splines in the video on top shows the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

Introduction

Machine learning algorithms for facial expression analysis systems often depend on having a set of high quality face images as training examples. To train the systems robustly, the database needs to be large and images need to have high variability in terms of facial features, pose and illumination, amongst other variables. 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. Unfortunately, collecting such databases is costly and time consuming.

In facial expression analysis, popular datasets include the Cohn-Kanade database [5], or CK+, and the MMI database [7], amongst a plethora of others [9]. Though widely used, these databases are greatly limited in the number of unique participants, and are mostly confined to laboratory settings. This is mainly due 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 [3], was released after almost 10 years.

In a recent work (Forbes dataset) [6], crowdsourcing [1] was used whereby participants were tasked to watch commercial videos whilst their face images were captured. They demonstrated the potential that crowdsourcing can be effective in collecting a large and varied image dataset. However, their dataset contained only joyful expressions, and extensibility is hard due to the dependence on choosing appropriate video content to solicit expressions. Reproducibility of the approach is also hard, as watching commercial videos is usually dreaded. Their success in

overcoming this might be attributed to the marketing advantage of being advertised on the highly visible Forbes website. 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 to protect the user's privacy.

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 [4]. 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 [8] was developed to enable massive crowdsourcing of facial expressions. It is a tablet game with a core gameplay mechanic based on a tile matching mechanic common in many popular casual games. For example Bejeweled (www.bejeweled.com) is an immensely popular puzzle game based on this mechanic, which has been downloaded over 150 million times. In BeFaced, an alternate 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.

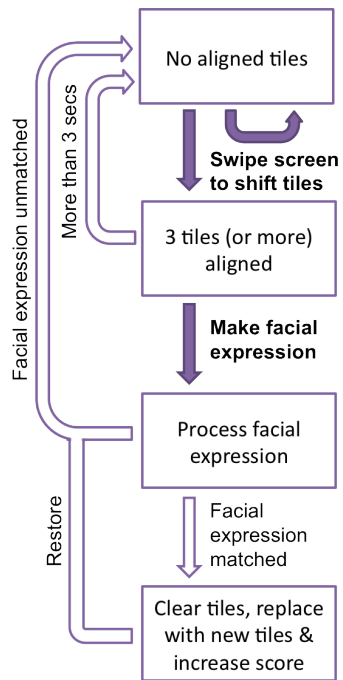


Figure 2: A high level overview of the user and system interactions in BeFaced during gameplay. Boxes depict game states and arrows depict user/system actions. User actions are differentiated by solid arrows and bold labels.

The BeFaced Game

The core gameplay of BeFaced involves matching facial expression tiles as shown in Figure 1. A broad overview of the user interaction process can be seen in Figure 2. Whenever three or more tiles are aligned, the user has to make the expression shown on the tiles in order to advance in the game. BeFaced is currently implemented on the Apple iPad. It uses the iPad's front device camera to capture facial expressions of the player, which is the primary input interface for the game.

As shown in Figure 2, the game starts in a default state where no tiles are aligned. The user first needs to perform swiping actions on the touchscreen in order to get three or more tiles in a line. When three tiles are aligned, the game then provides visual cues on the tiles (highlights tiles and overlays user's video on them) to prompt the player to make the expression shown on the tile. The user then has three seconds to make the expression. When the expression is made, the game then processes it which involves *capturing*, *labeling*, *uploading* and *matching* the expression.

Capturing and *labeling* simply grabs the facial feature points in the current video frame and labels them according to the expression shown on the aligned tiles. *Uploading* involves sending the data to a cloud database. Consent will be explicitly sought before the start of the game, along with some demographic data to tag the expressions. Ideally, the player would have indicated explicit permission for uploading face images as well, and the system will also upload the raw face images as part of the data record. Depending on the affective algorithm employed, we believe both types of data (tracked feature points and face images) would be useful to researchers. The ability to collect expressions data whilst protecting

the privacy of the user is a novel contribution of BeFaced as well.

During *matching*, the captured expression is passed into our dynamic facial expression classifier. If the classified face matches the aligned tiles, the player scores and these tiles are destroyed and replaced, otherwise the game is restored to the original state. Dynamic difficulty adjustment (DDA) [2] is employed in the classifier in order to keep the players interested in playing more, and hence providing more examples to our database. The goal of our expression recognition system can be perceived to be different from most computer vision applications. We are not aiming to provide accurate recognition, but instead are willing to reduce the accuracy of the recognition in order to keep the players engaged, so as to obtain more varied records in our dataset. Further details of the implementation can be found in a prior paper [8].

A pilot study has also been performed which we found that most users enjoyed playing BeFaced and were intrigued by the novelty of interacting with the facial expression gameplay mechanic.

Audience and Relevance

BeFaced is highly relevant to researchers and practitioners at CHI as it crosses the domains of crowdsourcing, game design as well as affective user interfaces. Firstly, it demonstrates a viable crowdsourcing platform that can be used for advancing computer vision algorithms. This might trigger discussions of applications beyond just facial expressions, for example applying the BeFaced concept to crowdsource dance gestures using popular games (e.g., using Kinect dance games). Secondly, it portrays the design of a game-with-a-purpose to enable massive crowdsourcing. The notion of using popularized gameplay

mechanics (i.e., tile matching in our case) to increase engagement in these games might be an area for discussion. Thirdly, it uses a novel facial expression input interface on the iPad to control gameplay. To the best of our knowledge, BeFaced is the first tablet game to have realtime facial expression tracking and classification as part of the core gameplay mechanic. This can provide inspirations for other forms of affective interface design.

Hands-on interaction with BeFaced will therefore provide first-hand experiences with the novel interaction interface as well as the potential for massive crowdsourcing. This can generate valuable insights and triggers for the discussion and exploration of the designs behind BeFaced. Crowdsourcing, game design and novel affective interfaces have been major topics of interest in CHI over the last few years, and BeFaced provides a platform to explore these topics further.

Acknowledgments

This work was supported by the Faculty of Engineering & IT travel grant at the University of Technology, Sydney. We also wish to thank the Centre for Human Centred Technology Design and the School of Software at the Faculty of Engineering & IT for their support in this work. We also thank the anonymous referees for their constructive comments.

References

- [1] Bernstein, M. Crowdsourcing and Human Computation : Systems , Studies and Platforms. In *Proc. CHI 2011 Ext. Abstracts*, ACM (2011), 53–56.
- [2] Hunicke, R., and Chapman, V. AI for Dynamic Difficulty Adjustment in Games. In *Proc. AIIDE 2004*, AAAI Press (2004), 91–96.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] McDuff, D., Kaliouby, R. E., and Picard, R. Crowdsourcing facial responses to online videos. *IEEE Trans. on Affective Computing* 6, 1 (2012), 1–14.
- [7] 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.
- [8] Tan, C. T., Rosser, D., and Harrold, N. Crowdsourcing facial expressions using popular gameplay. In *Proc. SIGGRAPH Asia 2013 Technical Briefs*, ACM Press (2013).
- [9] 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.