Assessment of Fun from the Analysis of Facial Expressions to Support Video Game Design

Luiz C. Vieira and Flávio S. Côrrea da Silva[1]

Abstract. Evaluation is arguably the most important step in the development of a game. Existing methods – involving observation, inquiring and the capture of utilitarian and non-utilitarian data from the game itself or from players through the use of bodily sensors – may tamper the experience or not be able to capture it as it occurs. This paper presents a work in progress of a software system intended to predict fun from the facial expressions of players, based on the detection of prototypic emotions and immersive state. Initial results of the emotion classifier obtained accuracies varying from 68% to 96%, depending on the prototypic emotion.

1 INTRODUCTION

The most important requirement for a game, including digital games, is to be fun. But fun is hard to design because it is a subjective experience essentially made of emotions. Human emotions result from the conscious judgement of events, and are complementary to reason in the decision process by making memory of past experiences more relevant, reinforcing intentions and preparing the body for action [19]. People interpret new experiences in different levels of emotional details depending upon context of use, past experiences, preferences and expectations, what makes emotions very ephemeral and hard to guarantee as a result of the interaction with a product [9].

Designing for fun requires a two-way communication between designers and players, in order to allow all used design patterns to be exercised throughout the entire game development and to gain insights into whether or not the aimed experiences are being achieved [7]. By creating prototypes of increasing quality, reviewing ideas, identifying potential players and asking them to test the game, a designer can have valuable feedback on her design choices and apply the needed changes as soon as possible [7]. Also, since the emotional experience is much more subjective than the satisfaction aspects of usefulness or safety, only the observation or inquiring of people playing can help identifying obscure aspects of the design and provide opportunities to incorporate good unexpected events discovered by players themselves [18]. Therefore, the practise of testing is fundamental to help improving the possibilities that a game is fun to broader audiences in different contexts.

The evaluation of games has been traditionally performed by specialists, either by observing people playing games or inquiring them about their experience with pre and post interviews and questionnaires. More recent approaches rely on physiological measurements collected from biometric sensors to help evaluating the emotional variations that accompany fun experiences [15]. But there are important difficulties with these techniques. The presence of an analyst may influence results because the acknowledgement of being observed can temper the player's emotions [11]. And even though self-reporting can reduce this influence, inquiring may produce wrong expectations if performed before the game is played, disrupt the experience if performed during interaction and not capture the player's real emotional state if performed after the game session is concluded [18]. Finally, the use of intrusive sensors applied to the player's body may also influence results by causing discomfort and making easier to break out of immersion (the involvement with the game) [22].

In that sense, the use of cameras for capturing facial expressions and extracting data about the emotional variations during game sessions seems to be a promising alternative. First of all, the face is an important channel for the expression of fun. The brain is hard-wired for facial recognition just as it is for language, being an important channel for communicating behavioural intentions [5]. Indeed, many studies performed by psychologists in the last four decades have lead to the general agreement that the expression of emotions in the human face is consistently interpreted among different cultures, being very important to the interaction between humans [1]. Also, facial expressions do not include only signs of emotions. The concentration in children's faces as they learn new skills – an important aspect of fun, related to challenges – is a good indication of the enjoyment they are feeling [5].

Additionally to that, the current technological state favours this approach because most of the video game consoles and almost all modern game-enabled mobile devices already have built-in cameras [22]. In the domain of digital games, the player attention is consistently focused on the output screen during playing, and even in full body games the participants observe the action on the screen most of the time [12]. These conditions remove or at least considerably reduce difficulties with image processing related to partial occlusion of the face (except regarding spectacles, hats or facial hair) and to low resolution (since the player face is commonly close to the device), and allow for an non-intrusive capture of data for automated analysis.

This paper describes a work in progress to develop a supporting tool for game designers, in order to help assessing the level of fun a person may experience while playing a prototype of a digital game. The measurement is based on just a low-cost camera for the capture of frontal face images, without requiring any data from the game itself. The system is focused on the assessment of fun from its emotional roots because fun is not only about the achievement of goals and the balance between challenges and skills (a more in-depth look on the psychological foundations of fun can be found elsewhere [24]).

[1] Laboratory of Interactivity and Digital Entertainment Technology, University of São Paulo, Brazil, email: {lvieira,fcs}@ime.usp.br

2 RELATED WORK

Since fun is subjective, it is more common to find research works attempting to measure it indirectly from its utilitarian (related to attention and performance) and non-utilitarian (related to emotions) aspects. A example of work based on the utilitarian aspect of fun was proposed by Levialdi et al. [13]. They attempted to assess attention in interactive systems from the analysis of prototypic emotions (neutral, joy, anger, disgust, fear, sadness and surprise) in facial expressions. Using a Naïve Bayes classifier, they detected the prototypic emotions and used them altogether with indicative features of cognitive effort (number of mouse clicks and keystrokes), to classify attention into three levels: low, medium and high. Results indicate correlation between the level of attention on a task and the production of facial expressions.

An example of work based on the non-utilitarian aspect of fun was proposed by Tan et al. [22]. They investigated weather sufficient facial expressions are elicited when games are played, and if those expressions can be robustly captured to help assessing the emotional states of players. Voluntaries played two mainstream commercial games as they tracked facial landmarks using a deformable fitting algorithm. Then, the prototypic emotions were predicted from an Artificial Neural Network trained from the local responses of a Gabor filter. After playing the two games, the participants filled the Game Experience Questionnaire (GEQ) [10] questionnaire. The authors observed that a good variety of facial expressions other than neutral were exhibited with rich variance. Comical scenes accurately followed the elevated detection of happiness, and anger also increased over time according to self-reported frustration in figuring out puzzles.

A more in-depth study on the assessment of fun can be found elsewhere [25].

3 MATERIALS AND METHODS

4 System Development

The general architecture planned for the system is presented in figure 1. In the first layer, the Face Tracker will record a video of the player in action and extract from each frame the face region altogether with the relevant features. From the segmented face region, the Eye Tracker will then extract the features related to the eyes. All features will be used by the modules on the next layer to perform three types of monitoring: the Immersion Monitor will produce an indication on the probability that the player is immersed in the game, and the Emotion Monitor will produce seven values as the probabilities that each of the six prototypic emotions and the neutral face are expressed in the face. Finally, these probabilities will be used by the Fun Classifier in the last layer to output the probability that the player is having fun at the moment. All values will be assessed in a predefined short time window, so they can be used to provide a timed-graph for visual inspection by a game designer.

The code will be developed with the C++ language and with OpenCV[2] for image processing and Qt[3] for the construction of the user interface. All code will be developed with portable standards, so it shall be usable in other operating systems, and will be released under the General Public License (GPL) [4].

[2] http://opencv.org/
[3] http://qt-project.org/
[4] http://www.gnu.org/licenses/gpl-3.0.en.html

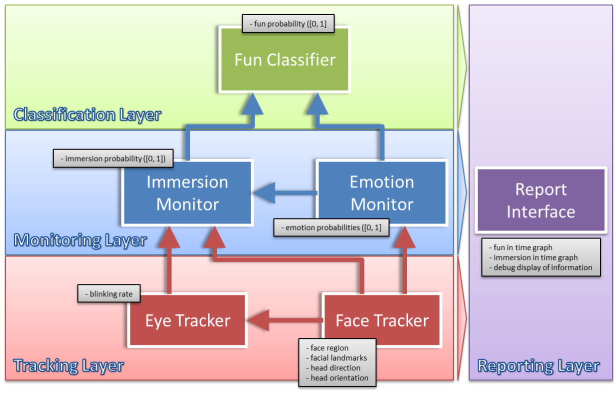

Figure 1: Overview of the system architecture.

4.1 Face and Eye Trackers

The Face Tracker will be based on the OpenCV implementation of the Viola and Jones algorithm [26] and on the Active Appearance Model (AAM) [4]. The Viola-Jones detector is going to be used for the initial face detection and the AAM algorithm will be used to track the face through a set of facial landmarks.

The Viola-Jones detector works by iteratively searching for an object in different scales (windows) of an image. It is trained from a large set of positive and negative image samples of the object of interest, from which it learns the best threshold values that will classify the images as positive or negative for each feature. The AAM tracker works from the adjustment (fitting) of a statistical model previously learnt in a new image. The model is represented in terms of the principal components of shape and grey-level. During the training phase, image samples annotated with connected landmarks are processed by Principal Component Analysis (PCA) to find the main modes of displacement and pixel grey level under the marks. New PCA is then performed to combine the two models into a single one. The model is initially guessed into a new image, and the fitting is done by iteratively computing the prediction on the corrections needed to reduce displacement error.

The OpenCV library does not have an implementation of the AAM algorithm, so one has been trained from the Extended Cohn-Kanade Dataset (CK+) of facial images [14]. This is a publicly accessible dataset containing 593 frontal images of posed and natural facial expressions. All images are annotated with the coordinates of 68 facial landmarks, and 327 of them are annotated with labels of the prototypic emotions.

The head pose will be estimated from the Pose from Orthography and Scaling with Iterations (POSIT) algorithm [23]. This algorithm estimates the orientation of a known object in three dimensions, by orthographically projecting four non-coplanar points from the image plane to a scaled plane, assuming that the object is far enough from the camera. The approximation is iteratively improved until it converges. It will be employed together with AAM to improve the tracking of the facial features, as well as to estimate the distance of the face from the camera as a measurement of the leaning.

The Eye Tracker will be responsible for calculating the open-close state of each eye and their blinking rate. The eyes will be segmented from the face using the information provided by the Face Tracker, particularly the landmarks related to the eyes. Blinks will be detected by simple comparison of frames. An efficient method is the binary segmentation of two sequential frames for comparison of their histogram values in the region of the eyes [16].

4.2 Emotion and Immersion Monitors

The Emotion Monitor will use the information provided by the Face Tracker to predict the probabilities of each individual prototypic emotion. A binary Support Vector Machine (SVM) classifier will be trained for each prototypic emotion, from the labelled images available at the CK+ dataset, and the scores will be transformed into probabilities through the Platt scaling procedure [17]. The SVM has been chosen because it is a classifier commonly used in the literature and it has good performance in the classification of emotions from the face [1].

The features used for training the SVMs will be the responses of a set of Gabor filters extracted from the facial landmarks tracked. Gabor filters are two dimensional matrices (kernels) formed by the complex signal of a sinusoidal carrier given by wavelength λ and orientation θ, attenuated by a Gaussian envelope with standard deviation σ. Filtering an image by a Gabor kernel is commonly performed by convolution in the spacial domain, producing responses proportional to how well the image features match the wavelength and orientation of the kernel used. A set of Gabor filters configured with different wavelengths and orientations provide a good descriptor of texture and are very robust to variations in lighting [6]. The set of Gabor filters used for training and classification of the prototypic emotions will be composed from the parameters in table 1.

Table 1: Parameter values that will be used to build the Gabor filters in the bank.

Parameter	List of Values							
$\boldsymbol{\theta}$	0	$\frac{\pi}{8}$	$2\frac{\pi}{8}$	$3\frac{\pi}{8}$	$4\frac{\pi}{8}$	$5\frac{\pi}{8}$	$6\frac{\pi}{8}$	$7\frac{\pi}{8}$
$\boldsymbol{\lambda}$	3	6	9	12				
$\boldsymbol{\sigma}$	0.56λ							
$\boldsymbol{\gamma}$	1							
$\boldsymbol{\psi}$	$\frac{\pi}{2}$							

The choice of these values come from the literature. A good bank of Gabor filters includes orientation (θ) varying from 0 to π in order to describe image features in all directions, and at least four variations of wavelength (λ). The values of λ were obtained from the work of Sigari and Fathy [20], who estimated the values that produce the best classification results. The standard deviation of the Gaussian envelope (σ) is automatically calculated from the wavelength. The human visual cortex responds to different edge orientations with the bandwidth between 1 and 1.8 octaves, so the ration σ/λ for a bandwidth of 1 octave leads to the value of $\sigma \simeq 0.56\lambda$ [8]. The aspect ratio as $\gamma = 1$ indicates that the Gaussian envelope is symmetrical (bell shaped), and the offset as $\psi = \pi/2$ indicates that the sinusoidal signal is shifted by $90\,\mathrm{deg}$.

The Immersion Monitor will be responsible for estimating the probability of the player being totally immersed by the game. Immersion is an state in which the player is totally involved with the game, to the point that the person's perceptual system might be tricked into believing she is somewhere else [2]. Hence, this state is strongly dependent upon attention and the ability to concentrate on a task. An SVM classifier will be used to try to predict immersion from features provided by the Face and Eye Trackers: proximity of the face to the screen and blinking rate.

Spontaneous eye blinking rate decreases during high-attention tasks in order to maximize stimulus perception and also when the "mental tension" is low due to task completion, serving as a relief mechanism [3]. Leaning towards an object of interest is also a display of overt attention, the one that is associated with external stimuli and involves the movement of a sensory organ to capture the stimuli data [21]. The movement of eyes is not an interesting feature because games have many different patterns on the screen, what would probably decrease the usefulness of this feature for the intended purpose.

Once the software module is developed, it will be used in a testing lab, with the games chosen and a group of voluntary players, to capture the features data. Players will also be requested to fill the GEQ questionnaire, since it contains a specific set of questions used to evaluate the sensory and imaginative immersion of players. The mean score of the answers will be used as the labels for training the SVM.

4.3 Fun Classifier and Report Interface

The Fun Classifier will be the one responsible for predicting if the player is either having or not fun while playing the game. It will also use a SVM classifier, but constructed upon the features provided by the modules in the monitoring layer, that is, the immersion probability provided by the Immersion Monitor and the probabilities of each prototypic emotion provided by the Emotion Monitor.

The training of the classifier will be made from the data collected in a different experiment, in which new volunteers will be asked to play the selected games and fill up the GEQ questionnaire. However, this time, the global score will be used as the labels for fun/not-fun. The classifier built will also be pre-evaluated by a cross-validation procedure, but when all that is done, a new experiment will be executed to try to correlate the responses given by the complete system in execution with the self-reports of a new set of volunteers. In this last experiment, it is intended to experiemtn with a game under development by a Brazilian game studio.

The Report Interface will simply be a graphical user interface which will control the initialization and termination of the whole system and present the results in a graphical form, easy for visual inspection of game designers.

5 PRELIMINARY RESULTS

A first version of the Face Tracker has already been started using just the Viola-Jones and the AAM algorithms. A video of a test session of this module working can be seen on Youtube[5].

The development of the Emotion Monitor has also been started, and the software has been pre-tested with cross-validation on static frames of the CK+ image dataset. The results, shown in figure 2, point that the easiest prototypic emotions to detect are fear and contempt, with respectively 96% and 94% of accuracy. The most difficult emotion to detect is by far surprise, with just 68%.

6 CONCLUSION

Evaluation is an important stage in the development process of any game. Traditional approaches have explored observation, interviews and the collection of both utilitarian and non-utilitarian data from the game itself or from biometric sensors applied to the players' bodies. All these approaches have their difficulties, as they may not be able to capture the experience as it happens or they can tamper the experience by being too intrusive. The analysis of facial expressions seem

[5] `https://www.youtube.com/watch?v=_tC7L_lGngc`

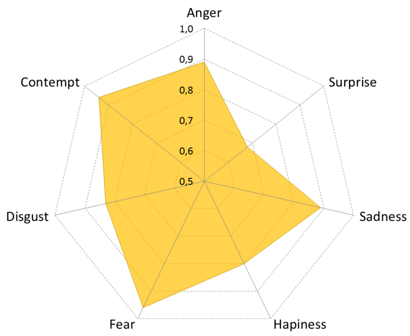

Figure 2: Per-class accuracy obtained with cross-validation tests with the CK+ face dataset

to be a interesting way to address these issues, since it is not intrusive and can be performed as a person plays a game. Also, it can help evaluating fun from the non-utilitarian aspect related to human emotions, since the human face is an important way for the expression of emotions and communication of intentions.

This paper described a work in progress of a system intended to help game designers to evaluate their games throughout the analysis of fun based on the automated detection of prototypic emotions and the state of immersion. The experiments performed so far have demonstrated that the proposed approach can work, even though there are still important issues to be addressed. First of all, the training data used in experiments with cross-validation is composed of both posed and non-posed expressions. Real non-posed expressions might be too subtle for a proper classification of emotions, and it is still to be verified if the immersion detection from blinking rate and body leaning can aid with that particular difficulty. Also, different types of games may induce fun from different levels of each prototypic emotion (for instance, horror versus comedy games regarding fear and happiness). Therefore, targeted evaluations shall be performed to verify the robustness of the fun classifier with different games.

ACKNOWLEDGEMENTS

The authors would like to thank CAPES for the financial support.

REFERENCES

[1] Vinay Bettadapura, 'Face Expression Recognition and Analysis : The State of the Art', *Computer Vision and Pattern Recognition*, **1203.6**, 1–27, (2012).

[2] Emily Brown and Paul Cairns, 'A Grounded Investigation of Game Immersion', in *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, 1297—-1300, ACM Press, New York, NY, USA, (2004).

[3] Siyuan Chen and Julien Epps, 'Automatic classification of eye activity for cognitive load measurement with emotion interference.', *Computer methods and programs in biomedicine*, **110**(2), 111–24, (may 2013).

[4] T. F. Cootes, G. J. Edwards, and C. J. Taylor, 'Active Appearance Models', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23**(6), 681—-685, (jun 2001).

[5] Mihaly Csikszentmihalyi, *Flow: The Psychology of Optimal Experience*, Harper Perennial, New York, New York, USA, 1 ed. edn., 1991.

[6] Beat Fasel and Juergen Luettin, 'Automatic facial expression analysis: a survey', *Pattern Recognition*, **36**(1), 259–275, (jan 2003).

[7] Tracy Fullerton, *Game Design Workshop: A Playcentric Approach to Creating Innovative Games*, CRC Press, 2 edn., 2008.

[8] Cosmin Grigorescu, Nicolai Petkov, and Michel a Westenberg, 'Contour detection based on nonclassical receptive field inhibition.', *IEEE transactions on image processing*, **12**(7), 729–739, (jan 2003).

[9] Marc Hassenzahl, 'Emotions can be quite ephemeral. We cannot design them', *Interactions*, **11**(5), 46–48, (sep 2004).

[10] W. A. IJsselsteijn, K. Poels, and Y. A. W. de Kort, 'The Game Experience Questionnaire: Development of a self-report measure to assess player experiences of digital games', Technical report, (2007).

[11] Katherine Isbister, 'Enabling Social Play: A Framework for Design and Evaluation', in *Evaluating User Experience in Games: Concepts and Methods*, ed., Regina Bernhaupt, Human-Computer Interaction Series, chapter 2, 11–22, Springer London, London, (2010).

[12] Mitja Koštomaj and Bojana Boh, 'Evaluation of User's Physical Experience in Full Body Interactive Games', *Haptic and Audio Interaction Design*, **5763**, 145–154, (2009).

[13] Stefano Levialdi, Alessio Malizia, Teresa Onorati, Enver Sangineto, and Nicu Sebe, 'Detecting attention through Telepresence', in *he 10th Annual International Workshop on Presence - PRESENCE 2007*, ed., Laura Moreno, pp. 233—-236, Barcelona, Spain, (2007). Starlab Barcelona S.L.

[14] Patrick Lucey, Jeffrey F. Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, and Iain Matthews, 'The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression', in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, number July, pp. 94–101. IEEE, (jun 2010).

[15] Lennart E. Nacke, 'An Introduction to Physiological Player Metrics for Evaluating Games', in *Game Analytics: Maximizing the Value of Player Data*, eds., Magy Seif El-Nasr, Anders Drachen, and Alessandro Canossa, 585–619, Springer London, London, England, (2013).

[16] Patrik P. Olatsek, 'Eye Blink Detection', in *Proceedings of the IIT.SRC 2013*, pp. 1—-8. Slovak University of Technology in Bratislava, (2013).

[17] John C. Platt, 'Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods', in *Advances in Large Margin Classifiers*, pp. 61—-74. MIT Press, (1999).

[18] Jesse Schell, *The Art of Game Design: A book of lenses*, CRC Press, 1st edn., 2008.

[19] Klaus R. Scherer, 'What are emotions? And how can they be measured?', *Social Science Information*, **44**(4), 695–729, (dec 2005).

[20] Mohamad Hoseyn Sigari and Mahmood Fathy, 'Best wavelength selection for Gabor wavelet using GA for EBGM algorithm', in *2007 International Conference on Machine Vision*, volume I, pp. 35–39. IEEE, (dec 2007).

[21] Darren Stanley, *Measuring Attention using Microsoft Kinect*, master, Rochester Institute of Technology, 2013.

[22] Chek Tien Tan, Daniel Rosser, Sander Bakkes, and Yusuf Pisan, 'A feasibility study in using facial expressions analysis to evaluate player experiences', *Proceedings of The 8th Australasian Conference on Interactive Entertainment Playing the System - IE '12*, 1–10, (2012).

[23] Kenneth Y. Tsai, 'Model-based object pose in 25 lines of code', *Seminars in cutaneous medicine and surgery*, **33**(2), 59, (jun 2014).

[24] Luiz Carlos Vieira and Flávio S. Corrêa da Silva, 'Understanding Fun', in *Videojogos 2014, Conferência de Ciências e Artes dos Videojogos*, Barcelos, Portugal, (2014). Sociedade Portuguesa de Ciências dos Videojogos.

[25] Luiz Carlos Vieira and Flávio S. Corrêa da Silva, 'Assessment of Fun in Interactive Systems: a Survey', in *10th International Brazilian Meeting on Cognitive Science*, São Paulo, (2015).

[26] P. Viola and M. Jones, 'Rapid object detection using a boosted cascade of simple features', *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, **1**, 511—-518, (2001).