Competitive Affective Gaming: Winning with a smile

André Mourão, João Magalhães
Departamento de Informática
Faculdade de Ciências e Tecnologia
Universidade Nova de Lisboa
Quinta da Torre, Caparica, Portugal
a.mourao@campus.fct.unl.pt, jm.magalhaes@fct.unl.pt

ABSTRACT
Human-computer interaction (HCI) is expanding towards natural modalities of human expression. Gestures, body movements and other affective interaction techniques can change the way computers interact with humans. In this paper, we propose to extend existing interaction paradigms by including facial expression as a controller in videogames. NovaEmötions is a multiplayer game where players score by acting an emotion through a facial expression. We designed an algorithm to offer an engaging interaction experience using the facial expression. Despite the novelty of the interaction method, our game scoring algorithm kept players engaged and competitive. A user study done with 46 users showed the success and potential for the usage of affective-based interaction in videogames, i.e., the facial expression as the sole controller in videogames. Moreover, we released a novel facial expression dataset with over 41,000 images. These face images were captured in a novel and realistic setting: users playing games where a player's facial expression has an impact on the game score.

Categories and Subject Descriptors
I.2.1 Applications and Expert Systems; **H.5.2** User Interfaces; **I.2.10** Vision and Scene Understanding.

General Terms
Human Factors, Design.

Keywords
Affective interaction, videogames, competitive games.

1 INTRODUCTION
Interaction between humans encompasses more than words, gestures or actions and the face is a fundamental part of communication. Facial expressions, for example, can help interaction by adding tone or intention to words. As Tian, Kanade and Cohn [28] put it, facial expressions are "the facial changes in response to a person's internal emotional states, intentions, or social communications". Facial expressions can represent a wide array of human behavior. Ekman et al. [7, 8] defined seven emotions that can be mapped into facial expressions. The face and the body can provide useful input in HCI. Steps towards this goal are gaining momentum and enabling users to control computers and game consoles with their bodies through facial expression [9, 20, 24, 25], gestures and other movements [21] (using depth-sensors, such as, Microsoft Kinect [12]) and to search information using their facial expressions [1]. This trend is also observed in recommendation systems that use text sentiment analysis [2].

In this paper, we propose to extend existing interaction paradigms by including facial expression based input in videogames. Affect-based interaction methods can greatly change the way videogames are designed and played. Thus, our objective is to research facial expressions as input in games. We argue that a game should be able to react to the player's facial expression as input. For example, in a fighting game, a punch thrown with an angry face could cause more damage; an adventure game could alter its outcome depending whether player acted surprised or not on a certain situation. To support our hypothesis, we researched affect-based interaction techniques in a computer. The proposed framework explores the player's facial expression as the sole controller of the game. Previous work has explored affective features in games [9, 21, 24, 25] or art exhibitions [11] but not at the level of interaction. The closest approaches, [7] and [13], are aimed at people with Autistic Spectrum Disorders (ASD). Their aim is to teach emotions by mimicking facial expressions.

NovaEmötions game is the central contribution of this article – it is an interactive game exploring affective features based on the face. Our scoring algorithm was designed to tackle the new challenges of affective interaction in videogames. We researched methods for measuring displays of affect, taking into account fair competitiveness and smoothed responses. Affect-based computer interaction still has many challenges to be researched and we believe the proposed game illustrates how such novel interaction paradigm can be embedded in computational systems. A game trial evaluation with 46 players showed that players got quickly engaged on the game and rapidly became competitive. We documented and captured the entire game sequences showing the player's affective responses during the game.

Other important contribution of this article is novel facial expression dataset made available to the scientific community for research purposes. We released more than 41,000 images of face images annotated with their facial expression. This dataset is unique in the following senses: captured in a realistic setting: user faces are not in fixed positions (about 50% of the face images are not front facing and are at different heights); users are playing a game where their facial expression play a role in the game, hence there is a true and direct interaction. We expect this dataset will help the development of new interaction technologies based on affective features.

This paper is organized as follows: Section 2 details the game dynamics and rules. The affect-based interaction implementation is detailed in Section 3. Section 4 reports the game evaluation and user trial. Section 5 we discuss work related to our approach.

MM '13, October 21 - 25 2013, Barcelona, Spain

http://dx.doi.org/10.1145/2502081.2502115

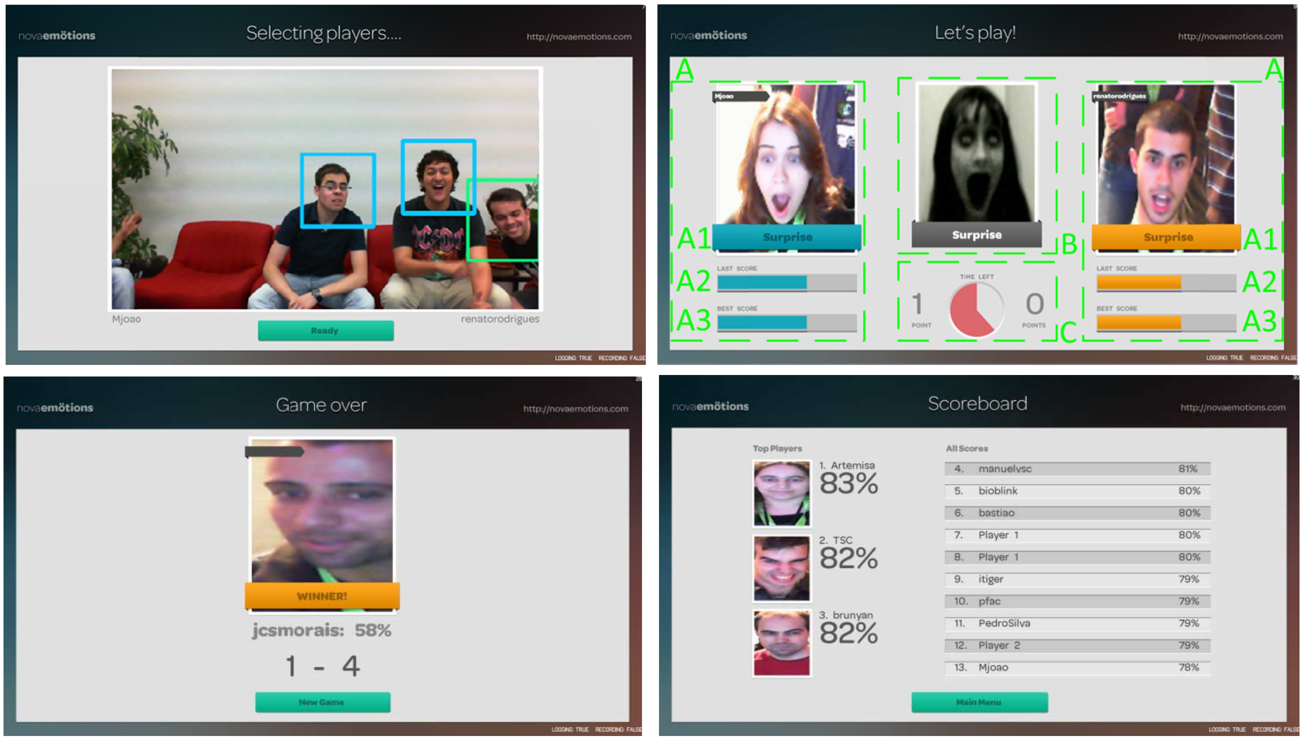

Figure 1. Top left: Player selection screen; top right: two players performing the *Surprise* expression; bottom left: winner screen; bottom right: high score screen.

2 GAME DESIGN

NovaEmötions is a game that explores players' facial expressions as the sole controller. Several challenges exist to accomplish such natural interaction [22]. The most critical one resides on designing smooth facial expressions recognition. The design of the game must not frustrate the user and must react homogeneously, independently of the expression being performed. With these requirements in mind, we identified the key design aspects of an affect-based interaction framework to be the following:

1. **Affective awareness** – users (i.e. players) must realize that the system (i.e. the game) is aware of them. This triggers a mindset that prepares them for the game interaction and the lack of a mechanical control;

2. **Affective challenge –** each game round must face a challenge to the player. This challenge assumes the form of an emotion that must be performed. The player displays the emotion through an intentional facial expression;

3. **Affect-based interaction** – the framework must limit the interaction to a well-defined set of affective user-responses. The reaction to players' displays of affection must be unambiguous and in real-time. Weakly specified interactions will frustrate users with unrelated actions-reactions.

We have found these design principles to be fundamental to achieve an effective interaction based on facial expression.

2.1 Gameplay

NovaEmötions is a two-players game where the objective is to perform a set of facial expressions. Players play simultaneously and facial expressions are competitively scored. The player that performs an expression closer to the one asked, wins a round – the player who wins more rounds wins the game. Each round consists of the following sequence (Figure 2):

1. A reaction image is displayed, along with the label of the expected expression;

2. When a face is detected, the facial expression is recognized, player's label is updated with the recognized expression and the scores are updated. The process is repeated through the round, until the timer ends;

3. At the end of the round, the player with the highest score wins a point. If the player's best score is the same, they both get one point;

4. The game will continue until a set number of images are displayed. In the end, a screen with the winner is displayed.

The main screen (displayed in Figure 1: Top right) is composed by three components: the player's information component (A) that contains the currently detected expression (A1), the last obtained score (A2) and the best round score (A3); the current round challenge expression label and stimulus image (B); and the game total round score and the current round timer (C). Players were asked to perform the expression of the expected label.

2.2 Affective awareness and challenge

Players must realize that the game is aware of their affective behavior. In the first step, everybody is detected by the game (square around the faces), Figure 1: Top left, and the two players are manually selected (blue square). This creates an affective setting where players quickly realize that the game is aware of them and of their facial expressions. This state of mind is reinforced by the fact that there is no mechanical control – the game tracks player's faces and monitors their facial expressions without interaction (other than the initial player selection).

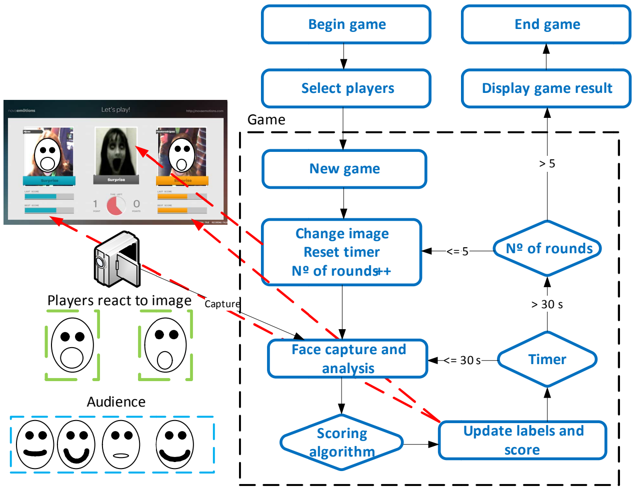

Figure 2. Game overview.

2.3 Assessing facial expressions

The previous section described the design aspects guiding our affect-based interaction game. In this section, we detail the researched framework at the core of the game: a computational representation of facial expressions and a facial expressions similarity computation. These elements will feed the most important element of the framework: the NovaEmötions scoring algorithm (section 2.4). The scoring algorithm is critical in the sense that it computes the game reaction to the player's affective expressions. Thus, it must offer a consistent response to all affective interactions, without favoring any particular facial expression due to algorithm accuracy issues.

2.3.1 Face detection and alignment.

The affective component of the proposed game runs in real-time and is unsupervised requiring no calibration (no need for user intervention to tune algorithms). Therefore, we use a face detection algorithm based on the Viola and Jones' algorithm [30]. To improve the alignment of the detected faces and reduce the false positives in face detection, the Viola and Jones' algorithm is also applied to eye detection. The alignment algorithm detects the eye pair and rotates the image with the assumption that the eyes have the same y-axis value. The variations in pitch and yaw are not corrected, due to a lack of reference points. Face detection is reliable with a ±10º roll and ±5º pitch and yaw. Faces rotated outside this range are ignored.

2.3.2 Facial expression detection

To represent facial expressions, we have chosen the expressions from the Emotional Facial Action Coding System (EMFACS) [8]. EMFACS is built on the assumption that there are seven facial expressions: *Happiness*, S*adness*, *Surprise*, *Fear*, *Anger*, *Disgust* and *Contempt* and a state of no expression: *Neutral*. These facial expressions are decomposed into a set of Action Units (AU). An AU corresponds to visible muscle movements that we can approximate with Gabor wavelets and face regions determined by Gabor energy clustering (e.g. mouth or eyebrows).

Using ideas already deployed in visual analysis of facial expressions [32] [15], we implemented a dictionary of Gabor filters to extract face contours. Combinations of Gabor filters are widely applied in facial expression recognition because of their ability to detect the contours of the facial features (i.e., eyes, nose, mouth, brows and wrinkles) and filter out most of the noise present in the image [3]. To extract information concerning the face contours and expression traces, twenty-four Gabor filters with six orientations and four scales capture the different details of a facial expression. Dictionaries with this configuration have been found to work well on a number of domains, namely, facial expressions analysis [3] and image retrieval [18].

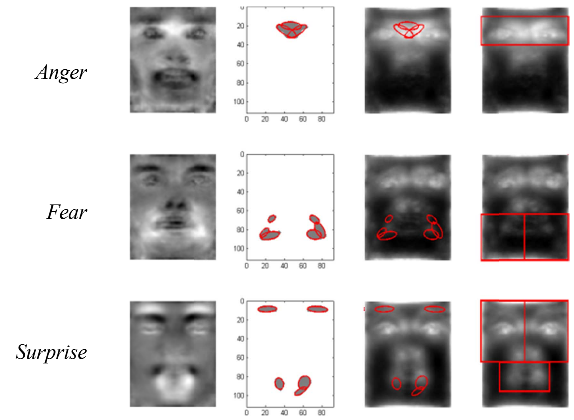

Figure 3. Creation of the expression regions.

In addition to the Gabor representation, our objective was to identify the face regions where the dominant facial expression changes are modelled by a *k*-means clustering [17]. We considered a sequence of images per expression/user, starting at neutral expression and ending at peak expression. To calculate the region with greater variation from the neutral expression, we took the frame with the maximal facial expression from each sequence and convolved it with the Gabor filters (Figure 3, column 1). To identify the regions with the dominant changes for the specific facial expression image, we subtracted the Gabor filters output of the neutral image of that sequence, and run the k-means clustering algorithm to detect the high-energy regions (Figure 3, column 2). Figure 3, column 3 exemplifies the clusters over original face images. For efficiency and robustness purposes, the resulting *k*-means clusters were approximated by rectangular regions (Figure 3, column 4). These regions are robust to small changes in the alignment of the facial features and simplify feature extraction, improving the algorithm's robustness.

Finally, we extract the average value and standard deviation of each rectangular region for each filter scale and orientation. To classify a facial expression and extract the expression label, we combine our representation with a multiclass SVM. The algorithm is described and evaluated in detail in [19]. Thus, the SVM classifier will classify each facial expression with a given expression label, i.e., the *playerEmoLabel* variable of the scoring algorithm.

2.3.3 Dissimilarity between facial expressions

To deploy affect-based interaction, we need to measure the intensity of a player's facial expression. We need to compare the player's facial expression to a set of reference facial expressions. This is done through a dissimilarity measure based on the L_0 norm. We use the L_0 norm since it measures the number of dimensions in a vector that are different from zero – thus, it is more robust. Using the representation described in the previous section, we compute the difference between the player's expression feature vector pf, and the average feature vector af_j of an expression j,

$$\Delta_j = \left(pf - af_j\right) = \left(\Delta_{j,1}, \dots, \Delta_{j,i}, \dots, \Delta_{j,n}\right)$$

and remove the noisy dimensions that have a magnitude smaller than a threshold ε_0 ($\Delta_{j,i} = 0$ if $-\varepsilon_0 < \Delta_{j,i} < \varepsilon_0$). Computing the L_0 norm of this vector, produces a measure of the intensity of the facial expression – this will be the *playerEmoDissim* variable of the scoring algorithm.

2.4 Scoring algorithm

The game scoring algorithm must provide a realistic and balanced reaction to the players' expression. This is critical to keep the user responsive and immersed in the game. A significant challenge was to obtain a meaningful score that would answer players' expectations. In this section, we argue that although both expression dissimilarities and SVM are incomplete, a smoothed and temporally limited combination of these techniques can deliver a competitive scoring algorithm for affect-based games.

A baseline scoring algorithm would take the label from the SVM and the expression dissimilarity score to directly update the game interface. This approach has several properties and drawbacks inherited from the previously presented techniques:

1. Some players are detected more often than other players, giving them more chance to score;
2. The SVM classifier is highly accurate for the on-screen label of the player's face but it is not adequate for computing an expression score;
3. The expression dissimilarity is a good indicator of how close the player's expression is to average expression, but it is not adequate for providing the on-screen label;
4. A player's expression labeled as *Anger* could still achieve the highest score for a *Disgust* label.

To overcome these challenges, our approach was to devise a novel game scoring strategy to increase the competitiveness of the game and overcome algorithmic limitations. At the core of our approach is the prior knowledge that we know what the expected expression is. We explore this information, along with the score and the label, to create improved labels and scores to design the game's scoring algorithm.

2.4.1 Fair competitiveness

The first goal towards a fair game scoring scheme, is to reward the best facial expressions and not a large number of average facial expressions. This decision was made to avoid frustrating players that are not detected as often as others. Thus, the game encourages the most acute facial expressions.

The emotion label, output variable *Label,* to be shown to the player (Figure 1: Top right A1), is an indicator of how the player is being interpreted by the game. This label is determined by one of two ways (Algorithm 1, steps 1 to 3):

1. If the dissimilarity between the player's expression and the current round expression (*playerEmoDissim*) is above the HIGH_CONFIDENCE threshold, the label corresponds to this round's expression. This ensures players receive the correct label if they achieve a high dissimilarity value and the SVM output is ignored.
2. If the dissimilarity value (*playerEmoDissim*) is not high enough, the Label is assigned by the SVM classifier (step 3).

The HIGH_CONFIDENCE threshold was determined experimentally to an accuracy level of 95% on the CK+ dataset.

Once the label has been determined, the score of the current facial expression, we calculate the *ScoreMeter*. The score meter is displayed in the game screen at position A2, see Figure 1, top right.

If the current label is correct (Algorithm 1, step 4), but the dissimilarity is too low we assume there's too much noise in the image (Algorithm 1, step 5), and adjust the value by adding a uniformly random value between 0 and BONUS (Algorithm 1, step 6). This solves two issues: (i) the disagreement between the dissimilarity and the SVM, and (ii) provides a meaningful response to the user to keep him responsive.

Algorithm 1. Calculating the player's score and label.

Input:
playerEmoLabel *label predicted by SVM classifier*
playerEmoDissim *L_0 dissimilarity for the captured face image*
roundExpression *label displayed to the user*
time *current round time*

Output:
Label *emotion label to display to the user*
ScoreMeter *score for the last image to be displayed the user*
BestScore *score of the player's best facial expression*

Algorithm:
1. **if** playerEmoDissim > HIGH_CONFIDENCE
2. Label = stimuliEmotion *assumes SVM label*
else
3. Label = playerEmoLabel *assumes dissimilarity label*
endif
4. **if** Label == stimuliEmotion
5. **if** playerEmoDissim < NOISY_EXPRESSION
6. ScoreMeter = playerEmoDissim + rand_uniform(0, BONUS)
else
7. ScoreMeter = playerEmoDissim
endif
8. ScoreMeter = temporalSmoothing(Score, time)
else
9. ScoreMeter = playerEmoDissim - rand_uniform(0, REACT)
endif
10. **if** BestScore < ScoreMeter
11. BestScore = ScoreMeter
endif

If the player's facial expression is not recognized as the stimuli emotion (Algorithm 1, step 9), the score will be penalized by an uniform random value. This happens when the expression performed is not close to the expected one (both the classifiers and the dissimilarity gave results different from the expected label), so it's fair to penalize the score and show the incorrect *playerEmoLabel* (different from *roundExpression*).

Since in some cases the detected faces are too noisy and the visual analysis is too slow (search for faces on the image) we add a jitter to the player's last score to inform the player that the game is responding to their expressions. This is done in Algorithm 1, steps 6 and 9 where small random values are added to the score. These random values are small and don't have an influence on the game.

Finally, the best score (output variable *BestScore*) is displayed in the game screen at position A3, see Figure 1, top right. This corresponds to the best scored obtained until a given moment. The score of the round (*BestScore*) is the value that will determine which player wins the round. To avoid rewarding players that are detected more often than others and to reward the best peak expression, the *BestScore* is the best score of the round, instead of a sum of individual images scores (Algorithm 1, steps 10 and 11).

Figure 4. User study room conditions: two players are playing the game surround by an audience.

2.4.2 Progressive competitiveness

To maintain the competitiveness throughout the entire round duration we imposed a limit on players' score. This will allow players to adjust themselves to the camera and to train their facial expressions. The limit is progressively removed until the score is no longer limited. This is implemented on Algorithm 1: step 8. The score is capped to 30% on the first 10 seconds of the round, by multiplying the score value from by 0.5. Between 10 and 20 seconds, the multiplier increases linearly from 0.5 to 1. On the last 10 seconds, the score is fully unlimited (multiplier equal to 1).

3 EVALUATION

The real setting evaluation assesses the scoring algorithm's performance with 46 players who played the game for at least five rounds. A user study was conducted with these players to assess the different aspects of the game. An initial trial was conducted with more than 100 players.

3.1 Methodology

The purpose of the game evaluation is to assess the effectiveness of the addressed hypothesis: using affect-based interaction in competitive games. This evaluation was conducted on a real gaming environment, where the 46 volunteers were asked to play the game and to answer a questionnaire about their experience. In this section we discuss the performance of the game algorithms in this setting. Section 3.5 presents the questionnaire results.

Table 1. General game statistics.

Number of players	46
Total number of games played	29
Average nº of games played per player	1.26
Nº of images player per game (fixed)	5
Time per image (fixed)	30 seconds

Initially, players were briefed about the objective of the game and the interface was explained. Then, they played the game by themselves. The trial was conducted like a regular game session on a social environment (i.e. multiple friends watching), contrasting with the strictly controlled conditions when capturing images for most face datasets (e.g. the CK+ dataset). Figure 4 illustrates the experiment conditions. In our trial, players moved their faces a lot (sometimes even stopped looking at the camera) and performed unexpected actions (mainly laughter). At the end of the interaction, they were asked to answer the questionnaire. Table 1 summarizes the captured data statistics.

The camera processed 9 images per second. Each player's face was processed and its score updated, over than 4 times per second.. Note that the facial expression and the game UI are running in parallel on the same computer.

The SVM classifier was trained on the CK+ dataset, see [16] for details. A round expression corresponds to the average of all images of that given expression on the CK+ dataset.

3.2 A Novel Facial Expressions Dataset

The captured face data is available to the scientific community. We captured and classified 42,911 facial expressions, with an average of 295 facial expressions per round. These face images were captured in a novel and realistic setting: humans competing in games where players' facial expression have an impact on the game These images offer a novel view of facial expression datasets: players were competing using their own facial expressions as an interaction mechanism, instead of performing well defined prototype expression.

Figure 5. Example of captured faces.

Part of the data is available for research purposes[1]. It is composed of over 41,000 images annotated with a facial expression. Each image contains the information regarding the expected expression, the expression detected by the scoring algorithm and human judgments obtained by crowdsourcing. Evaluators were asked to choose one expression from the set *Happiness*, S*adness*, *Surprise*, *Fear*, *Anger*, *Disgust, Contempt, Neutral* (state of no expression), and *Ambiguous* (does not fit any of the categories). The crowdsource labeling job is described in [26].

3.3 Affective interaction assessment

The importance of the facial expression recognition in this game is twofold: first, it must recognize the presence of the expected facial expression; second, it ought to provide players with adequate feedback about how the game is assessing their facial expression. Table 2 contains the confusion matrix for the label (expected) expression versus the detected expression. This value is determined by taking the displayed image label (*roundExpression*) and comparing it to the label computed by Algorithm 1.

Table 2. Scoring algorithm confusion matrix.

	Ang.	*Con.*	*Dis.*	*Fear*	*Hap.*	*Sad.*	*Sur.*
Ang.	**54.91**	1.13	0.00	5.00	19.01	1.63	18.32
Con.	16.52	**62.50**	0.52	1.72	8.16	0.16	10.42
Dis.	9.09	4.53	**60.54**	3.61	11.39	0.52	10.32
Fear	4.41	4.63	0.82	**64.61**	8.72	2.68	14.13
Hap.	2.37	2.90	0.24	4.39	**84.89**	0.86	4.35
Sad.	7.92	3.42	1.40	3.54	10.91	**63.94**	8.87
Sur.	10.42	2.67	2.57	5.56	20.52	0.96	**57.30**

The global facial expression recognition rate of Algorithm 1 was 68.23%. This value corresponds to the rate where the player's detected expression (Figure 1: Top right A2) corresponds to the label (Figure 1: Top right B). We consider this value to be the percentage of successful interactions. The value is consistent across all expressions, except for *Happiness,* where it is visibly higher. We consider that it is because the *Happiness* expression is easier expressions to express and detected because we had more training data for this facial expression

Although we expected players to be always performing the expected expression, this was not true. The competitive setting put players the under pressure of wanting to win. This pressed players who performed a wrong or ambiguous expression, or were not able of acting the expression naturally. Although these situations reduced the classification performance, we observed that it is a consequence of the player's engagement on the game. In the Scoring algorithm assessment and the User study we will return to this issue.

3.4 Scoring algorithm assessment

To evaluate the scoring algorithm, we selected the best rounds and the worst rounds, and compared their performances. All game rounds were divided into these two categories: the best rounds are the ones that finished with more than 80% of the highest possible score, the worst rounds are the ones with less than 80% of the highest possible score. This allows us to observe the different gaming strategies. The three curves on the above graphs represent the score values from Algorithm 1: (a) *Score* is the current score meter displayed to the player; (b) *BestScore* is the best score of the round and (c) the *PlayerEmoDissim* is the dissimilarity returned by the L_0 metric.

In the "best rounds" graph, Figure 6, the scores evolved as expected: the *BestScore* is very similar to the time penalty, except for the unlimited zone, where players where still able to increase the *BestScore* slowly. Other very interesting fact is that the *playerEmoDissim*) is very consistent across the entire round with a value range between 74 and 80%. As this value is only dependent on the captured face image (not affected by the scoring algorithm, this means players were able to maintain consistent expressions across the round. Thus, they were consistently competitive during the entire round.

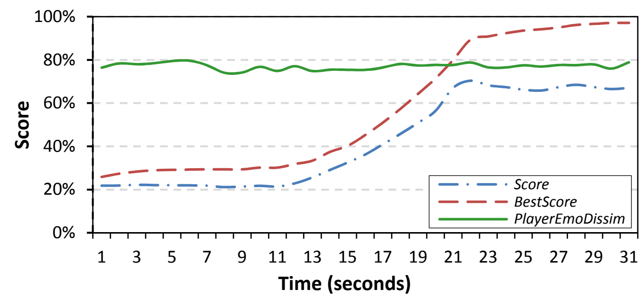

Figure 6. "Best rounds" score evolution.

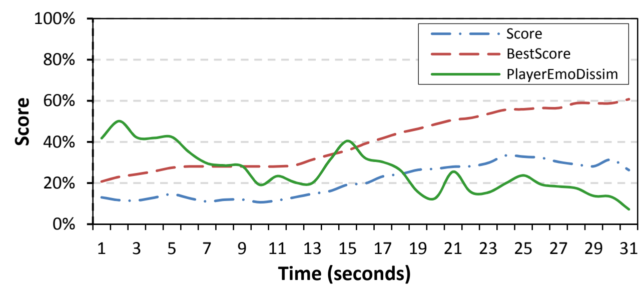

Figure 7. "Worst rounds" score evolution.

In the "worst round" graph, Figure 7, the most obvious difference is the scores range and the lack of consistency of the *PlayerEmoDissim*. It varied between 7% and 50% and it decreased as the round passed. The time penalty imposed on the score is also visible in both the *Score* and *BestScore*, but their values are much lower than the ones from the "best players". The *Score* value is much smoother than the *PlayerEmoDissim* value, showing that the Scoring algorithm was capable of smoothing out the dissimilarity inconsistencies and deliver a better gaming experience (players could be performing very poorly but their score would always correspond to their "*best*" facial expression).

Looking at both graphs, one can observe specific trends and differences across both categories. Best players had a high dissimilarity score throughout the round duration, meaning that they were highly competitive until the very last moment of the round. The worst players, showed a different trend: either they would stop performing the expected expression in the first moments or their attempts to perform the correct expressions got worse as the round went by.

[1] http://novasearch.org/datasets/

3.5 User Study

At the end of each game, players answered a questionnaire about their gaming experience. We took into account the heuristics from [5] that could be applied to NovaEmötions when preparing the questionnaire. Besides standard questions regarding player information (gender, age) and general gameplay (e.g. game objective, difficulties felt, enjoyment level), we also addressed NovaEmötions specific aspects such as interaction novelty, enjoyment and perceived accuracy (label accuracy and score accuracy). There were also open answer questions, where players could contribute with suggestions and critics.

Players were mostly undergraduate students, aged between 18 and 25 years old. The gender distribution was balanced (24 male and 22 female).

3.5.1 Game design assessment

Players found the game easy to understand (91% of the answers were "High" or "Very High") and enjoyed playing (91% of the answers were "High" or "Very High"), Figure 8. We consider that the difficulty level is adequate (not too high nor too low), as the majority of the answers regarding the difficulties felt during the game were "Medium". We consider this difficulty level adequate, as 91% of players enjoyed the game competitiveness.

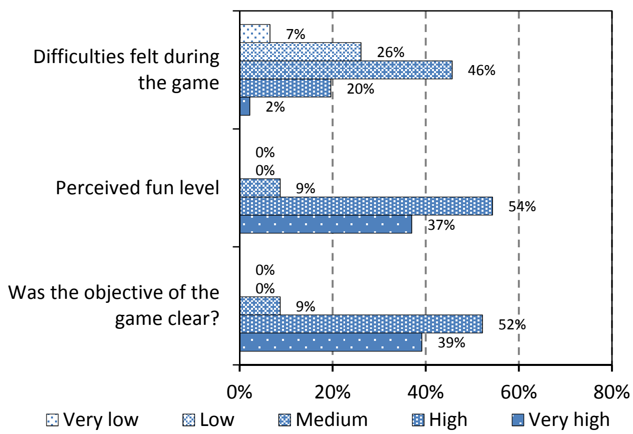

Figure 8. Game play assessment.

Regarding the image exhibition time (round duration) and number of rounds per game, Figure 9, 46% of the players wanted more images per game and 30% wanted less exhibition time per image. This is in line with what we observed: if one of the players got a very high score at the middle of the round, the competition on that round could end sooner. Thus, players expected affective stimuli to be shorter and more frequent.

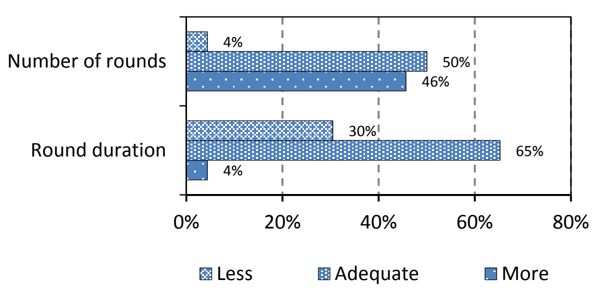

Figure 9. Rounds' duration and images.

When asked what would be the adequate number of players, 69% of the players chose "Two players", while the remaining 31% chose "More than Two players" (no one chose the option "One player"). One of the possible reasons for the relatively low number of "More than Two players" answers is that the users were presented with a two player version of the game, which could lead to a bias towards that answer.

3.5.2 Affect-based interaction

This group of questions concerns how players assess the various aspects of the game design (Figure 10). The large majority of players liked or loved the usage of facial expression as a controller (89%), and novelty of the controller type (98%). This is a highly positive result supporting the initial hypothesis of using the face as a game controller, and supports the effectiveness of the proposed solution.

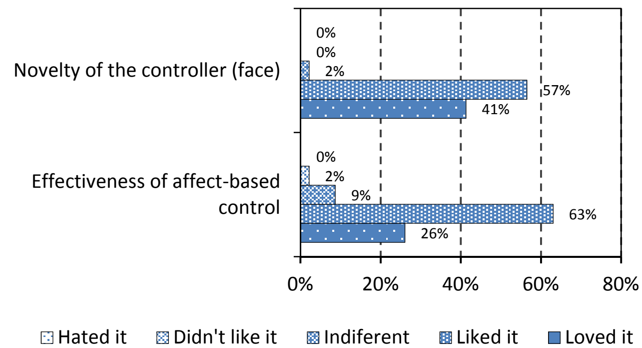

Figure 10. Affect-based interaction effectiveness.

Other critical aspect was the perceived accuracy of the score and the label by the players, Figure 11. Most of the players considered that the score was accurate most of the times (66%: 4 or over, average: 3.6), with a small reduction of when the question was about the label (43%: 4 or over, average: 3.3). This result is positive, as it shows that more than half of the players were satisfied with the label accuracy. It is important to compare these results to the algorithm's performance presented on Table 2. The perceived accuracy (3.3 in 5: 66% for label) is in line with the measured accuracy (68.23%) in our formal evaluation (the confusion matrix diagonal).

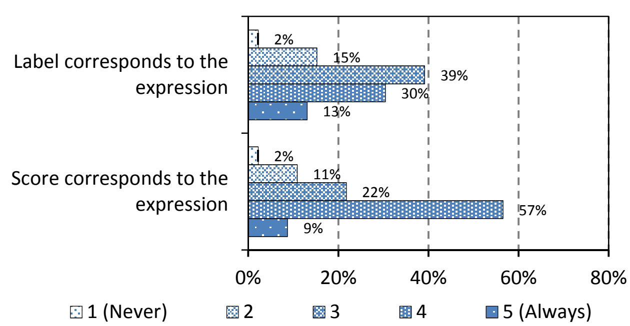

Figure 11. Perceived accuracy by the players.

We also investigated the accuracy of specific expressions and measured how players felt about the different expressions and expression-specific difficulties they encountered (Figure 12). There are three main important conclusions that can be drawn from these answers. Regarding the "Hardest expression to perform", the answers were quite distributed across expressions, except for contempt. In this specific question, 35% of the players reported that they did not know how to perform a *Contempt* expression. This result is backed up by our observations. Some players said that they did not knew how to make a *Contempt* expression, situation that did not happened with other expressions. All other expressions appealed for a player's reaction. Thus, we believe that *Contempt* is not a good expression for affect-based interaction gamming.

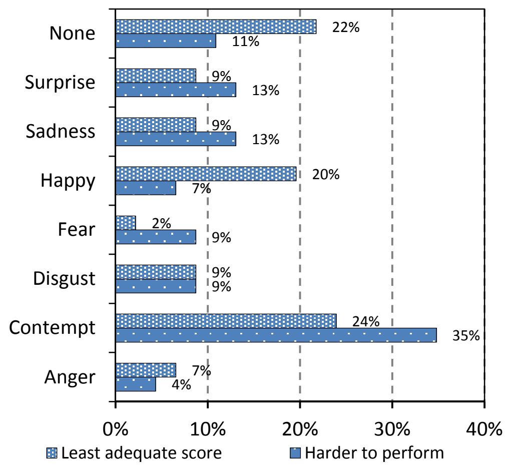

Figure 12. Expression specific assessment.

Regarding the expression whose score was least adequate, the two top results were *Contempt* (24%) and *Happiness* (20%), although the reason for the high percentage is different. In *Contempt*, there is possibly a relation with the previous question. Most players found this expression hard to perform and thought that the score was not accurate was a result. According to our study, volunteers found the *Happy* expression to be one of the easiest to perform and with the least adequate score. The cause for these answers is related to the classifier's good performance in detecting happy faces. Thus, the least adequate in this case, means players would get a higher score with less effort.

4 RELATED WORK

The Mimic Game [25] proposes an interesting application of facial expression recognition techniques: a synthetic agent mimics the facial expression and head pose of a person from a real-time video feed. The system detects the facial components and head pose using a calibrated Active Appearance Models (AAMs) [6] and maps the detected mesh into a two-dimensional emotion space. A point in this space corresponds to an emotion (from EMFACS) and intensity (from neutral face to full scale expression) pair. There are several games based on facial expressions aimed mainly at children with Autism Spectrum Disorders (ASD). What a Feeling [20] is another game based on the recognition of facial expressions by the player. The main idea is to make the player recognize the facial expression of the avatar in multiple situations (e.g. normal face, face with top omitted, micro-expressions). The therapist can control the exercises displayed to the patients. The Emotion Mirror [4] is a game designed to teach children with ASD about facial expressions. The main focus of the game is to mimic the facial expression of an avatar and to have the avatar mimic the player. The game's didactic approach cycles the direction of the interaction ("it mimics you" and "you mimic it") and rewards the player with a virtual ice cream that grows when the expression is correct. Both games use cartoonish avatars to produce the expressions, instead of the face of the player. Most of the studied facial expression based games are didactic and focused on helping people with ASD; their main objective is to reward well performed expressions and help players understand the different expressions. In contrast, our approach is based on competitiveness: it does not matter if your expression is completely correct, as long as it is better than the expression of your opponent. Thus, we argue that a facial expression can be used in non-didactic games.

Other types of affect based interaction are also being applied in games. Paiva et al. developed SenToy [21] a doll that recognizes emotion using motion sensors when manipulated by a user. Different gestures lead to the following emotions: *Anger*, *Fear*, *Surprise*, *Sadness*, *Gloating* and *Happiness*. The doll was evaluated in two trials: in the first trial (similar to a training stage), volunteers were asked to control the emotions of a virtual character by interacting with SenToy. In the second trial, the researchers tested the detected gestures from the first trial in a game. The users' response towards SenToy was very positive.

Novel techniques are being developed to evoke emotions on people using innovative media. Wang and Marsella [31] developed a video game called Emotion Evoking Game (EEG), designed to provoke emotions on the player. The game was created to aid the development of systems that analyze emotion and facial expression. They made a small study that consisted on having a small pool of volunteers that played a version of the game design to provoke four different emotions (boredom, surprise, joy and anger) at specific stages throughout the duration of the gamming session. The player's face was being captured with a webcam and they were asked to fill a form at the beginning and end of the game regarding their emotional state at the key moments of the game. The video from the camera was analyzed by the researchers and compared with the answers to the forms. The results were not consistent amongst emotions, producing some unexpected reactions to the programmed events in the game.

SOEmote [24] is the facial expression detection component for Everquest 2 that enables the game character to mimic the player's facial expression. The game detects the position of several facial features using a webcam and maps them into the face of the character. SOEmote also features a voice modulator integrated in the game's voice chat, allowing the player to alter its voice tone to better match the virtual character. Everquest 2 is a fantasy massively multiplayer online role playing game where players are encouraged to cooperate to defeat their enemies. SOEmotion allows players to extend immersion even further. Reception by the players was mixed [10, 27]: some players praised it for the innovation while others argued that it did not add anything to the game experience.

A key component of the proposed game is the facial expression analysis technique. Humans are able to recognize different facial expressions and infer what emotion an expression conveys. *Happiness*, *Anger* and *Surprise* are some of these emotion specific expressions [7]. One of most used systems to define facial expressions and the one we use is the Facial Action Coding System (FACS) [28]. FACS primary goal was "to develop a comprehensive system which could distinguish all possible visually distinguishable facial movements" [8]. FACS can be seen as an index of Action Units (AUs). An AU is an individual action that humans are able to distinguish, that can be performed by one or more muscles of the face. FACS is widely used as a standard since its introduction. It combines sets of different positions in face muscles and features to determine an underlying facial expression.

The EMFACS (Emotion FACS) system [8] was created by the developers of FACS to map AU combinations into emotions. EMFACS was built under the assumption that a facial expression conveys how the person is feeling. However, mapping facial expressions into emotions is challenging. Some facial expressions can be represented by various combinations of AU, e.g. Sadness can be represented as "AU1+4+15 or AU1+4+16" [16]. Some AU are only used for one particular expression (for example, AU20 is only present in the expression *Fear* [16]).

We deal with facial expression captured on an entertainment environment. For that reason, the player expression and crowdsourced expressions might differ from the expected prototype expressions. Nevertheless, we relied on the prototypes to give us a baseline expectation of the facial components actions. The physiological discussion of emotion versus facial expression is outside the scope of this paper.

After choosing a representation, it is necessary to turn an image into a set of features for facial expression recognition. Gabor wavelets were a popular choice at the First Facial Expression Recognition Challenge [29]. Littlewort et al. [15] used Gabor wavelets based features in automatic facial expression recognition in real-time video. They detected faces using an algorithm based on Viola and Jones' [30] face detection algorithm, extracted the facial features using Gabor filters and then used two approaches for recognition: multiclass decisions using SVMs and AdaBoost. They achieved good results with 91.5% recognition rate with SVMs. Gabor wavelets allow for automatic facial expression recognition; they eliminate manual selection of facial features.

Psychologists use images [14] to study facial expressions and emotional response on people. The international affective picture system (IAPS) [13] is a database of pictures used in the medical community to study emotional responses in people. It was built by showing various images to people and measuring their emotional response. Given the usage restrictions of this dataset, we could not use it in our gaming scenario with multiple persons being exposed to the images simultaneously. Savva et al. [23] studied body expression for emotion analysis using videogames to elicit movement. They tracked the positions of body parts across time to detect a series of abstract emotion categories (e.g. High-intensity negative emotions). They achieved an overall accuracy of 61.1%, comparable to the human observers' body expression agreement (61.49%).

Iacobini et al. [11] studied the contagion of emotions using interactive art. The art exhibition used facial expression recognition to measure the emotional state of the visitors and respond with a video from a database of emotional video portraits. At the end of the exhibition, interviews showed that the interaction with the exhibition was able to invoke emotional response and that the presence of more people added richness to interaction, as multiple people tried to collaborate influence the system. Some visitors also tried to control the responses of the exhibition through posed facial expressions.

5 CONCLUSION

In this paper, we proposed an emotion-based game exploring facial expressions as the sole interaction mechanism. The main conclusions to be drawn from our contributions concern three main points:

Gamification. Game trials illustrated how affective interaction can be successfully used as a computer game controlling mechanism. Despite the limitations of current state-of-the-art image-processing techniques, the proposed game design was able to deliver an emotion-based game. In particular, the NovaEmötions scoring algorithm is the key component implementing the game competitiveness: the affective interaction.

Affective-interaction and competitiveness. The images related to some expressions required a more difficult reaction. In line with the previous point, players reported that some of the expressions were hard to perform (in particular *Contempt*). The competitiveness factor of the game distorted the link between emotion and facial expression. *Contempt* did not work as well as other expressions in our game and we have considered removing it entirely. Thus, game designers must take into account if the expressions are recognizable and easy to perform by players when integrating facial expression recognition in a game.

From a **facial expression analysis** point of view, we observed that the game response time to players' facial expressions and round time was quite critical. A smoothed combination of a facial expression detector and similarity computation (Algorithm 1), provides an appropriate scoring of players' facial expressions. We believe more games will explore facial expressions to measure rage, strength or tiredness and let it play an active role on the game's outcome.

Social component is key. We observed that when players came in larger groups (5 or more people), they had more fun. The higher the number of people watching, the higher would be the enjoyment of the players (bursting into laugher would be more common). Thus, the social environment allowed people to explore the affective interaction more freely.

From the user study (Figure 10), we believe that affective interaction can be quite effective in social/party games, promoting interaction instead of isolation. Having the face on the screen (no "avatars") was critical. We observed that players started laughing when the face appears on the screen before the start of the game (in the "Player Selection screen").

The final contribution of this article is a **dataset of facial-expression images** of users controlling a game with their facial expression. As far as we are aware of, this is the first publicly available dataset captured in such affective gaming scenario.

Acknowledgements. We would like to thank the volunteers who tested the game and answered the questionnaires for the game evaluation and user study. We also would like to thank Bruno Cardoso for the insightful comments and useful discussions. This work has been partially funded by the Portuguese National Foundation under the projects UTA-Est/MAI/0010/2009 and PEst-OE/EEI/UI0527/2011.

6 REFERENCES

[1] Arapakis, I., Jose, J.M. and Gray, P.D. 2008. Affective feedback. *ACM SIGIR* (2008), 395–402.

[2] Arapakis, I., Moshfeghi, Y., Joho, H., Ren, R., Hannah, D. and Jose, J.M. 2009. Integrating facial expressions into user profiling for the improvement of a multimodal recommender system. *IEEE International Conference on Multimedia and Expo (ICME)* (2009), 1440–1443.

[3] Dahmane, M. and Meunier, J. 2011. Continuous emotion recognition using Gabor energy filters. *Affective Computing and Intelligent Interaction* (2011), 351–358.

[4] Deriso, D., Susskind, J., Krieger, L. and Bartlett, M. 2012. Emotion Mirror: A Novel Intervention for Autism Based on Real-Time Expression Recognition. *Computer Vision – ECCV Workshops and Demonstrations*. A. Fusiello, V. Murino, and R. Cucchiara, eds. Springer. 671–674.

[5] Desurvire, H., Caplan, M. and Toth, J. 2004. Using heuristics to evaluate the playability of games. *Human factors and computing systems (CHI)* (Apr. 2004), 1509–1512.

[6] Edwards, G.J., Taylor, C.J. and Cootes, T.F. 1998. Interpreting face images using active appearance models. *IEEE International Conference on Automatic Face and Gesture Recognition* (1998), 300–305.

[7] Ekman, P. 1999. Basic Emotions. *Handbook of Cognition and Emotion*. John Wiley & Sons.

[8] Ekman, P., Friesen, W. V. and Hager, J.C. 1978. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press.

[9] EverQuest 2 Receives Facial Tracking Upgrade: 2012. *http://www.tomshardware.com/news/SOEmote-EverQuest-SOE-MMORPG-Live-Driver,16701.html*. Accessed: 2013-07-31.

[10] EverQuest II: SOEmote Released: *http://www.mmorpg.com/discussion2.cfm/thread/359738/*. Accessed: 2013-07-31.

[11] Iacobini, M., Gonsalves, T., Bianchi-Berthouze, N. and Frith, C. 2010. Emotional contagion in interactive art. *Kansei Engineering and Emotion Research* (2010), 1975–1984.

[12] Kinect for Xbox 360: *http://www.xbox.com/Kinect*. Accessed: 2013-07-31.

[13] Lang, P.J. 2008. International affective picture system (IAPS): Affective ratings of pictures and instruction manual. *Technical Report A-8, University of Florida, Gainesville* (2008).

[14] Lang, P.J. 1995. The emotion probe: Studies of motivation and attention. *American psychologist*. 50, 5 (1995), 372–385.

[15] Littlewort, G. and Fasel, I. 2002. Fully automatic coding of basic expressions from video. *INC MPLab Technical Report*. (2002), 6.

[16] Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z. and Matthews, I. 2010. The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. *IEEE Computer Vision and Pattern Recognition - Workshops* (2010), 94–101.

[17] MacQueen, J.B. 1967. Some Methods for Classification and Analysis of MultiVariate Observations. *Berkeley Symposium on Mathematical Statistics and Probability* (1967), 281–297.

[18] Manjunath, B.S. and Ma, W.Y. 1996. Texture features for browsing and retrieval of image data. *IEEE Trans on Pattern Analysis and Machine Intelligence*. 18, 8 (1996), 837–842.

[19] Mourão, A., Borges, P., Correia, N. and Magalhães, J. 2013. Facial Expression Recognition by Sparse Reconstruction with Robust Features. *Image Analysis and Recognition* (2013), 107–115.

[20] Orvalho, V., Miranda, J. and Sousa, A. 2009. What a Feeling: Learning Facial Expressions and Emotions. *Videojogos Conference* (2009), 1–11.

[21] Paiva, A., Costa, M., Chaves, R., Piedade, M., Mourão, D., Sobral, D., Höök, K., Andersson, G. and Bullock, A. 2003. SenToy: an affective sympathetic interface. *International Journal of Human-Computer Studies*. 59, 1-2 (2003), 227–235.

[22] Pantic, M., Sebe, N., Cohn, J.F. and Huang, T. 2005. Affective multimodal human-computer interaction. *ACM Multimedia* (2005), 669–676.

[23] Savva, N., Scarinzi, A. and Bianchi-Berthouze, N. 2012. Continuous Recognition of Player's Affective Body Expression as Dynamic Quality of Aesthetic Experience. *IEEE Trans on Computational Intelligence and AI in Games*. 4, 3 (2012), 199–212.

[24] SOEmote Guide: 2012. *http://soemote.com/guide/index.vm*. Accessed: 2013-07-31.

[25] Stoiber, N., Aubault, O., Seguier, R. and Breton, G. 2010. The mimic game: real-time recognition and imitation of emotional facial expressions. *ACM SIGGRAPH Talks*. ACM.

[26] Tavares, G., Mourão, A. and Magalhães, J. 2013. Crowdsourcing for affective-interaction in computer games. *Workshop on Crowdsourcing for Multimedia at ACM MM* (2013).

[27] The Daily Grind: Do you really want MMO innovation?: *http://massively.joystiq.com/2012/08/16/the-daily-grind-do-you-really-want-mmo-innovation/*. Accessed: 2013-07-31.

[28] Tian, Y.L., Kanade, T. and Cohn, J.F. 2005. Facial expression analysis. *Handbook of face recognition*. S.Z. Li and A.K. Jain, eds. Springer. 247–275.

[29] Valstar, M.F., Mehu, M., Jiang, B., Pantic, M. and Scherer, K. 2012. Meta-Analysis of the First Facial Expression Recognition Challenge. *IEEE Trans on Systems, Man, and Cybernetics, Part B: Cybernetics*. 42, 4 (2012), 966 – 979.

[30] Viola, P. and Jones, M.J. 2004. Robust Real-Time Face Detection. *International Journal of Computer Vision*. 57, 2 (2004), 137–154.

[31] Wang, N. and Marsella, S. 2006. Introducing EVG: An emotion evoking game. *Intelligent Virtual Agents*. 4133, (2006), 282–291.

[32] Yeasin, M., Bullot, B. and Sharma, R. 2006. Recognition of facial expressions and measurement of levels of interest from video. *IEEE Trans on Multimedia*. 8, 3 (2006), 500–508.