Facial Affect Recognition for Cognitive-behavioural Therapy

Maher Ben Moussa[1], Nadia Magnenat-Thalmann[1], Dimitri Konstantas[2]
[1]*MIRALab* / [2]*Institute of Services Science, University of Geneva, Route de Drize 7, Carouge, Switzerland*

Juan J. Santamaría[3], Fernando Fernández–Aranda[3], Susana Jiménez-Murcia[3]
[3]*Department of Psychiatry, University Hospital of Bellvitge-IDIBELL, and Ciber Fisisologia Obesidad y Nutricion (CIBEROBN), Instituto Salud Carlos III, Barcelona, Spain*

Keywords: Cognitive-behavioural Therapy, Facial Affect Recognition, Affective Computing, Health.

Abstract: In this paper we present facial affect recognition specifically designed and developed for clinical use in Cognitive-Behavioural Therapy. We describe our hybrid approach for facial affect recognition that combines geometric and appearance based features of the face. For geometric features we have developed a robust Active Shape Model based facial point tracking system for real-time use in clinical environments. Gabor wavelets are employed to extract the appearance-based features from appearance changes of the face skin. Further, we present an evaluation of our approach and we describe the systems integration with serious games that are being used for mental disorder therapy.

1 INTRODUCTION

Cognitive-Behavioural Therapy (CBT) is the evidence-based treatment of choice of several mental disorders and addictive behaviours (Fernandez-Aranda et al, 2008) (Jimenez-Murcia et al, 2007). However, there is a lack of effective strategies and adequate psychotherapy tools to remediate some shared dysfunctional emotional regulatory processes and disinhibited personality traits in impulse related disorders (Alvarez-Moya et al, 2009) (Fernandez-Aranda et al, 2006). This had led us to consider new approaches based on serious games for patient treatment. In the scope of the EU initiative Playmancer a videogame prototype (Fig 1) for treating specific mental disorders (namely eating disorders and impulse control disorders) has been developed by the company Serious Games Interactive. It is being applied at the Department of Psychiatry (University Hospital of Bellvitge, Barcelona, Spain) for mental disorders (mainly eating disorders and behavioural addictions) and introduces the player to an interactive scenario (named Islands), where the final goal is to increase emotional self-control skills in patients and control over their general impulsive behaviours.

Figure 1: Playmancer video game.

Next to the keyboard and mouse, which are the regular game control components, the videogame is being influenced by physiological signals and affect recognition components. The physiological signals are measured by a stationary measurement system from MobiHealth B.V that measures and processes the signals: galvanic skin, oxygen saturation, heart rate and heart rate variation, temperature, breathing frequency signal processing. The patient's affect can be measured by the facial, the speech-based or the physiological affect recognition component. These three components can also be combined to provide a single fused output to the therapeutic video game. From these three components, this paper covers

Magnenat-Thalmann N., Santamaría J., Moussa M., Konstantas D., Fernández–Aranda F. and Jiménez-Murcia S. (2012).
Facial Affect Recognition for Cognitive-behavioural Therapy.
In *Proceedings of the Sixth International Symposium on e-Health Services and Technologies and the Third International Conference on Green IT Solutions*, pages 64-68
DOI: 10.5220/0004474200640068

mainly the facial affect recognition component and its integration into the Playmancer system. In the next section, we will begin with an explanation of the related concepts and previous work that has been undertaken in the area of facial affect recognition. In Section 3 we present our implementation. In section 4, we evaluate our system. Section 5 describes the clinical application of the system. Finally, we include our paper and discuss the limitations and future work in section 6.

2 BACKGROUND

Facial affect recognition usually involves several steps. Firstly, the faces are detected and extracted from static images or video input, followed by the extraction of the features from these faces. The resulting facial features are analysed and used to classify the affective state of a person. For facial extraction there are several notable approaches. However, (Viola and Jones, 2004) is the most commonly used facial detection system. For facial feature extraction we can distinguish between two types of facial features: *geometric* features and *appearance*-based features. Geometric features represent the shape of the facial components (such as mouth, nose, eyes, etc) and the locations of these facial points (corners of mouth, nose, eyes, etc) (Cohen et al, 2003). Appearance-based features represent the appearance changes of the face skin (wrinkles, furrows, etc), such as in (Bartlett, 2003) where Gabor-wavelets are employed to extract the features. There are different approaches of geometric feature extraction. Some approaches employ variations of the optical flow algorithm (Tian et al 2005), some use observation models (Patras and Pantic, 2005) and others employ Active Shape Models (ASM) or Active Appearance Models (AAM) (Cootes and Taylor, 2004)(Matthews and Baker, 2004)(Stegmann, 2000). Different machine learning techniques have been employed to classify facial actions and human affect on static frames as well on the temporal features of facial motion. This includes Bayesian Networks (Cohen et al, 2003). Support Vector Machines (SVM) (Ford, 2002), Hidden Markov Models (HMM) (Bartlett, 2003) and even rule-based classifiers (Cohn et al, 2004).

3 SYSTEM OVERVIEW

Recent research involving eating disorders patients and pathological gamblers (Vanderlinden et al, 2004) (Ledgerwood, 2006) showed that anger is one of the emotions most commonly described as "triggering" factors of linked impulsive behaviour's (binging, going out to gamble, etc.), while understanding and attaining a calm disposition is often a clinical goal for emotion regulation.

For the Playmancer video game, it was assumed that three basic emotions (one positive, one negative and one neutral) should be sufficient for the target population at this point (Fernandez-Aranda et al, 2012). The choice has been made to recognize the emotions joy, anger and neutral. Based on this assumption, a facial affect recognition system has been developed that integrates geometric and appearance-based facial feature extraction and employs SVMs for classification of emotions.

3.1 Geometrical Feature Extraction

Given video input, our facial tracker detects and tracks the facial points over time. We employ the Active Shape Model (ASM) to extract the shape information of the face in a video sequence. ASM, which was introduced by (Cootes and Taylor, 2004), is an algorithm that matches a set of shape points to an image constrained by the statistical model of the shape. A shape is represented as $s = (x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n)$ where n is the number of points. The shape's statistical model allows linear shape variation. This means that a shape s can be expressed as a base shape s_0 plus a linear combination of N shape vectors s_i:

$$s = s_0 + \sum_{i=1}^{N} p_i s_i$$

where the coefficients $p = (p_1, \dots, p_n)^T$ are the shape parameters and s_i are orthonormal eigenvectors. Furthermore, on each shape Procrustes alignment (Cootes and Taylor, 2004) is employed to estimate the base shape s_0. This algorithm removes rigid body distortions such as translation, rotation and scaling.

The detection and the tracking using ASM involves several steps. In the initialisation stage, a Viola & Jones based face detection (Viola and Jones, 2004) and feature detection (eyes, nose and mouth) (Castrillón et al, 2007) is performed on the video frame. The ASM shape is aligned with respect to the position of the detected facial features. After the alignment, the shape is fitted iteratively to the face, where for every model point the best match (by taking gradients into consideration) within a small image patch around the location of the corresponding point is searched for with respect to

the shape parameter constraints. Our approach is based on (Jia 2010) where instead of using one gradient vector along one direction, two gradient vectors along two orthogonal directions are used. This means more flexibility of the shape. During the tracking, in each new frame the shape of the previous frame is used for the fitting. In each frame, the eyes and the nose are also tracked using the template matching technique to ensure the accuracy of the shape fitting. In case of lower accuracy, the ASM shape is automatically reinitialised.

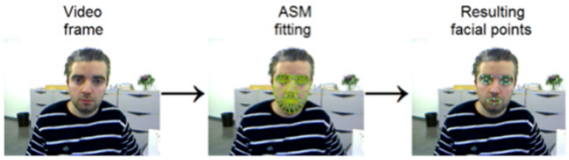

Figure 2: Geometrical points detection and tracking.

To improve the results of the tracking in the clinical experiment, we have compiled our own database to train the ASM algorithm. Our database consists of images we recorded in the clinical experiment room with its usual lighting condition (Kostoulas et al, 2010).

From the resulting ASM, 17 facial points are extracted as shown in Figure 2. Using these facial points, the feature vector $\vec{v}$ of 9 elements is composed as shown in Table 1. We have chosen to use only 9 features because we have mainly selected those features most visible and most relevant to the chosen emotions according to Ekman's FACS definitions. Further all the features are relative to the calibration frame where the face should be in a neutral position.

Table 1: Feature vector.

v_1	$\text{LeftEyeInnerCorner}_y - \text{LeftEyesbrowCenter}_y$
v_2	$\text{LeftEyeInnerCorner}_y - \text{LeftEyesbrowInnerCorner}_y$
v_3	$\text{RightEyeInnerCorner}_y - \text{RightEyesbrowCenter}_y$
v_4	$\text{RightEyeInnerCorner}_y - \text{RightEyesbrowInnerCorner}_y$
v_5	$\text{LeftEyeBottom}_y - \text{LeftEyeTop}_y$
v_6	$\text{RightEyeBottom}_y - \text{RightEyeTop}_y$
v_7	$\text{MouthBottom}_y - \text{MouthTop}_y$
v_8	$\text{MouthCenter}_y - \text{MouthTop}_y$
v_9	$\text{MouthRight}_x - \text{MouthLeft}_x$

3.2 Appearance-based Feature Extraction

Although different appearance-based features can be extracted, we have chosen to mainly use the brow furrow because it is one of the most relevant features when distinguishing between positive and negative emotions.

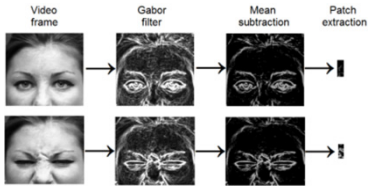

Figure 3: Gabor filter based feature extraction.

In our approach we combine 4 orientations (0, 2, 4, and 6) of Gabor wavelets that we employ on the upper face image patch as shown in Figure 3. The patch is subtracted by its own pixels mean and a brow furrow patch is extracted from that image. The *L2 norm* of the brow furrow patch then becomes element v_{10} of feature vector $\vec{v}$.

3.3 Training and Classification

Using features vector $\vec{v}$, support vector machines (SVM) are trained to recognize emotions. Because our recorded database (Kostoulas et al, 2010) does not contain posed emotions, we have chosen to train our system using the Cohn-Kanade database (Lucey et al, 2010). Using the ASM-annotations that are available for the Cohn-Kanade database, we have trained our SVMs for affect recognition. This means that our training data is not dependent on the accuracy of the ASM tracker on a certain database and therefore it should perform well on other databases. Other emotions (*sadness*, *fear*, *surprise*) have been tagged *neural*.

Furthermore, although there are different studies where temporal data is being used for affect classification, we have chosen to focus on frame-based classification instead of temporal. Temporal classification requires the detection of the different stages of a facial expression (onset, apex and offset) and would only work well on a database. In a real life scenario facial tracking fails sometimes in case of quick movements or head rotations. However, we still consider the temporal aspect by relating the accuracy of the detection of an emotion to its occurrence in a window of time.

4 TECHNICAL EVALUATION

We have tested our system using the Cohn-Kanade database. According the database manual, we have identified 42 samples of *anger*, 65 samples of *joy* and 38 samples of *neutral*.

Figure 4: ASM with Cohn-Kanade database.

Figure 4 shows the tracking results on Cohn-Kanade database images and Table 2 shows the tracking confusion matrix of our software on this database. The table shows that *neutral* is well detected. *Anger* and *Joy* are detected 73.81% / 78.46% of the times correctly. The results are satisfying. However, it has to be said that we expect to have higher accuracy in the clinical setup because of the better calibration stage and because of the same quality of videos (Cohn-Kanade database consists of images with different qualities). Furthermore, fusion with other affect recognition modalities will also increase the accuracy.

Table 2: The confusion matrix for the affect recognition.

	Anger	Joy	Neutral
Anger	**73.81**	11.9	14.28
Joy	4.62	**78.46**	16.92
Neutral	0	10.53	**89.47**

5 CLINICAL SETUP

Controlled prospective longitudinal clinical studies using the software have been going on for some time (Fernandez-Aranda et al, 2012) (Jimenez-Murcia et al, 2009). The Playmancer video game is being used as an additional therapeutic tool (combined with usual therapy), with consecutively admitted outpatients. Up to now, over 30 impulsive spectrum disorder patients (eating disorders and pathological gamblers) have been recruited. Besides usability scores of over 85% in patients, promising results are being obtained in case-control studies. Upcoming medical papers will explain these results in more detail.

The procedure goes as follows: During the initial session, the therapist explains the rationale behind the study and the affect recognition to the patient during 20 minutes. At the beginning of each session, the affect recognition components are calibrated

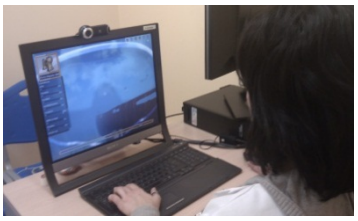

Figure 5: Subject playing the therapeutic game.

with the patients. For the facial affect recognition component a *neutral*, *joy* and *anger* expression is recorded and used later on for recognition. By knowing the patient specific maximum and minimum value of the features (especially the appearance-based feature), the affect recognition rate can be improved significantly.

6 CONCLUSIONS & DISCUSSION

In this paper we have described our approach for affect recognition intended for a Cognitive-Behavioural Therapy. We developed an affect recognition component that employs active shape models to extract the geometric features. We have chosen to add one appearance-based feature to improve the recognition rate. Support Vector Machines (SVM) have been employed for affect recognition and have been trained and evaluated with the Cohn-Kanade database. Further experiments fusing different affect recognition components are being conducted to improve the results. Clinical experiments using the software are being conducted the results of which will be part of future publications.

ACKNOWLEDGEMENTS

The research has been funded by the European research projects PlayMancer (FP7-ICT-215839-2007), 3DLife NoE (IST-FP7 247688) and the Swiss National Science Foundation project AerialCrowds (CRSI20-122696).

REFERENCES

Moore, R., Lopes, J., 1999. Paper templates. In *TEMPLATE'06, 1st International Conference on Template Production*. SciTePress.

Smith, J., 1998. *The book*, The publishing company. London, 2nd edition.

Fernandez-Aranda, F., Nunez, A. et al. 2008. Internet-Based Cognitive-Behavioral Therapy for Bulimia Nervosa: A Controlled Study. *Cyberpsychol Behav.* 12.

Jimenez-Murcia S, A.-M. E., Granero R, et al.,2007, Cognitive-behavioral group treatment for pathological gambling: analysis of effectiveness and predictors of therapy outcome. *Psychotherapy Research*, 17(5), 544 – 552.

Alvarez-Moya, E. M., Jimenez-Murcia, S., Moragas, L., Gomez-Pena, M., Aymami, M. N., Ochoa, C., et al., 2009, Executive functioning among female pathological gambling and bulimia nervosa patients: preliminary findings. *J Int Neuropsychol Soc*, 15(2), 302-306.

Fernandez-Aranda, F., Jimenez-Murcia, S., Alvarez-Moya, E. M., Granero, R., Vallejo, J., & Bulik, C. M., 2006, Impulse control disorders in eating disorders: clinical and therapeutic implications. *Compr Psychiatry*, 47(6), 482-488.

Viola, P. and Jones, M. J., 2004, Robust Real-Time Face Detection, *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137-154

Cohen, I., Sebe, N., Garg, A., Chen, L. S., and Huanga, T. S., 2003, Facial expression recognition from video sequences: temporal and static modeling, *Computer Vision and Image Understanding*, vol. 91, no. 1-2, pp. 160-187

Bartlett, M. S., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., and Movellan, J., 2006, Fully Automatic Facial Action Recognition in Spontaneous Behavior, *in 7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, pp. 223-230

Tian, Y.-li, Kanade, T., and Cohn, J. F., 2005, Facial Expression Analysis, *Handbook of Face Recognition*, vol. 3, no. 5, pp. 247-275

Patras, I. abd Pantic, M., 2005, Tracking deformable motion, *International Conference on Systems, Man and Cybernetics,* IEEE, vol.2, no., pp.1066-1071

Cootes, T. and Taylor, C., 2004, Statistical models of appearance for computer vision, *Technical report, Imaging Science and Biomedical Engineering*, University of Manchester

Matthews, I. and Baker, S., 2004, Active appearance models revisited, *International Journal of Computer Vision*, 60(2):135–164

Stegmann, M. B., 2000, Active Appearance Models: Theory, Extensions and Cases

Ford, G., 2002, Fully automatic coding of basic expressions from video, *Technical Report INC-MPLab-TR-2002.03*, Machine Perception Lab, Institute for Neural Computation, University of California, San Diego

Cohn, J. F., Reed, L. I., Ambadar, Z., and Moriyama, T., 2004, Automatic analysis and recognition of brow actions and head motion in spontaneous facial behavior, *in IEEE International Conference on Systems, Man and Cybernetics* (IEEE Cat. No.04CH37583), vol. 1, pp. 610-616.

Vanderlinden, J., Dalle Grave, R., Fernandez, F., Vandereycken, W., Pieters, G., & Noorduin, C., 2004. Which factors do provoke binge eating? An exploratory study in eating disorder patients. *Eat Weight Disord*, 9(4), 300-305

Ledgerwood, D. M., & Petry, N. M., 2006, Psychological experience of gambling and subtypes of pathological gamblers. *Psychiatry Res,* 144(1), 17-27.

Castrillón, M., Déniz, O., Guerra, C., and Hernández, M., 2007, ENCARA2: Real-time detection of multiple faces at different resolutions in video streams, *Journal of Visual Communication and Image Representation*, vol. 18, no. 2, pp. 130–140

Jia, P., 2010, 2D Statistical Models, *Technical Report of Vision Open Working Group*, 2st Edition, Oct 21

Jia, P., Audio-visual based HMI for an Intelligent Wheelchair. *PhD thesis*, University of Essex, 2010

Kostoulas T. et al., 2010, The PlayMancer Database: A Multimodal Affect Database in Support of Research and Development Activities in Serious Game Environment, *in Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010),* Valetta, Malta

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, *in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, pp. 94-101

Jiménez-Murcia S, Fernández-Aranda F, Kalapanidas E, Konstantas D, Ganchev T, Kocsis O, Lam T, Santamaría JJ, Raguin T, Breiteneder C, Kaufmann H, Davarakis C., 2009, Playmancer project: a serious videogame as an additional therapy tool for eating and impulse control disorders. *Stud Health Technol Inform*,144:163-6.

Fernández-Aranda F, Jiménez-Murcia S, Santamaría JJ, Gunnard K, Soto A, Kalapanidas E, Bults R, Davarakis C, Ganchev T, Granero R, Konstantas D, Kostoulas T, Lam T, Lucas M, Masuet-aumatell C, Ben Moussa M, Nielsen J & Penelo E., 2012,Video games as a complementary therapy tool in mental disorders: PlayMancer, a European multicentre study, *Journal of Mental Health*