Recognizing Emotions from Video in a Continuous 2D Space

Sergio Ballano[1], Isabelle Hupont[1], Eva Cerezo[2] and Sandra Baldassarri[2]

[1] Aragon Institute of Technology, Department of R&D and Technology Services,
Zaragoza. 5018, María de Luna 7-8, Spain
[2] Universidad de Zaragoza, Computer Science and Systems Engineering Department,
Zaragoza. 50018, María de Luna 3, Spain
{sballano, ihupont}@ita.es, {ecerezo, sandra}@unizar.es

Abstract. This paper proposes an effective system for continuous facial affect recognition from videos. The system operates in a continuous 2D emotional space, characterized by evaluation and activation factors. It makes use, for each video frame, of a classification method able to output the exact location (2D point coordinates) of a still facial image in that space. It also exploits the Kalman filtering technique to control the 2D point movement along the affective space over time and to improve the robustness of the method by predicting its future locations in cases of temporal facial occlusions or inaccurate tracking.

Keywords: Affective computing, facial expression analysis.

1 Introduction

Facial expressions are often evaluated by classifying still face images into one of the six universal "basic" emotions proposed by Ekman [1] which include "happiness", "sadness", "fear", "anger", "disgust" and "surprise". This categorical approach fails to describe the wide range of emotions that occur in daily communication settings and ignores the intensity of emotions.

Given that humans inherently display facial emotions following a continuous temporal pattern [2], more recently attention has been shifted towards sensing facial affect from video sequences. The study of facial expressions' dynamics reinforces the limitations of categorical approach, since it represents a discrete list of emotions with no real link between them and has no algebra: every emotion must be studied and recognized independently.

This paper proposes a method for continuous facial affect recognition from videos. The system operates in a 2D emotional space, characterized by evaluation and activation factors. It combines a classification method able to output, frame per frame, the exact location (2D point coordinates) of the shown facial image and a Kalman filtering technique that controls the 2D point movement over time through an "emotional kinematics" model. In that way, the system works with a wide range of intermediary affective states and is able to define a continuous emotional path that characterizes the affective video sequence.

2 Facial Images Classification in a Continuous 2D Affective Space

The starting point of the system is the method for facial emotional classification presented in authors' previous work [3]. The inputs to this method are the variations with respect to the "neutral" face of the set of facial distances and angles shown in Fig. 1. This initial method combines through a majority voting strategy [3] the five most commonly used classifiers in the literature (Multilayer Perceptron, RIPPER, SVM, Naïve Bayes and C4.5) to finally assign at its output a confidence value CV(E_i) of the facial expression to each of Ekman's six emotions plus "neutral".

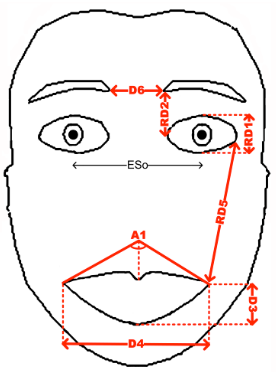

Fig. 1. System's facial inputs.

To enrich the emotional output information from the system in terms of intermediate emotions, the evaluation-activation 2D model proposed by Whissell has been used. In her study, Whissell assigns a pair of values <evaluation, activation> to each of the approximately 9000 affective words that make up her "Dictionary of Affect in Language" [4]. The next step is to build an emotional mapping so that an expressional face image can be represented as a point on this plane whose coordinates (x,y) characterize the emotion property of that face.

The words corresponding to each of Ekman's six emotions have a specific location (x_i, y_i) in the Whissell space. Thanks to this, the output of the classifiers (confidence value of the facial expression to each emotional category) can be mapped onto the space. This emotional mapping is carried out considering each of Ekman's six basic emotions plus "neutral" as weighted points in the evaluation-activation space. The weights are assigned depending on the confidence value CV(E_i) obtained for each emotion. The final coordinates (x,y) of a given image are then calculated as the centre of mass of the seven weighted points.

3 From Still Images to Video Sequences

Thanks to the use of the 2-dimensional description of affect, which supports continuous emotional input, an emotional facial video sequence can be viewed as a point (corresponding to the location of a particular affective state in time *t*) moving through this space over time. In that way, the different positions taken by the point (one per frame) and its velocity over time can be related mathematically and modeled, finally obtaining an "emotional path" in the 2D space that reflects intuitively the emotional progress of the user throughout the video.

3.1 Modeling Emotional Kinematics with a Simple Kalman Filter

For real-time "emotional kinematics" control, the Kalman filter is exploited. Analogously to classical mechanics, the "emotional kinematics" of the point in the Whissell space (x-position, y-position, x-velocity and y-velocity) is modeled as the system's state in the Kalman framework at time t_k. In this way, the Kalman iterative estimation process -that follows the well-known recursive equations detailed in Kalman's work [5]- can be applied to the recorded user's emotional video sequence, so that each iteration corresponds to a new video frame (i.e. to a new sample of the computed emotional path). One of the main advantages of using Kalman filter for the 2D point emotional trajectory modeling is that it can be used to tolerate small occlusions or inaccurate tracking so that, when a low level of confidence in the facial tracking is detected, the measurement will not be used and only the filter prediction will be taken as the 2D point position.

3.2 Experimental Results

In order to demonstrate the potential of the proposed "emotional kinematics" model, it has been tested with a set of complex video sequences recorded in an unsupervised setting (VGA webcam quality, different emotions displayed contiguously, facial occlusions, etc.). A total of 15 videos from 3 different users were tested, ranging from 20 to 70 seconds, from which a total of 127 key-frames were extracted to evaluate different key-points of the emotional path.

These key-points were annotated in the Whissell space thanks to 18 volunteers. The collected evaluation data have been used to define a region where each image is considered to be correctly located. The algorithm used to compute the shape of the region is based on Minimum Volume Ellipsoids (MVE) and follows the algorithm described by Kumar and Yildrim [6]. MVE looks for the ellipsoid with the smallest volume that covers a set of data points. The obtained MVEs are used for evaluating results at four different levels, as shown in Table 1. As can be seen, the success rate is 61.90% in the most restrictive case, i.e. with ellipse criteria and rises to 84.92% when considering the activation axis criteria. Finally, Fig. 2 shows an example of emotional path obtained after applying the "emotional kinematics" model.

Table 1. Results obtained in an uncontrolled environment.

	Ellipse criteria (success if inside the ellipse)	Quadrant criteria (success if in the same quadrant as the ellipse centre)	Evaluation axis criteria (success if in the same evaluation semi-axis as the ellipse centre)	Activation axis criteria (success if in the same activation semi-axis as the ellipse centre)
Success%	61.90%	74.60%	79.37%	84.92%

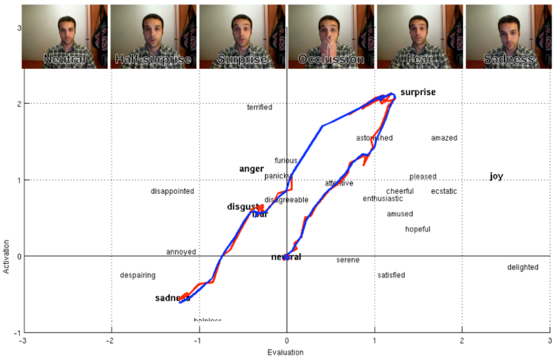

Fig. 2. "Emotional kinematics" model response during the different affective phases of the video and the occlusion period. In dashed red, emotional trajectory without Kalman filtering; In solid blue, reconstructed emotional trajectory using Kalman filter.

4 Conclusions

This paper describes an effective system for continuous facial affect recognition from videos. The main distinguishing feature of our work compared to others is that the output does not simply provide a classification in terms of a set of emotionally discrete labels, but goes further by extending the emotional information over an infinite range of intermediate emotions and by allowing a continuous dynamic emotional trajectory to be detected from complex affective video sequences.

References

1. Keltner, D., Ekman, P.: Facial Expression Of Emotion. Handbook of emotions. págs. 236-249. New York: Guilford Publications, Inc. (2000).

2. Petridis, S., Gunes, H., Kaltwang, S., Pantic, M.: Static vs. dynamic modeling of human nonverbal behavior from multiple cues and modalities. Proceedings of the 2009 international conference on Multimodal interfaces. 23-30 (2009).

3. Hupont, I., Cerezo, E., Baldassarri, S.: Sensing facial emotions in a continuous 2D affective space. Presented at the Systems Man and Cybernetics (SMC), Istanbul Octubre 10 (2010).

4. Whissell, C.M.: The Dictionary of Affect in Language, Emotion: Theory, Research and Experience. New York Academic (1989).

5. Kalman, R.: A New Approach to Linear Filtering and Prediction Problems. Transactions of the ASME – Journal of Basic Engineering. 35-45 (1960).

6. Kumar, P., Yildirim, E.A.: Minimum-Volume Enclosing Ellipsoids and Core Sets. Journal of Optimization Theory and applications. 126, 1-21 (2005).