

REAL TIME FACIAL EXPRESSION RECOGNITION FROM IMAGE SEQUENCES USING SUPPORT VECTOR MACHINES

I. Kotsia and I. Pitas

Aristotle University of Thessaloniki
Department of Informatics
Box 451
54124 Thessaloniki, Greece

ABSTRACT

In this paper, a novel real-time method is proposed as a solution to the problem of facial expression classification in image sequences. The user manually places some of the Candide grid's points to the face depicted at the first frame. The grid adaptation system, based on deformable models, tracks the entire Candide grid as the facial expression evolves through time, thus producing a grid that corresponds to the greatest intensity of the facial expression, as shown at the last frame. Certain points that are involved into creating the Facial Action Units movements are selected. Their geometrical displacement information, defined as the coordinates' difference between the last and the first frame, is extracted to be the input to a six class Support Vector Machine system. The output of the system is the facial expression recognized. The proposed novel real-time system, recognizes the 6 basic facial expressions with an approximately 98% accuracy.

1. INTRODUCTION

Several research efforts have been done regarding facial expression recognition during the past two decades, due to its importance for human centered interfaces. The facial expressions under examination were defined as a set of six basic facial expressions (anger, disgust, fear, happiness, sadness and surprise), whose combinations produce every other "complex" facial expression [1]. In order to make the recognition procedure more standardized, a set of muscle movements (known as Action Units) that produce each facial expression, was created by psychologists, thus forming the so called *Facial Action Coding System (FACS)* [2].

A survey on automatic facial expression recognition can be found in [3]. Gabor wavelets have been thoroughly used in facial expression recognition. They were used as a pre-processing step in order to create the input for multi-layer

neural networks [4]. Gabor wavelets were also combined with elastic graph matching techniques [5]. Facial expression recognition has also been investigated using Tree-Augmented-Naive Bayes (TAN) trees in [6].

In the current paper, a novel real-time method for recognizing facial expressions using Support Vector Machines (SVM), is proposed. The system is semi-automatic, in the sense that the user has to manually place in the beginning some of the Candide grid's points to a face depicted at the first frame of the image sequence under examination. The tracking system follows the facial expression evolving through time to reach its highest intensity, producing at the same time the grid that corresponds to it. A subset of the Candide grid's point is chosen, as the one that is responsible for the formation of movement as described by the Facial Action Units. The geometrical displacement of those points, defined as the difference of each point's coordinates between the first and the last frame of the image sequence, are used as an input to a multiclass SVM system. Each classification class represents one of the 6 basic facial expressions. The experiments were performed using the Cohn-Kanade database and the results show that the above mentioned novel real-time system can achieve an accuracy of 97.75% when recognizing 6 basic facial expressions.

2. SYSTEM DESCRIPTION

The diagram of the system used for the experiments is shown in Figure 1. The system is composed of two subsystems, one for geometrical information extraction and one for geometrical information classification.

Facial expressions can be described as combinations of Facial Action Units (FAUs), as proposed by [7]. As can be seen from the rules proposed, the FAUs that are necessary for fully describing all facial expressions are the FAUs 1, 2, 4, 5, 6, 7, 9, 10, 12, 15, 16, 17, 20, 23, 24, 25 and 26. Therefore, these 17 FAUs are responsible for creating movement according to the Facial Action Coding System (FACS).

This work has been supported by the FP6 European Union Network of Excellence MUSCLE "Multimedia Understanding Through Semantic Computation and Learning" (FP6-507752).

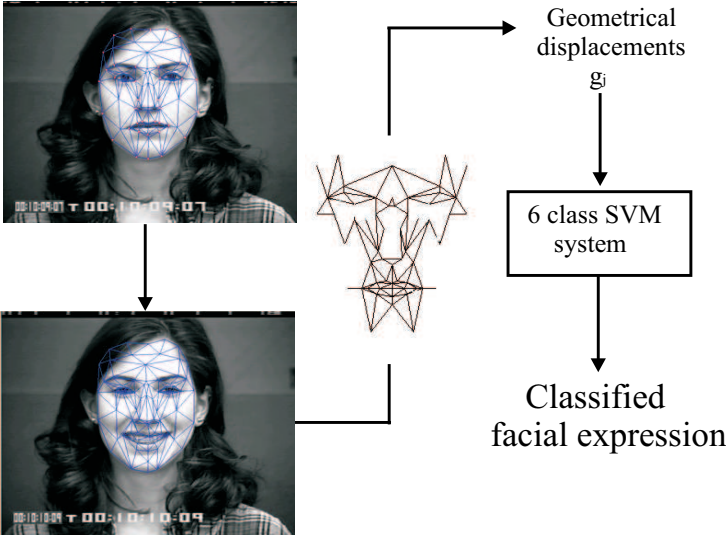


Fig. 1. System description

2.1. Geometrical displacement information extraction

The geometrical information extraction is done by a grid adaptation system, based on deformable models. The Candidate grid is randomly initialized on the first frame of the image sequence, being in its neutral state. The user has to manually place some of its points to the face depicted. The points around the eyes, eyebrows and mouth are the ones with the greatest importance, since they are the ones responsible for the formation of movement according to FACS. The software automatically adjusts the grid to the face and then tracks it through the image sequence, following the facial expression evolving through time [9]. At the end, the grid adaptation software produces the deformed Candidate grid that corresponds to the facial expression appearing at the image sequence.

The deformed Candide grid produced by the grid adaptation software, that corresponds to the greatest intensity of the facial expression shown, is constructed by 104 points. Not all of these points are important for the recognition of the facial expression, for example the contour of the face doesn't affect the way the eyes or the mouth points move. Thus, a subset of 62 points are chosen, as those that control the movement described by the 17 FAUs used for describing facial expressions. The grid that is composed of these points can be seen in figure 1.

The classification is performed based only in geometrical information, without taking into consideration any luminance or color information. The geometrical displacement information of the subset's points coordinates is extracted to be used for facial expression classification.

Let \mathcal{U} the database that contains the geometrical displacement information separated into the 6 different classes, $\mathcal{U}_k (k \in \{1, \dots, 6\})$, each one representing one of six basic facial expressions (anger, disgust, fear, happiness, sadness and surprise).

The geometrical information used is the displacement of one point \mathbf{d}_j^i , defined as the difference between the last and the first frame's coordinates

$$\mathbf{d}_j^i = \begin{bmatrix} \Delta x_j^i \\ \Delta y_j^i \end{bmatrix}, \quad i \in \{1, \dots, K\} \quad \text{and} \quad j \in \{1, \dots, N\} \quad (1)$$

where i is the number of points taken under consideration, here K , equal to 62 and j is the the number of image sequences to be examined, here N , equal to 222.

In that way, for every image sequence to be examined, a feature vector \mathbf{g}_j that belongs to one of the six facial expression classes \mathcal{U}_k is constructed, containing the geometrical displacement of every point taken into consideration, thus having the following form

$$\mathbf{g}_j = \begin{bmatrix} \mathbf{d}_j^1 \\ \mathbf{d}_j^2 \\ \vdots \\ \mathbf{d}_j^K \end{bmatrix}. \quad (2)$$

where the vector \mathbf{g}_j has $F = 62 \cdot 2 = 124$ dimensions.

2.2. Geometrical displacement information classification

The feature vector $\mathbf{g}_j \in \mathcal{R}^F$ is used as an input to a multi class Support Vector Machine system. Six classes were considered for the experiments, each one representing one of the basic facial expressions (anger, disgust, fear, happiness, sadness and surprise). The SVM system, classifies each set of the grid's geometrical displacements to one of the six basic facial expressions.

More specifically, as an input for the SVM system, the feature vector \mathbf{g}_j is used, labelled properly with the true corresponding facial expression. The output of the SVM system is a label that classifies the grid under examination to one of the six basic facial expressions.

3. SUPPORT VECTOR MACHINES

We want to classify one test \mathbf{g}_j displacement vector to one of the six facial expressions. This is done using multiclass SVMs [10] that are a generalization of the binary SVM [8].

The SVMs creates a decision function $f(\mathbf{g}_j, \boldsymbol{\alpha})$ which classifies a vector \mathbf{g}_j into one of the six basic facial expression classes. The vector $\boldsymbol{\alpha}$ should be chosen in such a way that for any \mathbf{g}_j the function should be able to provide a classification $l_j \in \{1 \dots 6\}$ (class label).

The main idea of an SVM system is to construct a hyperplane that will separate the desired classes, in such a way that the margin (defined as the distance between the hyperplane and the nearest point) is maximal. Therefore, to generalize, the following equation should be minimized

$$\Phi(\mathbf{w}, \xi) = 1/2 \sum_{m=1}^6 (\mathbf{w}_m^T \cdot \mathbf{w}_m) + c \cdot \sum_{i=1}^N \sum_{m \neq l_i} \xi_i^m \quad (3)$$

with constraints

$$(\mathbf{w}_{l_i}^T \cdot \mathbf{g}_i) + b_{l_i} \geq (\mathbf{w}_m^T \cdot \mathbf{g}_i) + b_m + 2 - \xi_i^m \quad (4)$$

$$\xi_i^m \geq 0, \quad i \in \{1, \dots, N\} \quad m \in \{1, \dots, 6\} \setminus l_i. \quad (5)$$

The decision function that is derived from equation 3 is the following

$$f(\mathbf{g}_j) = \arg \max_n [(\mathbf{w}_n^T \cdot \mathbf{g}_j) + b_n], \quad n \in \{1, \dots, 6\} \quad (6)$$

The solution to this optimization problem in dual variables can be found by the saddle point of the Lagrangian

$$\begin{aligned} L(\mathbf{w}, \mathbf{b}, \xi, \alpha, \beta) = & 1/2 \sum_{m=1}^6 (\mathbf{w}_m^T \cdot \mathbf{w}_m) + c \sum_{i=1}^N \sum_{m=1}^6 \xi_i^m \\ & - \sum_{i=1}^N \sum_{m=1}^6 \alpha_i^m [((\mathbf{w}_i - \mathbf{w}_m)^T \cdot \mathbf{g}_i) + b_{l_i} - b_m - 2 + \xi_i^m] \\ & - \sum_{i=1}^N \sum_{m=1}^6 \beta_i^m \xi_i^m \quad (7) \end{aligned}$$

with the variables

$$\alpha_i^{l_i} = 0, \quad \xi_i^{l_i} = 2, \quad \beta_i^{l_i} = 0, \quad i = \{1, \dots, N\} \quad (8)$$

and constraints

$$\alpha_i^m \geq 0, \quad \beta_i^m = 0, \quad \xi_i^m \geq 0, \quad (9)$$

$$i \in \{1, \dots, N\} \quad m \in \{1, \dots, 6\} \setminus l_i \quad (10)$$

which has to be maximized with respect to α and β and be minimized with respect to \mathbf{w} and ξ .

By further processing [10] equation (6) is finally expressed as

$$f(\mathbf{g}_j, \alpha) = \arg \max_n \left[\sum_{i:l_i=n} A_i (\mathbf{g}_i^T \cdot \mathbf{g}_j) - \sum_{i:l_i \neq n} \alpha_i^n (\mathbf{g}_i^T \cdot \mathbf{g}_j) + b_n \right] \quad (11)$$

where α is the vector of Lagrangian multipliers in equation (7) and A_i is defined as

$$A_i = \sum_{m=1}^6 \alpha_i^m. \quad (12)$$

The previous analysis is used for linear decision surfaces. For the proposed method, nonlinear SVMs were considered. To achieve that, a nonlinear mapping to a high dimensional feature mapping, $Z(\mathbf{g}_j)$ was used. This mapping is defined by a positive kernel function, $k(\mathbf{g}_j^T, \mathbf{g}_j)$, specifying an inner product in the feature space

$$Z(\mathbf{g}_j^T) \cdot Z(\mathbf{g}_j) = k(\mathbf{g}_j^T, \mathbf{g}_j). \quad (13)$$

The kernel used for the experiments was a d degree polynomial function, defined in general as

$$k(\mathbf{g}_j^T, \mathbf{g}_j) = (\mathbf{g}_j^T \cdot \mathbf{g}_j + 1)^d. \quad (14)$$

4. EXPERIMENTAL RESULTS

The database used for the experiments was the Cohn-Kanade database [2], which is encoded into combinations of Action Units. These combinations were translated into facial expressions according to [7]. For each person, the image sequence was created and processed by the grid adaptation system, based on deformable models, to a total of 222 image sequences. In figure 2, a sample of image sequences of one person from the database used for the experiments, is shown.

The database created for the experiments was of limited size, therefore the classifier accuracy was measured using the leave-one out method in order to make maximal use of the available data and produce averaged accuracy results. The classification accuracy was measured as the percentage of the correctly classified facial expressions. The polynomial function used for the creation of the polynomial kernel, was of degree 3.

The experiments indicated that the whole system is fast enough to fulfill a real-time system's requirements, since it is able to classify 20 frames per second. The accuracy achieved was equal to 97,75% when the 6 basic facial expressions were under examination.

The accuracy obtained is averaged over all facial expressions and does not provide any information with respect to a particular expression. The confusion matrix [11] has been computed to handle this problem. It is a $n \times n$ matrix containing the information about the actual class label l_j , $j = 1, \dots, n$ (in its rows) and the label obtained through classification p_j , $j = 1, \dots, n$ ones (in its columns). The diagonal entries of the confusion matrix are the number of facial expressions that are correctly classified, while the off-diagonal entries correspond to misclassification. The confusion matrix showed that the ambiguous facial expression was anger, since it was the only one misclassified as another one of the remaining 5 basic facial expressions. More specifically, anger was mostly misclassified as sadness and then as disgust, as shown from the confusion matrix below.

$lab_{ac} \setminus lab_{clas}$	an	di	fe	ha	sa	su
an	32	0	0	0	0	0
di	1	37	0	0	0	0
fe	0	0	37	0	0	0
ha	0	0	0	37	0	0
sa	4	0	0	0	37	0
su	0	0	0	0	0	37

where an, di, fe, ha, sa and su represent anger, disgust, fear, happiness, sadness and surprise respectively, and lab_{ac} , lab_{clas} represent the actual and the classified label of the video sequence, respectively.

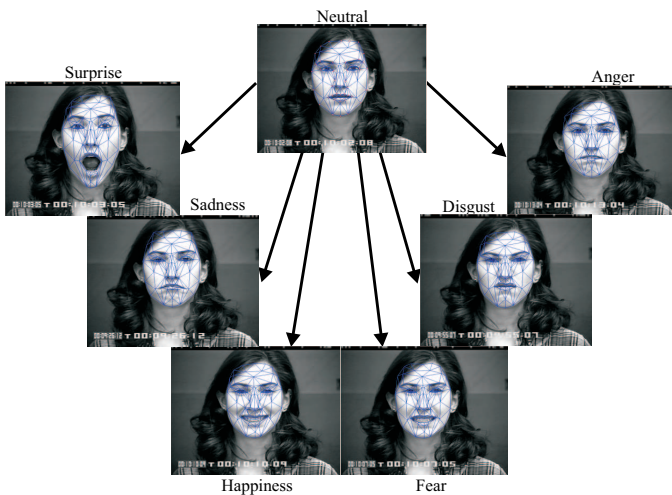


Fig. 2. System description

5. CONCLUSION

Facial expression recognition using Support Vector Machines has been investigated in this paper. The user adjusts the Candide grid to the face depicted at the first frame of the image sequence, by manually placing some of its points. The grid adaptation system, based on deformable models, tracks the grid, as the facial expression progresses through the time, thus producing a deformed grid that corresponds to the highest intensity of the facial expression under examination. A subset of the deformed grid's points is chosen, as those that are the most important for the Facial Action Units formation. Their geometrical displacement information, defined as the difference between the first and the last frame, is used as an input to a six class (one for each facial expression) Support Vector Machine System. The output of the Support Vector is the facial expression recognized from the image sequence. The above mentioned novel real-time system achieves a recognition rate of approximately 98%, which is the best achieved, according to the authors knowl-

edge of the facial expression recognition literature.

6. REFERENCES

- [1] P. Ekman, and W.V. Friesen, "Emotion in the Human Face," *Prentice Hall*, 1975.
- [2] T. Kanade, J. Cohn, and Y. Tian, "Comprehensive Database for Facial Expression Analysis," *Proceedings of IEEE International Conference on Face and Gesture Recognition*, 2000.
- [3] M. Pantic, and L.J.M. Rothkrantz, "Automatic Analysis of Facial Expressions: The State of the Art," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000.
- [4] W. Fellenz, J. Taylor, N. Tsapatsoulis, and S. Kollias, "Comparing Template-based, Feature-based and Supervised Classification of Facial Expressions from Static Images," *Computational Intelligence and Applications, World Scientific and Engineering Society Press*, 1999.
- [5] L. Wiskott, J.-M. Fellous, N. Kruger, and C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997.
- [6] I. Cohen, N. Sebe, A. Garg, M.S. Lew and T.S. Huang, "Facial Expression Recognition from Video Sequences," *Proceedings of International Conference on Multimedia & Expo*, 2002.
- [7] M. Pantic, and L. J. M. Rothkrantz, "Expert System for Automatic Analysis of Facial Expressions," *Image and Vision Computing*, 2000.
- [8] V. Vapnik, "The nature of statistical learning theory," *Springer Verlag*, 1995.
- [9] S. Krinidis, and I. Pitas, "Statistical Analysis of Facial Expressions for Facial Expression Synthesis," *submitted to IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004.
- [10] J. Weston, and C. Watkins "Multi-class Support Vector Machines," *Technical report CSD-TR-98-04*, 2004.
- [11] M. J. Lyons, J. Budynek, and S. Akamatsu, "Automatic Classification of Single Facial Images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1999.