FACIAL EXPRESSION RECOGNITION IN VIDEOS USING A NOVEL MULTI-CLASS SUPPORT VECTOR MACHINES VARIANT

Irene Kotsia[†]*, Nikolaos Nikolaidis*[†] *and Ioannis Pitas*[†]

[†]Aristotle University of Thessaloniki
Department of Informatics
Box 451, 54124 Thessaloniki, Greece

ABSTRACT

In this paper, a novel class of Support Vector Machines (SVM) is introduced to deal with facial expression recognition. The proposed classifier incorporates statistic information about the classes under examination into the classical SVM. The developed system performs facial expression recognition in facial videos. The grid tracking and deformation algorithm used tracks the Candide grid over time as the facial expression evolves, until the frame that corresponds to the greatest facial expression intensity. The geometrical displacement of Candide nodes is used as an input to the bank of novel SVM classifiers, that are utilized to recognize the six basic facial expressions. The experiments on the Cohn-Kanade database show a recognition accuracy of 98.2%.

Index Terms— Facial Expression Recognition, Facial Action Coding System, Support Vector Machines, Candide Grid.

1. INTRODUCTION

Facial expression recognition has attracted a great interest during the past two decades, due to its importance for human centered interfaces. A set of six basic facial expressions (anger, disgust, fear, happiness, sadness and surprise) were defined [1]. A set of muscle movements, known as $Facial\ Action\ Units\ (FAUs)$, that produce each facial expression when combined following specific rules [2], was created by psychologists, thus forming the so called $Facial\ Action\ Coding\ System\ (FACS)$ [3]. A survey on automatic facial expression recognition can be found in [4].

In this paper, a method for recognizing facial expressions in videos using geometrical information and a novel class of SVM, is proposed. The geometrical displacements of the Candide facial model grid points (being tracked on the face through time), defined as the difference of each point's coordinates between the first and the last frame of the video, are used as an input to a novel multi-class SVM system that incorporates statistic information about the classes under examination. Essentially the paper improves the system proposed in [5] by utilizing the novel multi-class SVM system. The acquired experimental results justify the improved performance of the system.

This work was supported by the "SIMILAR" European Network of Excellence on Multimodal Interfaces of the IST Programme of the European Union (www.similar.cc).

2. GEOMETRICAL DISPLACEMENT INFORMATION EXTRACTION

The geometrical information extraction is performed by a grid adaptation system, based on deformable models [5]. The Candide grid is semi-automatically adjusted to the face on the first video frame and then tracked through the video, following the facial expression evolving through time. At the end, the grid tracking algorithm produces the deformed Candide grid that corresponds to the facial expression appearing at the last frame of the video, i.e. the one with the greatest intensity. The diagram of the system is shown in Figure 1.

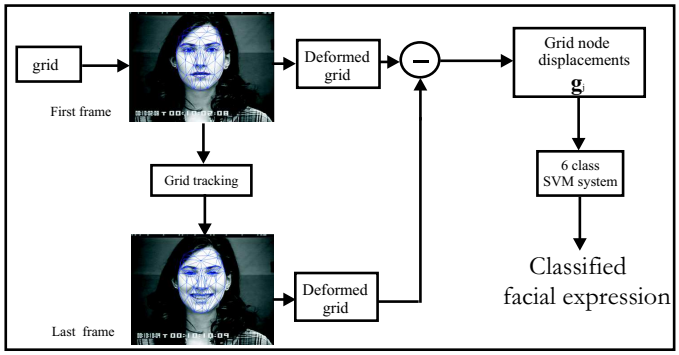

Fig. 1. System architecture for facial expression recognition in facial videos.

The geometrical information used are the displacements $\mathbf{d}_j^i$ of the Candide grid points, defined as the difference between the last and the first frame's coordinates of the point:

$$\mathbf{d}_j^i = [\Delta x_j^i \quad \Delta y_j^i]^T, \quad i \in \{1, \ldots, K\} \quad \text{and} \quad j \in \{1, \ldots, N\} \tag{1}$$

where i is the point index ($K = 104$ Candide grid points were used in our case) and j is the index of the video examined.

In that way, for every video, a feature vector $\mathbf{g}_j$ is constructed:

$$\mathbf{g}_j = [\mathbf{d}_j^1 \quad \mathbf{d}_j^2 \quad \dots \quad \mathbf{d}_j^K]^T. \tag{2}$$

where the vector $\mathbf{g}_j$ has $F = 104 \cdot 2 = 208$ dimensions. While training the SVM system, a set of feature vectors $\mathbf{g}_j \in \Re^F$ is used as an input, labelled properly with the true corresponding facial expression. To perform testing, an unlabelled feature vector $\mathbf{g}_p$ is used as an input. The trained SVM system provides a label that classifies $\mathbf{g}_p$ to one of the six basic facial expressions.

In this paper, a new multi-class SVM approach is used for this purpose. Before proceeding to the description of this new SVM variant, a brief outline of the standard two-class and multi-class SVM will be provided.

3. GEOMETRICAL DISPLACEMENT INFORMATION CLASSIFICATION USING CLASSICAL SVM SYSTEMS

3.1. Two-class SVM systems

A two-class SVM classifier finds a hyperplane or surface that separates the two-classes $\mathcal{F}^1$ and $\mathcal{F}^2$ with the maximum margin [6]. In order to train a two-class SVM network using soft margin formulation, the following minimization problem has to be solved [6]:

$$\min_{\mathbf{w},b,\boldsymbol{\xi}} \qquad \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{j=1}^{N}\xi_j \tag{3}$$

subject to the separability constraints:

$$y_i(\mathbf{w}^T\phi(\mathbf{x}_j) + b) \geq 1 - \xi_j, \xi_j \geq 0, \quad j = 1, \dots, N \tag{4}$$

where $\mathbf{w}$ is the vector of hyperplane coefficients, b is the bias, $\boldsymbol{\xi} = [\xi_i, \dots, \xi_w]$ is the slack variable vector, C is the term that penalizes the training errors and y_i is the class label of the vector $\mathbf{x}_i$ that takes values in $\{-1, 1\}$.

After solving the optimization problem (3) subject to the separability constraints (4), the decision function that can be used to classify unlabelled samples is:

$$f(\mathbf{g}) = \text{sign}(\mathbf{w}^T\phi(\mathbf{g}) + b). \tag{5}$$

In this formulation, a non-linear mapping ϕ is used. On the other hand, if a linear SVM system is to be constructed then $\phi(\mathbf{g}) = \mathbf{g}$. The non-linear mapping is defined by a positive kernel function, $h(\mathbf{g}_i, \mathbf{g}_j)$, specifying an inner product in the feature space and satisfying the Mercer condition [6]:

$$h(\mathbf{g}_i, \mathbf{g}_j) = \phi(\mathbf{g}_i)^T\phi(\mathbf{g}_j). \tag{6}$$

Typical kernels include the polynomial and Radial Basis Functions (RBF) kernels:

$$\begin{aligned} h(\mathbf{x}, \mathbf{y}) &= \phi(\mathbf{x})^T\phi(\mathbf{y}) = (\mathbf{x}^T\mathbf{y} + 1)^d \\ h(\mathbf{x}, \mathbf{y}) &= \phi(\mathbf{x})^T\phi(\mathbf{y}) = e^{-\gamma(\mathbf{x}-\mathbf{y})^T(\mathbf{x}-\mathbf{y})} \end{aligned} \tag{7}$$

where d is the degree of the polynomial kernel and γ is the spread of the Gaussian cluster. These kernels have been used in the experiments conducted in this paper.

3.2. Multi-class SVM

The multi-class SVM is a generalization of two-class SVM systems in order to deal with multi-class problems. In facial expression recognition this multi-class SVM constructs 6 facial expressions rules, where the $k-$th function $\mathbf{w}_k^T\phi(\mathbf{g}_j)+b_k$ separates training vectors of the facial expression class k from the rest of the vectors, by minimizing the objective function:

$$\min_{\mathbf{w},\mathbf{b},\boldsymbol{\xi}} \qquad \frac{1}{2}\sum_{k=1}^{6}\mathbf{w}_k^T\mathbf{w}_k + C\sum_{j=1}^{N}\sum_{k\neq l_j}\xi_j^k \tag{8}$$

subject to the constraints:

$$\begin{aligned} \mathbf{w}_{l_j}^T\phi(\mathbf{g}_j) + b_{l_j} &\geq \mathbf{w}_k^T\phi(\mathbf{g}_j) + b_k + 2 - \xi_j^k \\ \xi_j^k \geq 0, \quad j = 1, \dots, N, \quad k &\in \{1, \dots, 6\}\backslash l_j \end{aligned} \tag{9}$$

where l_j is the label of the geometrical displacement vector $\mathbf{g}_j$ which takes values in $\{1, \dots, 6\}$. Then, the decision function is:

$$h(\mathbf{g}) = \underset{k=1,\dots,6}{\text{argmax}}(\mathbf{w}_k^T\phi(\mathbf{g}) + b_k). \tag{10}$$

4. GEOMETRICAL DISPLACEMENT INFORMATION CLASSIFICATION USING THE PROPOSED SVM SYSTEMS

In this section, a novel multi-class classifier will be presented. In order to smoothly introduce the proposed variant which is an extension of the two-class SVM proposed in [7] to multiple classes, the method in [7] will be firstly described.

4.1. The two-class SVM incorporating class information

The two-class classifiers in [7] have been inspired by the optimization of the Fisher's discriminant ratio. That is, motivated by the fact that the Fisher's discriminant optimization problem for two-classes is a constrained least-squares optimization problem, the problem of minimizing the within-class variance has been reformulated in [7], so that it can be solved by constructing the optimal separating hyperplane for both separable and nonseparable cases. More details about the motivations of this modified SVM can be found in [7].

4.1.1. The Linear Case

In order to form the optimization problem of the modified SVM proposed in [7], the within class scatter matrix of the training set should be defined in the two-class case:

$$\mathbf{S}_w = \sum_{\mathbf{x}_i\in\mathcal{F}^1}(\mathbf{x}_i-\boldsymbol{\mu}_1)(\mathbf{x}_i-\boldsymbol{\mu}_1)^T + \sum_{\mathbf{x}_i\in\mathcal{F}^2}(\mathbf{x}_i-\boldsymbol{\mu}_2)(\mathbf{x}_i-\boldsymbol{\mu}_2)^T \tag{11}$$

where $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are the mean vectors of the classes $\mathcal{F}^1$ and $\mathcal{F}^2$, respectively. The optimization problem of the proposed SVM is [7]:

$$\min_{\mathbf{w},b,\boldsymbol{\xi}} \quad \mathbf{w}^T\mathbf{S}_w\mathbf{w} + C\sum_{j=1}^{N}\xi_j \tag{12}$$

subject to the separability constraints (4). The solution of this constrained optimization problem is given by the saddle point of the Lagrangian:

$$L(\mathbf{w},b,\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\xi}) = \mathbf{w}^T\mathbf{S}_w\mathbf{w} + C\sum_{i=1}^{N}\xi_i - \\ -\sum_{i=1}^{N}\alpha_i[y_i(\mathbf{w}^T\mathbf{x}_i - b) - 1 + \xi_i] - \sum_{i=1}^{N}\beta_i\xi_i \tag{13}$$

where $\boldsymbol{\alpha} = [\alpha_1,\ldots,\alpha_N]$ and $\boldsymbol{\beta} = [\beta_1,\ldots,\beta_N]$ are the vectors of Lagrangian multipliers for the constraints (4). The linear decision function is:

$$f(\mathbf{x}) = \text{sign}(\mathbf{w}^T\mathbf{x} + b) = \text{sign}(\frac{1}{2}\sum_{j=1}^{N} y_i\alpha_i\mathbf{x}_j^T{\mathbf{S}_w}^{-1}\mathbf{x} + b). \tag{14}$$

4.1.2. The Non-Linear Case

By applying the non-linear function ϕ to the vectors ${\mathbf{S}_w}^{-\frac{1}{2}}\mathbf{x}_i$, it is derived that $h({\mathbf{S}_w}^{-\frac{1}{2}}\mathbf{x}_i, {\mathbf{S}_w}^{-\frac{1}{2}}\mathbf{x}_j) =$
$= \phi({\mathbf{S}_w}^{-\frac{1}{2}}\mathbf{x}_i)^T\phi({\mathbf{S}_w}^{-\frac{1}{2}}\mathbf{x}_j)$ [7]. Then, kernel functions are applied and the Wolf dual problem [7] can be written as:

$$W(\boldsymbol{\alpha}) = \sum_{i}^{N}\alpha_i - \frac{1}{4}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j h({\mathbf{S}_w}^{-\frac{1}{2}}\mathbf{x}_i, {\mathbf{S}_w}^{-\frac{1}{2}}\mathbf{x}_j). \tag{15}$$

The corresponding non-linear decision function is given by:

$$f(\mathbf{x}) = \text{sign}(\frac{1}{2}\sum_{j=1}^{N} y_i\alpha_i h({\mathbf{S}_w}^{-\frac{1}{2}}\mathbf{x}_j, {\mathbf{S}_w}^{-\frac{1}{2}}\mathbf{x}) + b). \tag{16}$$

4.2. The proposed multi-class Classifier

As mentioned above, the proposed classifier is a generalization of the classifier presented in Section 4.1 towards handling multiple classes. The linear and non-linear cases of this classifier are described below.

4.2.1. The Linear Case

Let the within class scatter matrix of the grid deformation feature vectors $\mathbf{g}_i$ be defined as:

$$\mathbf{S}_w = \sum_{k=1}^{M}\sum_{\mathbf{g}_i\in\mathcal{U}_k}(\mathbf{g}_i - \boldsymbol{\mu}_k)(\mathbf{g}_i - \boldsymbol{\mu}_k)^T \tag{17}$$

where M is the number of facial expression classes (here equal to six), $\boldsymbol{\mu}_k$ is the mean geometrical displacement vector for the class k and $U_k, k \in \{1,\ldots,6\}$ the k-th facial expression class. The within class scatter matrix $\mathbf{S}_w$ is assumed to be invertible, which holds for the case under examination since the feature vector dimension is smaller than the available training examples.

By extending (12) the proposed constrained optimization problem is:

$$\min_{\mathbf{w}_k,\mathbf{b},\boldsymbol{\xi}} \quad \sum_{k=1}^{6}\mathbf{w}_k^T\mathbf{S}_w\mathbf{w}_k + C\sum_{j=1}^{N}\sum_{k\neq l_j}\xi_j^k \tag{18}$$

subject to the separability constraints in (9). The solution of the above constrained optimization problem can be given by finding the saddle point of the Lagrangian:

$$L(\mathbf{w},\mathbf{b},\boldsymbol{\xi},\boldsymbol{\alpha},\boldsymbol{\beta}) = \textstyle\sum_{k=1}^{6}\mathbf{w}_k^T\mathbf{S}_w\mathbf{w}_k + +C\sum_{i=1}^{N}\sum_{k=1}^{6}\xi_i^k - \\ -\sum_{i=1}^{N}\sum_{k=1}^{6}\alpha_i^k[(\mathbf{w}_{l_i} - \mathbf{w}_k)^T\mathbf{g}_i + b_{l_i} - b_k - 2 + \xi_i^k] \\ -\sum_{i=1}^{N}\sum_{k=1}^{6}\beta_i^k\xi_i^k \tag{19}$$

where $\boldsymbol{\alpha} = [\alpha_1^1,\ldots,\alpha_i^k,\ldots,\alpha_N^6]$ and $\boldsymbol{\beta} = [\beta_1^1,\ldots,\beta_i^k,\ldots,\beta_N^6]$ are the Lagrangian multipliers for the constraints (9) with :

$$\alpha_i^{l_i} = 0,\ \xi_i^{l_i} = 2,\ \beta_i^{l_i} = 0,\ i = 1,\ldots,N \tag{20}$$

and constraints:

$$\alpha_i^k \geq 0,\ \beta_i^k \geq 0,\ i = 1,\ldots,l,\ k \in \{1,\ldots,6\}\backslash l_i. \tag{21}$$

The Lagrangian (19) has to be maximized with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ and minimized with respect to $\mathbf{w}$ and $\boldsymbol{\xi}$. In order to produce a more compact equation form let us define the following variables:

$$A_i = \sum_{k=1}^{6}\alpha_i^k \ \text{ and } \ c_i^k = \begin{cases} 1, & \text{if } l_i = k \\ 0, & \text{if } l_i \neq k. \end{cases} \tag{22}$$

After a series of algebraic manipulations, the search of the saddle point of the Lagrangian (19) is reformulated to the maximization of the Wolf dual problem:

$$W(\boldsymbol{\alpha}) = 2\textstyle\sum_{i=1}^{N}\sum_{k=1}^{6}\alpha_i^k + \frac{1}{4}\sum_{i,j,k}(-\frac{1}{2}c_j^{l_j}A_iA_j + \\ +\alpha_i^k\alpha_i^{l_i} - \frac{1}{2}\alpha_i^k\alpha_j^k)\mathbf{g}_i\mathbf{S}_w^{-1}\mathbf{g}_j \tag{23}$$

which is a quadratic function in terms of $\boldsymbol{\alpha}$ with the linear constraints:

$$\sum_{i=1}^{N}a_i^k = \sum_{i=1}^{N}c_i^kA_i,\ k = 1,\ldots,6. \tag{24}$$

The corresponding decision hyperplane can be proven to be:

$$f(\mathbf{g}) = \text{argmax}_{k=1,\ldots,6}(\mathbf{w}_k^T\mathbf{g} + b_k) = \\ = \text{argmax}_{k=1,\ldots,6}[\tfrac{1}{2}\textstyle\sum_{i=1}^{N}(c_i^kA_i - \alpha_i^k)\mathbf{g}_i^T\mathbf{S}_w^{-1}\mathbf{g} + b_k]. \tag{25}$$

4.2.2. The Non-Linear Case

The non-linear multi-class decision surfaces can be created in the same manner as the two-class non-linear decision surfaces that have been proposed in [7] and described in Section 4.1.2. The fact that the term $\mathbf{g}_i^T \mathbf{S}_w^{-1} \mathbf{g}_j$ can be written in terms of dot products as $(\mathbf{S}_w^{-\frac{1}{2}} \mathbf{g}_i)^T (\mathbf{S}_w^{-\frac{1}{2}} \mathbf{g}_j)$, is exploited. Then, kernels can be applied in (23) as:

$$W(\boldsymbol{\alpha}) = 2 \sum_{i=1}^{N} \sum_{k=1}^{6} \alpha_i^k + \frac{1}{4} \sum_{i,j,k} (-\frac{1}{2} c_j^{l_j} A_i A_j + \\ + \alpha_i^k \alpha_i^{l_i} - \frac{1}{2} \alpha_i^k \alpha_j^k) h(\mathbf{S}_w^{-\frac{1}{2}} \mathbf{g}_i, \mathbf{S}_w^{-\frac{1}{2}} \mathbf{g}_j). \quad (26)$$

The corresponding decision surface can be proven to be:

$$f(\mathbf{g}) = \underset{k=1,\ldots,6}{\operatorname{argmax}} \frac{1}{2} [\sum_{i=1}^{N} (c_i^k A_i - \alpha_i^n) h(\mathbf{S}_w^{-\frac{1}{2}} \mathbf{g}_i, \mathbf{S}_w^{-\frac{1}{2}} \mathbf{g}) + b_k]. \quad (27)$$

5. EXPERIMENTAL RESULTS

The Cohn-Kanade database [3] was used to perform experiments regarding facial expression recognition in six basic facial expressions classes. The combinations of FAUs this database is annotated with, were translated into facial expressions according to [2], in order to define the corresponding ground truth for the facial expressions. All the subjects were used so as to form a database of over 400 videos.

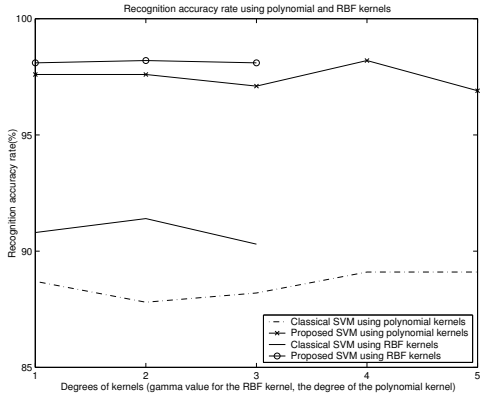

Fig. 2. Accuracy rates obtained for facial expression recognition using multi-class SVM.

The following procedure was followed for the experiments. All videos contained in the database were divided into 6 classes, each one corresponding to one of the 6 basic facial expressions to be recognized. A set containing 20% of the data for each class, chosen randomly, was created and used as the test set, while the remaining samples formed the training set. The procedure was repeated five times, each time with a different test set until all samples were used in the test set. The performance metric that was used was the average classification accuracy, i.e. the mean value of the percentages of the correctly classified facial expressions.

When the classical six class SVM were applied to feature vectors derived from the Candide grid, the best facial expression recognition accuracy achieved was equal to 91.4%. The best facial expression recognition accuracy when the proposed six class SVM was used, was equal to 98.2% (with different values if gamma and different degrees of the polynomial kernels). Figure 2 shows the accuracy rates achieved when polynomial and RBF kernels were used in the classical and proposed SVM.

6. CONCLUSION

Facial expression recognition in videos using SVM and geometrical information has been investigated in this paper. A novel multi-class SVM classifier that incorporates statistic information about the classes under examination into the classical SVM, has been proposed. The experiments yielded an accuracy rate equal to 98.2% which corresponds to a 6.4% increase from the recognition accuracy obtained when classical SVM were used.

7. REFERENCES

[1] P. Ekman and W. V. Friesen, *Emotion in the Human Face*, Prentice Hall, New Jersey, 1975.

[2] M. Pantic and L. J. M. Rothkrantz, "Expert system for automatic analysis of facial expressions," *Image and Vision Computing*, vol. 18, no. 11, pp. 881–905, August 2000.

[3] T. Kanade, J. Cohn and Y. Tian, "Comprehensive database for facial expression analysis," in *Proceedings of IEEE International Conference on Face and Gesture Recognition*, pp., 46–53.

[4] B. Fasel AND J. Luettin, "Automatic facial expression analysis: A survey," *Pattern Recognition*, vol. 36, no. 1, pp. 259–275, 2003.

[5] I. Kotsia, and I. Pitas, "Real time facial expression recognition from video sequences using support vector machines," in *Proceedings of Visual Communications and Image Processing (VCIP 2005)*, 2005.

[6] V. Vapnik, *Statistical learning theory*, Wiley, New York, 1998.

[7] A. Tefas, C. Kotropoulos and I. Pitas, "Using support vector machines to enhance the performance of elastic graph matching for frontal face authentication," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 7, pp. 735–746, 2001.