ELSEVIER

An analysis of facial expression recognition under partial facial image occlusion

Irene Kotsia *, Ioan Buciu, Ioannis Pitas

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

Abstract

In this paper, an analysis of the effect of partial occlusion on facial expression recognition is investigated. The classification from partially occluded images in one of the six basic facial expressions is performed using a method based on Gabor wavelets texture information extraction, a supervised image decomposition method based on Discriminant Non-negative Matrix Factorization and a shape-based method that exploits the geometrical displacement of certain facial features. We demonstrate how partial occlusion affects the above mentioned methods in the classification of the six basic facial expressions, and indicate the way partial occlusion affects human observers when recognizing facial expressions. An attempt to specify which part of the face (left, right, lower or upper region) contains more discriminant information for each facial expression, is also made and conclusions regarding the pairs of facial expressions misclassifications that each type of occlusion introduces, are drawn.
© 2008 Published by Elsevier B.V.

1. Introduction

During the past two decades, facial expression recognition has attracted a significant interest in the scientific community due to its importance for human centered interfaces. Facial expression recognition plays an important role in human–computer interfacing for many application areas, such as virtual reality, video conferencing, user profiling and customer satisfaction, for broadcasting and web services [1,2].

Several research efforts have been performed regarding facial expression recognition in the recent years. The facial expressions are usually classified in one of the six basic facial expressions (anger, disgust, fear, happiness, sadness and surprise) [3,4]. In order to make the recognition procedure more standardized, a set of facial muscle movements (known as Action Units) that produce each facial expres-

sion, was created by psychologists, thus forming the so-called *Facial Action Coding System* (FACS) [5].

A survey on automatic facial expression recognition can be found in [6]. Scientists conducted experiments in order to recognize the facial expressions from unoccluded facial images taken under controlled laboratory conditions. Unfortunately, at times, the human subject may be talking, thus altering his facial features or his face may be partially occluded. For example, sunglasses or virtual reality masks occlude the eyes region, while a scarf or a medical mask occlude the mouth region. Recognition of facial expressions in the presence of occlusion is investigated in [7]. In this paper, the lower, or higher, or left half or right half facial area is occluded. The approach in [7] is based on a localized representation of facial expression features and on the fusion of classifier outputs. Facial points are automatically tracked over an image sequence and are used to represent a face model. The classification of facial expressions is then performed by using decision level fusion that combines local interpretation of the face model into a general classification score. A study of four facial expressions

---

* Corresponding author. Tel./fax: +30 231 099 63 04.
 E-mail addresses: ekotsia@aiia.csd.auth.gr (I. Kotsia), pitas@aiia.csd.auth.gr (I. Pitas).

(anger, happiness, sadness and surprise) is made, in order to reveal how the occlusion of the eyes region, mouth region, left and right face side occlusion affect the robustness of the method used. An extensive discussion of our findings in comparison to the equivalent ones proposed in [7] will be given in Section 5.4.

In the current paper, the impact of facial occlusion for all six basic facial expressions is studied systematically. The simplest case of partial occlusion is considered, where only frontal facial views are available. Partial facial occlusion is simulated by superimposing graphically generated glasses/mouth or left/right region masks on unoccluded facial expression databases. The facial eyes occlusion mask simulates a pair of black sunglasses or virtual reality glasses, while the facial mouth occlusion mask simulates a medical mask or a scarf. The left/right occlusion masks are black rectangles covering the equivalent half part of the face. Texture and shape information extraction methods were chosen to achieve facial expression recognition. An extended analysis of the experiments using partially occluded faces was made, in order to investigate which part of the face affects the recognition rate of each facial expression, as well as to define the pairs of facial expressions that are usually confused with each other. Furthermore, an attempt to specify which part of the face contains the most discriminant information for every facial expression and to define the overall performance rate, was made.

An unsupervised method based on Gabor wavelets was used for texture information. The motivations and the appropriateness of the Gabor information for facial image analysis and facial expression recognition are discussed in [8–15] and in the references therein. Gabor filtering has been thoroughly used as a pre-processing step to be subsequently used in combination with neural networks to recognize facial expressions. Oriented quadrature phase Gabor wavelets were used as a pre-processing step in order to create the input for a multi-layer perceptron network [11] or a Support Vector Machine (SVM) classifier [12]. Multi-scale and multi-orientation Gabor filters were extracted from the fiducial points of facial images to be used as an input to a multi-layer perceptron [13,14]. Gabor filters were also applied to facial images in [15]. Their results were downsampled and their vectors were normalized to unit length, performing in that way a divisive contrast normalization, so as to classify 12 facial actions in the upper and lower face. The sampling of the Gabor filter bank output at various facial grid vertices was used to construct an unique labelled graph vector, consisting of the amplitude of complex valued Gabor transform coefficients sampled on a grid. Its dimensionality was reduced using Principal Component Analysis (PCA) and Linear Discrimination Analysis (LDA) to be used in facial expression recognition [8].

The second method regarding texture information is based on a supervised feature extraction method the so-called *Discriminant Non-negative Matrix Factorization* (DNMF) [16–18]. The use of DNMF algorithm for facial expression recognition has been motivated by the fact that it can achieve a sparse discriminant decomposition of faces in which almost all features found by its basis images are represented by the salient face features, such as eyes, eyebrows or mouth, as noted in [16,17]. The use of the DNMF algorithm to facial expression recognition problem is well motivated, since the algorithm is capable of decomposing the images into facial parts that play a vital role to facial expression [16,17]. Additional information regarding the suitability of the DNMF algorithm in facial image representation can be found in [18]. Moreover, some researchers claim that there is a connection between the sparsity of the bases images and the robustness of a method to partial image occlusion [19]. Thus, we anticipate the DNMF method, that produces sparse bases, to have fair performance in cases of facial image occlusion.

Finally, apart from texture-based methods, a shape-based method using geometrical displacement vectors as features accompanied by a multi-class SVM system for classification was also used. The shape information is extracted by fitting and tracking the Candide facial model in image sequences using the method presented in [20]. The use of geometrical displacement vectors with SVM classifiers been proven successful for facial expression recognition by various researchers [21,22].

The rest of the paper is organized as follows: the use of all the aforementioned methods in facial expression recognition is briefly presented in Section 2. The description of the database used for the experiments, as well as the procedure followed and the experiments conducted on unoccluded (to serve as a baseline) and occluded images are presented in Section 3. A study regarding the effect of left/right occlusion in facial expression recognition accuracy is presented in Section 4. Experiments conducted with human observers when eyes/mouth facial region occlusion is present are described in Section 5, where the set of facial expressions affected by occlusion is empirically predicted. Moreover, an attempt to specify which part of the face contains more discriminant information and the pairs of facial expressions that are mostly confused when eyes/mouth region is occluded is made. General conclusions are drawn in Section 6.

2. System description

Various approaches were followed to achieve facial expression recognition. The methods chosen can be divided in two categories: those that use holistic texture information, having as an input the whole facial image (Gabor filters [23] and DNMF algorithm [16]) and those that use shape-based information, taking under consideration the displacement of certain points on the facial region using (a multi-class SVM method [22]). The flow diagrams of the system used in each case (holistic and feature-based approach) are shown in Fig. 1. The texture information extraction subsystems using either Gabor filters or the DNMF algorithm are shown at the top, while the shape
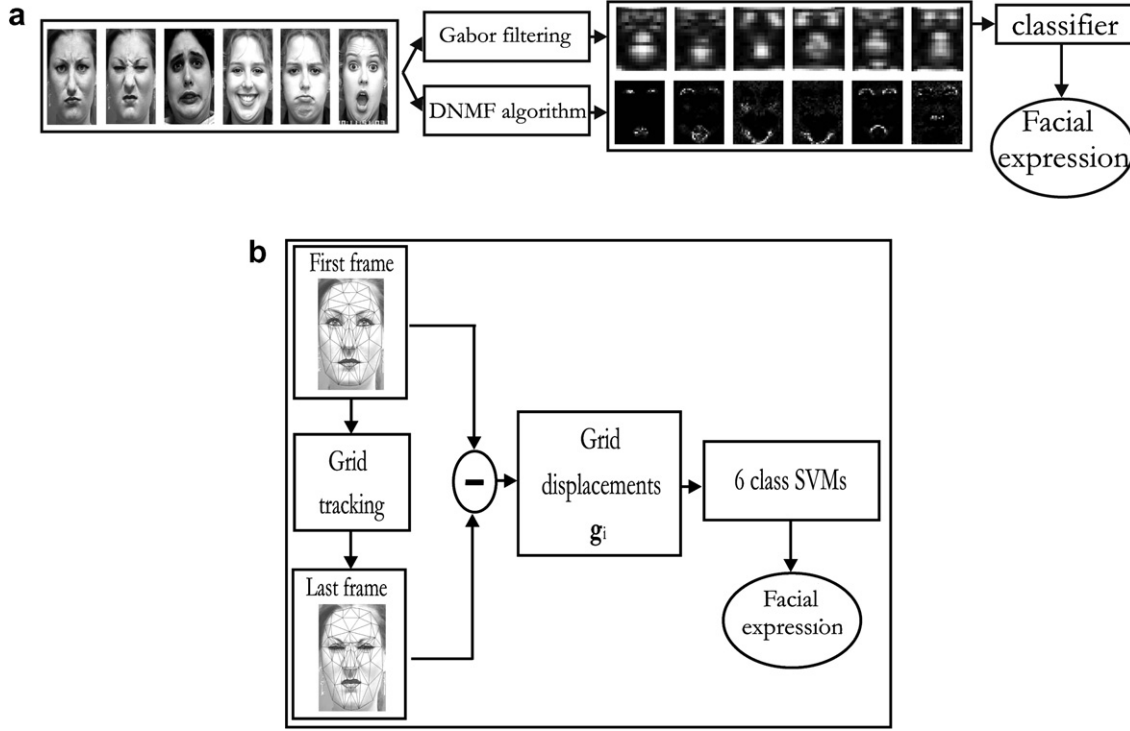
Fig. 1. Diagram block for facial expression recognition systems. (a) The texture information extraction subsystems. (b) The shape information extraction subsystem.

information extraction subsystem using SVMs, is shown at the bottom of the figure. The input images (or image sequences for the shape information extraction case) are the original here, but for the experiments the equivalent occluded images were used. In the following, we will briefly describe the use of the above mentioned methods in facial expression recognition.

2.1. Gabor filters for the extraction of texture information

Gabor filters (GF) were applied on the entire face instead of specific facial regions, thus avoiding the manual selection of regions of interest to extract facial features. A 2D Gabor wavelet transform is defined as the convolution of the image $\mathcal{I}(\mathbf{z})$:

$$\mathcal{J}_{\mathbf{k}}(\mathbf{z}) = \int_{\mathbf{z}} \mathcal{I}(\mathbf{z})\psi_{\mathbf{k}}(\mathbf{z} - \mathbf{z}')\,\mathrm{d}\mathbf{z} \tag{1}$$

with a family of Gabor filters [24]:

$$\psi_{\mathbf{k}}(\mathbf{z}) = \frac{\|\mathbf{k}\|^2}{\sigma^2}\exp\left(-\frac{\|\mathbf{k}\|^2\|\mathbf{z}\|^2}{2\sigma^2}\right)\left(\exp(i\mathbf{k}^{\mathrm{T}}\mathbf{z}) - \exp\left(-\frac{\sigma^2}{2}\right)\right), \tag{2}$$

where $\mathbf{z} = (x, y)$, σ is equal to 2π and $\mathbf{k}$ is the characteristic wave vector:

$$\mathbf{k} = \left[k_v\cos\varphi_\mu, k_v\sin\varphi_\mu\right]^{\mathrm{T}} \tag{3}$$

with

$$k_v = 2^{-\frac{v+2}{2}}\pi, \quad \varphi_\mu = \mu\frac{\pi}{8}. \tag{4}$$

The parameters v and μ define the frequency and orientation of the Gabor filter. In the implementation used, four orientations $0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}$ were used as well as two frequency ranges: high frequencies (hfr) for $v = 0, 1, 2$ and low frequencies (lfr) for $v = 2, 3, 4$ [15]. The middle frequency band $v = 2$ appears on both frequency bands, as it is considered an intermediate frequency.

A feature vector is formed by convolving a 80×60 facial image $I(\mathbf{z})$ with 12 Gabor filters corresponding to the chosen frequency range (low or high frequency) and orientation. It is then downsampled to an image of 20×15 pixels and scanned row by row to form a vector of dimension 300×1 for each Gabor filter output. Only the magnitude of the Gabor filter output was used in this representation, as it varies slowly with the face position, while the phase of the Gabor filter output is very sensitive to face position. The 12 Gabor filter output vectors have been concatenated to form a new long feature vector $\mathbf{s}$ of dimension m (in that case m is equal to 3600×1). The classification procedure is performed using directly the feature vectors.

For the experiments, the six basic facial expressions form six facial expression classes. Hence, $\mathcal{L} = \{anger(an), disgust(di), fear(fe), happiness(ha), sadness(sa), surprise(su)\}$. Lets denote the classes by $\mathcal{L}_j, j = 1, 2, \ldots, 6$. The label of the class $\mathcal{L}_j$ is denoted by l_j. Thus, each image $\mathbf{y}$ in the facial image database belongs to one of the six basic facial expression classes $\{\mathcal{L}_1, \mathcal{L}_2, \ldots, \mathcal{L}_6\}$.

Two common types of classifiers were used in order to classify a new test sample, using the nearest neighbor rule the cosine similarity measure and the Euclidean measure. Those classifiers were chosen as they were the easiest to implement, taking under consideration that this paper studies the effect of occlusion in facial expression recognition and does not propose a novel method to deal with the problem. Thus, only the classifier that holds the best results is presented. The classification uses the Euclidean distance between a new test vector $\mathbf{s}_{(\text{test})}$ and the mean value $\mathbf{m}_j$ of all labelled sample vectors belonging to class $\mathcal{L}_j$:

$$\mathcal{L}_j = \underset{j=1,\dots,n}{\text{argmin}}\{\|\mathbf{s}_{(\text{test})} - \mathbf{m}_j\|\}. \tag{5}$$

The best recognition rates, using this approach, have been achieved when using the low frequency representation and the nearest neighbor Euclidean classifier. Therefore, from now on the conclusions will be made based on the recognition rates achieved for this set up.

2.2. The DNMF algorithm for the extraction of texture information

Facial expressions are very hard to define when a facial expression is evolving. The study of these intermediate states of facial expressions would be feasible only if ground truth was provided by psychologists. In order for that to be achieved, the evolution of Action Units (AUs) needs to be measured precisely in terms of intensity with respect to the greatest intensity of the facial expression. However, the existing databases do not provide that kind of information. The current databases depict the person at the neutral state at the beginning of the video sequence while the facial expression evolves to reach its highest intensity through time. However, the performance in the case of not fully expressed images is expected to be significantly lower since the facial expression depicted is not fully formed and thus can be easily confused with the others. Due to the lack of intermediate state ground truth, only the images that correspond to the greatest intensity of the facial expression were used.

For the application of the DNMF algorithm [16], the images that depict the facial expression in its greatest intensity are taken under consideration i.e. the last frame of the image sequence, to form the facial image database U that is comprised of L facial images. Each image $\mathcal{I}(\mathbf{z})$ is scanned row-wise to form a vector $\mathbf{x} \in \Re_+^F$.

Let $\mathbf{x}, \mathbf{q}$ be positive vectors $x_i > 0, q_i > 0$, then the Kullback–Leibler (KL) Divergence (or relative entropy) between $\mathbf{x}$ and $\mathbf{q}$ is defined [25] as:

$$KL(\mathbf{x}\|\mathbf{q}) \triangleq \sum_i \left(x_i \ln \frac{x_i}{q_i} + q_i - x_i \right). \tag{6}$$

The Non-negative Matrix Factorization (NMF) method tries to approximate the facial expression image $\mathbf{x}$ by a linear combination of a set of basis images. In order to apply NMF, the matrix $\mathbf{X} \in \Re_+^{F \times L} = [x_{i,j}]$ should be constructed, where $x_{i,j}$ is the ith element of the jth image. In other words,

the jth column of $\mathbf{X}$ is the $\mathbf{x}_j$ facial expressive image. NMF aims at finding two matrices $\mathbf{Z} \in \Re_+^{F \times M} = [z_{i,k}]$ and $\mathbf{H} \in \Re_+^{M \times L} = [h_{k,j}]$ such that:

$$\mathbf{X} \approx \mathbf{ZH}. \tag{7}$$

After the NMF decomposition, the facial expressive image $\mathbf{x}_j$ can be written as $\mathbf{x}_j \approx \mathbf{Zh}_j$, where $\mathbf{h}_j$ is the jth column of $\mathbf{H}$. Thus, the columns of the matrix $\mathbf{Z}$ can be considered as basis images and the vector $\mathbf{h}_j$ as the corresponding weight vectors. The $\mathbf{h}_j$ vectors can also be considered as the projected vectors of a lower dimensional feature space for the original facial expressive vector $\mathbf{x}_j$. The non-negativity constraints in the NMF decomposition yield a set of bases that correspond better to the intuitive notion of facial parts [26]. In order to measure the error of the approximation $\mathbf{x} \approx \mathbf{Zh}$ the $KL(\mathbf{x}\|\mathbf{Zh})$ divergence can been used [26]. Discriminant constraints in the DNMF algorithm are incorporated in the cost of the decomposition [16]. The discriminant constraints are employed for the weight matrix $\mathbf{H}$ of the decomposition, which is considered a projection to a lower dimensional space. Let the vector $\mathbf{h}_j$ that corresponds to the jth column of the matrix $\mathbf{H}$, be the coefficient vector for the ρth facial image of the rth facial expression class, which will be denoted as $\boldsymbol{\eta}_\rho^{(r)} = [\eta_{\rho,1}^{(r)} \dots \eta_{\rho,M}^{(r)}]^{\text{T}}$. The mean vector of the vectors $\boldsymbol{\eta}_\rho^{(r)}$ for the class r is denoted as $\boldsymbol{\mu}^{(r)} = [\mu_1^{(r)} \dots \mu_M^{(r)}]^{\text{T}}$, the mean of all classes as $\boldsymbol{\mu} = [\mu_1 \dots \mu_M]^{\text{T}}$ and the cardinality of each facial class $\mathcal{L}_r$ as N_r. Then, the within scatter matrix for the coefficient vectors $\mathbf{h}_j$ is defined as:

$$\mathbf{S}_w = \sum_{r=1}^6 \sum_{\rho=1}^{N_r} (\boldsymbol{\eta}_\rho^{(r)} - \boldsymbol{\mu}^{(r)})(\boldsymbol{\eta}_\rho^{(r)} - \boldsymbol{\mu}^{(r)})^{\text{T}} \tag{8}$$

whereas the between scatter matrix is defined as:

$$\mathbf{S}_b = \sum_{r=1}^6 N_r(\boldsymbol{\mu}^{(r)} - \boldsymbol{\mu})(\boldsymbol{\mu}^{(r)} - \boldsymbol{\mu})^{\text{T}}. \tag{9}$$

The defined discriminant cost of the discriminant decomposition of a facial database is the sum of all KL divergences for all images in the database plus the minimization of $\text{tr}[\mathbf{S}_w]$ and the maximization of $\text{tr}[\mathbf{S}_b]$:

$$D(\mathbf{X}\|\mathbf{ZH}) = \sum_j KL(\mathbf{x}_j\|\mathbf{Zh}_j) + \gamma\text{tr}[\mathbf{S}_w] - \delta\text{tr}[\mathbf{S}_b]. \tag{10}$$

The DNMF is the outcome of the following optimization problem:

$$\min_{\mathbf{Z},\mathbf{H}} \quad D(\mathbf{X}\|\mathbf{ZH})$$
$$\text{subject to} \quad z_{i,k} \geqslant 0, \quad h_{k,j} \geqslant 0, \quad \sum_i z_{i,j} = 1, \quad \forall j. \tag{11}$$

Following the same EM approach used by NMF [26] and LNMF [19] techniques, the following update rules for the weight coefficients $h_{k,j}$ that belong to the rth facial expression class are derived from [16]:

$$h_{k,j}^{(t)} = \frac{T_1 + \sqrt{T_1^2 + 4\left(2\gamma - (2\gamma + 2\delta)\frac{1}{N_r}\right)h_{k,j}^{(t-1)}\sum_i z_{i,k}^{(t-1)}\frac{x_{i,j}}{\sum_l z_{i,l}^{(t-1)}h_{l,j}^{(t-1)}}}}{2\left(2\gamma - (2\gamma + 2\delta)\frac{1}{N_r}\right)},$$

(12)

where T_1 is given by:

$$T_1 = (2\gamma + 2\delta)\left(\frac{1}{N_r}\sum_{\lambda, \lambda \neq l} h_{k,\lambda}\right) - 2\delta\mu_k - 1,$$

(13)

whereas, for the $z_{i,k}$, the update rules are given by:

$$\acute{z}_{i,k}^{(t)} = z_{i,k}^{(t-1)} \frac{\sum_j h_{k,j}^{(t)} \frac{x_{i,j}}{\sum_l z_{i,l}^{(t-1)}h_{l,j}^{(t)}}}{\sum_j h_{k,j}^{(t)}}$$

(14)

and

$$z_{i,k}^{(t)} = \frac{\acute{z}_{i,k}^{(t)}}{\sum_l \acute{z}_{l,k}^{(t)}}.$$

(15)

The above decomposition is a supervised non-negative matrix factorization method that decomposes the facial expression images into parts while enhancing the class separability. The matrix $\mathbf{Z}_D^\dagger = (\mathbf{Z}_D^T \mathbf{Z}_D)^{-1}\mathbf{Z}_D^T$, which is the pseudo-inverse of $\mathbf{Z}_D$, is then used for extracting the discriminant features as $\acute{\mathbf{x}} = \mathbf{Z}_D^\dagger \mathbf{x}$.

In order to make a decision about the facial expression class the image under consideration belongs to, the image is projected to the lower dimensional feature space derived applying the DNMF algorithm. The Euclidean distance between the projection $\acute{\mathbf{x}} = \mathbf{Z}_D^\dagger \mathbf{x}$ and the center of each facial expression class is calculated and the image is then classified to the facial expression class whose distance is the smallest, just as in the Gabor-based approach in Section 2.1.

The bases of NMF are localized features that correspond better to the intuitive notion of facial parts [27]. The belief that NMF produces local representations is mainly intuitive (i.e. addition of different non-negative bases using non-negative weights). Recently, some theoretical work has been performed in order to determine whether NMF provides a correct decomposition into parts [28].

The DNMF method achieves a decomposition of the facial images, whose basis images represent salient facial features, such as eyes, eyebrows or mouth. Hence, there is a correlation between the features discovered by DNMF algorithm and the facial expression classification framework. It is demonstrated that the DNMF basis images are salient facial parts that preserve discriminant information for every facial expression, like smile, lowered eyebrows etc. in contrast to the NMF basis images that do not display spacial locality of such high quality. An example of the basis images extracted for the NMF and DNMF algorithms is shown in Fig. 2.

An interested reader can refer to [29] and [30] for more details on the way NMF and DNMF algorithms affect the facial expression recognition performance.

2.3. A shape-based approach for Facial Expression Recognition

The displacement of the ith Candide grid node of the jth facial image, $\mathbf{d}_{i,j}$ is defined as the difference of the grid node coordinates at the first and the fully formed expression facial image sequence frame:

$$\mathbf{d}_{i,j} = \begin{bmatrix} \Delta x_{i,j} & \Delta y_{i,j} \end{bmatrix}^T \quad i = 1, \ldots, E \quad \text{and} \quad j = 1, \ldots, N$$

(16)

where $\Delta x_{i,j}$, $\Delta y_{i,j}$ are the x, y coordinate displacements of the ith node, respectively. E is the total number of nodes ($E = 104$ for the unoccluded Candide model) and N is the number of the facial image sequences. This way, for every facial image sequence in the training set, a feature vector $\mathbf{g}_j$ is created, called grid deformation feature vector containing the geometrical displacement of every grid node:

$$\mathbf{g}_j = \begin{bmatrix} \mathbf{d}_{1,j} & \mathbf{d}_{2,j} \ldots \mathbf{d}_{E,j} \end{bmatrix}^T, \quad j = 1, \ldots, N$$

(17)

having $L = 104 \times 2 = 208$ dimensions. We assume that each grid deformation feature vector $\mathbf{g}_j \ j = 1, \ldots, N$ belongs to one of the six facial expression classes (for facial expression recognition multi-class SVMs were used). The deformation of the grid is tracked using the algorithm presented in [20] where the facial feature tracking and facial expression synthesis problem is treated. The geometric displacement vectors have proven to be very useful and have been successfully combined with SVM classifiers by many researchers [21,22]. A brief conversation about the optimization problem of the multi-class SVMs will be given below. The interested reader can refer to [31–33] and the references therein for formulating and solving multi-class SVMs optimization problems.

The training data are $(\mathbf{g}_1, l_1), \ldots, (\mathbf{g}_N, l_N)$, where $\mathbf{g}_j \in \mathfrak{R}^L$ are the grid deformation vectors and $l_j \in \{1, \ldots, 6\}$ are the facial expression labels of the feature vector. The multi-class SVMs problem solves only one optimization problem [31,33]. It constructs six facial expressions rules, where the kth function $\mathbf{w}_k^T\phi(\mathbf{g}_j) + b_k$ separates training vectors of the class k from the rest of the vectors, by minimizing the objective function:

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

(18)

subject to the constraints:

$$\mathbf{w}_{l_j}^T\phi(\mathbf{g}_j) + b_{l_j} \geqslant \mathbf{w}_k^T\phi(\mathbf{g}_j) + b_k + 2 - \xi_j^k$$
$$\xi_j^k \geqslant 0, \quad j = 1, \ldots, N, \quad k \in \{1, \ldots, 6\} \setminus l_j$$

(19)

where $\phi : \mathfrak{R}^L \to \mathcal{H}$ is the function that maps the deformation vectors to a higher dimensional space in which the data
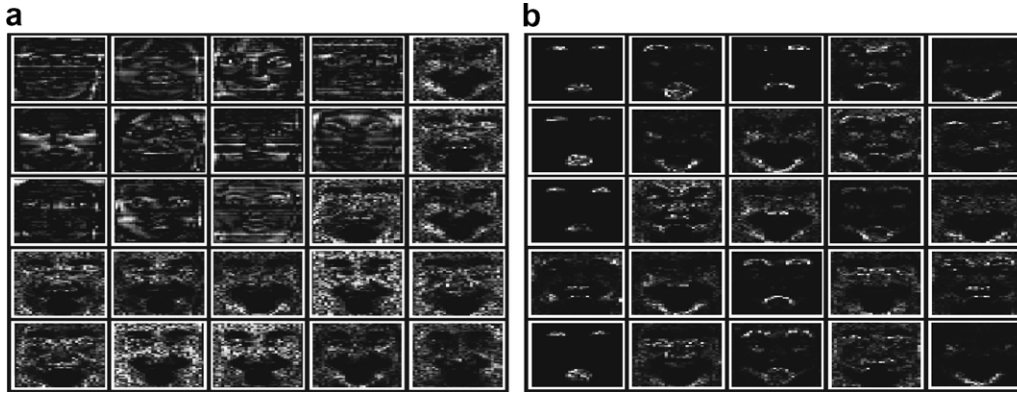
Fig. 2. Basis images extracted for (a) NMF (b) DNMF algorithms.

are supposed to be linearly or near linearly separable and $\mathcal{H}$ is an arbitrary dimensional Hilbert space. C is the term that penalizes the training errors. The vector $\mathbf{b} = [b_1 \dots b_6]^{\mathrm{T}}$ is the bias vector and $\boldsymbol{\xi} = [\xi_1^1, \dots, \xi_i^k, \dots, \xi_N^6]^{\mathrm{T}}$ is the slack variable vector. To solve the optimization problem (18) subject to (19), it is not necessary to compute ϕ but only dot products in $\mathcal{H}$. The dot products are defined by means of a positive kernel function, $h(\mathbf{g}_i, \mathbf{g}_j)$, specifying an inner product in the feature space and satisfying the Mercer condition [31,34]. The functions used as SVMs kernels in the experiments were the $d = 3$ degree polynomial function:

$$h(\mathbf{g}_i, \mathbf{g}_j) = \phi(\mathbf{g}_i)^{\mathrm{T}} \phi(\mathbf{g}_j) = (\mathbf{g}_i^{\mathrm{T}} \mathbf{g}_j + 1)^d. \qquad (20)$$

The decision function is defined as:

$$p(\mathbf{g}) = \operatorname*{argmax}_{k=1,\dots,6}(\mathbf{w}_k^{\mathrm{T}} \phi(\mathbf{g}) + b_k). \qquad (21)$$

Using this procedure, a test grid deformation feature vector is classified to one of the six facial expressions using (21). Once the six-class SVMs system is trained, it can be used for testing i.e. for recognizing facial expressions on new deformation vectors.

3. Discussion

In the beginning of this section, an empirical attempt to predict which part of the face is more important for every facial expression using the AUs that participate in the rules proposed in [35] and visual observations, is made. The database used for the experiments, as well as the procedure followed in order to form the training and test sets are described afterwards. The experimental procedure was firstly conducted with humans, experts and non-experts, to make an assumption regarding the facial expressions confused when eyes/mouth/left/right region occlusion is present. The same experiments were also conducted using the systems described in Section 2. The experiments performed on unoccluded images are presented in order to serve as a baseline. The occlusion experiments are presented afterwards in order to draw any occlusions regarding the effect of the occlusion of each facial region on

facial expression recognition. General conclusions regarding the effect of each facial region occlusion per facial expression are drawn. An attempt to specify the pairs of facial expressions that are mostly confused when no occlusion is present or a facial region occlusion is introduced, is presented.

3.1. Facial Action Coding System and facial expressions

The Facial Action Coding System (FACS) consists of Action Units (AUs), each one representing a possible facial muscle motion. The six basic expressions are produced by the combination of a set of AUs, following specific rules, as proposed in [35]. As can be seen in the rules (second column in Table 1), only AUs 1, 2, 4, 5, 6, 7, 9, 10, 12, 15, 16, 17, 20, 23, 24, 25 and 26 participate in the facial expression synthesis rules proposed in [35]. These AUs can be divided in three sets: the responsible for the upper face motion FAUs (1, 2, 4, 5, 6 and 7), the responsible for the middle facial area motion FAU (9) and the responsible for the lower face motion AUs (10, 12, 15, 16, 17, 20, 23, 24, 25 and 26) (see Fig. 3). The operator $+$ refers to the logical operator *AND* meaning here that the presence of both AUs has to be ensured for the condition to be true. The operator *or* refers to the logical operator *OR* meaning that the presence of only one of the two AUs is necessary for the condition to be true.

Two attempts were made to predict empirically which facial region is the most important for each facial expres-

Table 1
The AUs combination rules for describing the six basic facial expressions [35]

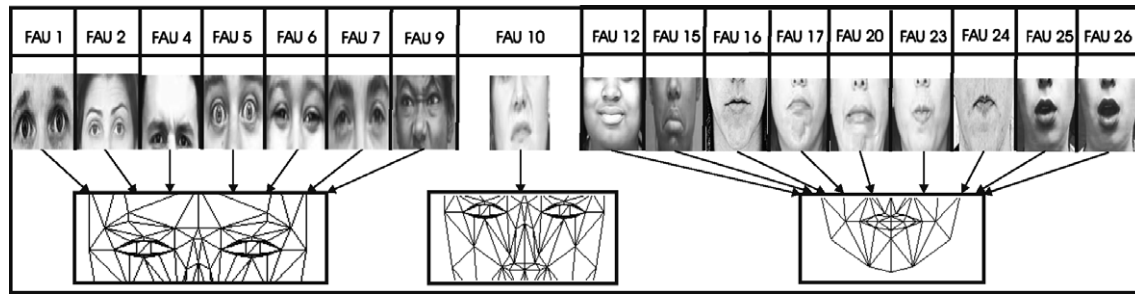| Facial expression	AUs coded description
Anger	$4 + 7 + (((23 \text{ or } 24) \text{ with or not } 17) \text{ or } (16 + (25 \text{ or } 26)) \text{ or } (10 + 16 + (25 \text{ or } 26))) \text{ with or not } 2$
Disgust	$((10 \text{ with or not } 17) \text{ or } (9 \text{ with or not } 17)) + (25 \text{ or } 26)$
Fear	$(1 + 4) + (5 + 7) + 20 + (25 \text{ or } 26)$
Happiness	$6 + 12 + 16 + (25 \text{ or } 26)$
Sadness	$1 + 4 + (6 \text{ or } 7) + 15 + 17 + (25 \text{ or } 26)$
Surprise	$(1 + 2) + (5 \text{ without } 7) + 26$

Fig. 3. Set of FAUs that form the rules in [35] and the corresponding part of the facial grid.

sion, based on the analysis of the rules proposed in [35] and on visual observation.

From the rules proposed in [35], we tried to specify the most characteristic AUs for every facial expression, in the sense that these AUs are the first one that someone notices when given an image of facial expression. Therefore, for anger it can be seen that the most distinctive AUs, in the sense that they are more visible than the others, are the AUs 23 and 24. Since those AUs belong to the third group that is responsible for mouth area motion, the mouth region will mostly affect the most the recognition of anger. The same assumption is valid for fear (AU 20) and sadness (AUs 15 and 17). On the other hand, for disgust, the most discriminative AUs seem to be 9 and 10, which control the middle and lower part of the face equivalently. However, AU 10 is more visible for the upper area of the mouth region. In combination with AU 9, the changes in the eyes and middle area seem to be more visible. Therefore, eyes occlusion would be expected to affect facial expression recognition more than mouth occlusion. The same assumption is valid for surprise as well (AUs 1, 2, 5 and 26). Happiness seems to be influenced by AUs 6 and 12 the most, that control the motion in different facial regions (eyes and mouth, respectively). However, the motion is more visible on the mouth region, thus mouth occlusion affects more facial expression recognition.

Assumptions regarding the most discriminant facial region for every facial expression were also made by visually observing. The difference images for each facial expression were calculated to serve that purpose (see Fig. 4). The difference images are derived by subtracting the last frame of the image sequence of each facial expression from the first one, corresponding to the greatest intensity of the facial expression and the neutral state, respectively. The motion that appears in each facial expression is emphasized that way, making it easier to predict which facial part should be more important by visual observation.

For anger facial expression, a frown is observed at the upper region of the nose. However, the mouth occlusion was simulated as if the poser wears a medical mask, that covers that region. Therefore, it's presence is taken under consideration when the mouth is present. Therefore, one should expect the mouth occlusion to affect more its recognition accuracy. The same conclusion can me made about
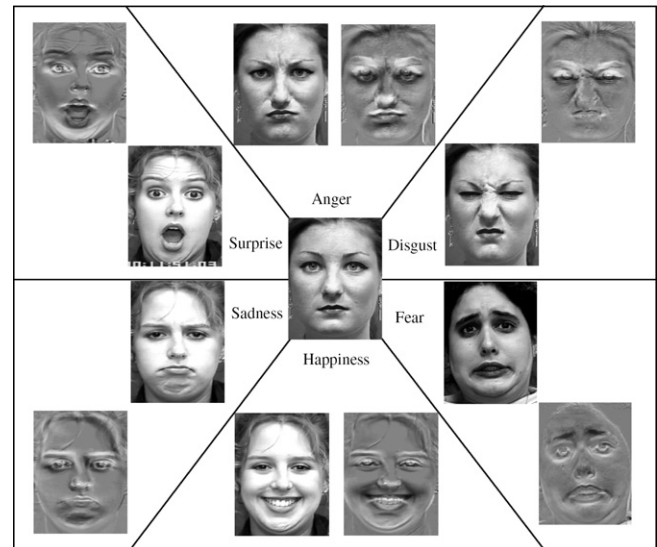

Fig. 4. Facial parts that affect movements the most, as derived from differences images.

fear, happiness and sadness, since the mouth region is more emphasized in the corresponding difference images. Disgust is a special kind of facial expression. The mouth region seems to attract the most interest, however the motion that appears between the eyes, in the bridge of the nose, is unique for the specific facial expression. Thus, the eyes facial region occlusion should be the one to play the most vital role in disgust's recognition. The same assumption is valid for surprise. The mouth region attracts the interest in the beginning, to notice from a more thorough examination that the eyes region motion is indeed quite distinctive. Therefore, the eyes region should be the one that plays the most vital role in surprise recognition.

The predictions acquired from the above mentioned two approaches are in par. Summarizing, it can be seen that the eyes occlusion is expected to affect more:

- disgust and
- surprise,

while the mouth occlusion should affect more

- anger
- fear

- happiness and
- sadness.

Regarding left/right facial region occlusion, someone can see that the absence of half the facial image should not really influence the recognition accuracy rate, since the representation of each facial expression is almost symmetrical on both facial regions. Therefore, the presence of left/right facial region is expected to preserve the results more or less the same.

3.2. Database pre-processing

There is no database available that contains facial expressions under various poses head rotations for all six facial expressions. Thus, only frontal poses will be taken under consideration for facial expression recognition. Future work includes experiments performed using the MMI database that introduces profile views for all six basic facial expressions. However, from the databases that include all six basic facial expressions, there is no database available that contains facial expressions with eyes/mouth occlusion for all six facial expressions for many posers. Therefore, facial expression occlusions had to be simulated by superimposing graphically generated glasses/mouth or left/right region masks on unoccluded facial expression databases. The case of occlusion studied in the paper is the simplest one. Occlusion was applied at all frames in every image sequence to produce that way the optimal simulation. The artificial masks used may lead to different results from those acquired when e.g. hands were used, due to their color. Thus, the results presented in this case are valid for these types of data. It is out of the scope of this paper and constitutes a subject of future research to investigate the effects of different types of occlusion on facial expression. Nevertheless, the applied type of occlusion simulates in realistic way the occlusion introduced when black sunglasses or a medical mask are used. Even if the entire face is available, occlusion can be assumed e.g. when the features in the eyes/mouth, left/right region cannot be efficiently tracked. The produced facial expressions are quite realistic and can be used to infer facial recognition performance under real occlusion (e.g. when wearing glasses or using the palm to shut the poser's mouth). However, new experiments should be performed when such databases containing real condition occlusion (including of course head rotation occlusion) are available to reconfirm or amend any findings.

Three databases exist that include all six basic facial expressions, the JAFFE database [14], the Cohn–Kanade database [5] and the MMI database [36]. The JAFFE and Cohn–Kanade databases were taken under consideration to form the database used for the experiments. The MMI database will be used for future research.

The JAFFE database is comprised of static images depicting the greatest intensity of each facial expression. All the subjects from JAFFE database were taken under consideration to form the database for the experiments. However, this database does not consist of image sequences, therefore the application of the shape-based method for the extraction of results was not possible. Thus, the assumptions made regarding the JAFFE database were extracted using the Gabor filtering method and the DNMF algorithm.

The Cohn–Kanade database consists of image sequences depicting the evolvement of every facial expression from the neutral state until it reaches its highest intensity in the last frame. All three methods were therefore possible to be applied. The Cohn–Kanade database is encoded into combinations of Action Units. These combinations were translated into facial expressions according to [35] in order to define the corresponding ground truth for the facial expressions. All the subjects were taken under consideration to form the database for the experiments.

Each original image has been aligned with respect to the eyes location. This was done in a automatic way using a method that performs eyes localization on a face, based only on geometrical information. A face detector is first applied to detect the bounding box of the face, and the edge map is extracted. A vector is assigned to every pixel, pointing to the closest edge pixel. Length and slope information for these vectors is consequently used to detect and localize the eyes. An interested reader can refer to [37] for more details regarding the method used.

The extracted eyes area was used to align the remaining images of the image sequence. A pair of black glasses and a mouth mask, as well as left and right face area masks were created using a graphics computer program, to be superimposed on the eyes or mouth regions, respectively, to simulate partial occlusion. The glasses were similar to black sun glasses, while the mouth mask was similar to a medical mask that covers the nose, cheeks, mouth and chin. A black rectangle covering the half facial area, either the right or the left one, was superimposed to the face to simulate right/left facial region occlusion. Then, each image was cropped and downsampled in a such way that the final image size is 80×60 pixels. In the case of images sequences, required for the shape-based facial expression recognition approach that uses SVMs, the simulated occlusion was applied to all frames of each image sequence, so that the produced image sequence would be more realistic. The tracking procedure was then applied to the occluded image sequence.

An example of a poser from the Cohn–Kanade database under eyes and mouth region occlusion for all facial expressions is shown in Fig. 5. Fig. 6 presents one expresser from Cohn–Kanade database posing for the six basic facial expressions. On each image, the Candide grid has been superimposed and deformed to correspond to the depicted facial expression, as it is used for the facial expression classification using shape information. The first and last row show the facial part that is taken under consideration when mouth and eyes occlusion is present. The equivalent subset of the Candide grid used for classification is also depicted.

Fig. 5. A poser from the Cohn–Kanade database under eyes and mouth region occlusion for all facial expressions.
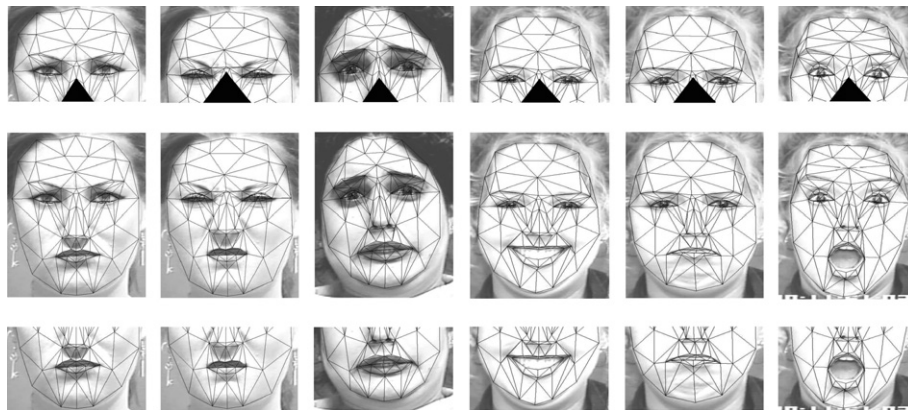

Fig. 6. A poser example from the Cohn–Kanade database, depicting the grid taken under consideration in the original image (second row) and when mouth and eyes occlusion is present (first and last row, respectively).

In Fig. 7, a poser example from the Cohn–Kanade database is shown, with or without the grid taken under consideration when left/right occlusion is present (first and second row, respectively).

The classifier accuracy was measured using the leave-one cross-validation approach [38], in order to make maximal use of the available data and produce averaged classification accuracy results. The term leave-one-out cross-validation, does not correspond to the classic leave-one-out definition here, as a variant of leave-one-out is used (i.e. leave 20% out) for the formation of the test dataset. How-

ever, the procedure followed will be called leave-one-out from now on. More specifically, all image sequences contained in the database are divided into six classes, each one corresponding to one of the six basic facial expressions to be recognized. Five sets containing 20% of the data for each class, chosen randomly, were created. One set containing 20% of the samples for each class is used for the test set, while the remaining sets form the training set. After the classification procedure is performed, the samples forming the testing set are incorporated into the current training set, and a new set of samples (20% of the samples for each

Fig. 7. A poser example with and without the grid under right/left facial region occlusion for all facial expressions.

class) is extracted to form the new test set. The remaining samples create the new training set. This procedure is repeated five times. The average classification accuracy is the mean value of the percentages of the correctly classified facial expressions [39].

The confusion matrices [8] have been computed for the experiments conducted have been calculated. The confusion matrix is a $n \times n$ matrix containing the information about the actual class label (in its columns) and the class label obtained through classification (in its rows). The diagonal entries of the confusion matrix are the percentages (%) that correspond to the rate of facial expressions that are correctly classified, while the off-diagonal entries are the percentages (%) corresponding to misclassification rates. The abbreviations $an, di, fe, ha, sa, su, no, ey$ and mo represent anger, disgust, fear, happiness, sadness, surprise, no occlusion, eyes occlusion and mouth occlusion, respectively. The confusion matrices obtained when no occlusion, eyes and mouth occlusion is present are merged into one, due to space limitations. In this representation, for every facial expression, three columns are presented aside, representing the results acquired when no occlusion, or eyes and mouth occlusion was present. The representation of the rows remains the same. That way, the comparison of the accuracy rates achieved for every facial expression under each type of occlusion is easier to be made.

4. Experiments in the presence of left/right facial region occlusion

Experiments were conducted under left/right facial region occlusion. Those experiments included both human observer experiments (experts and non-experts) and machine experiments (using all of the described methods). In all cases, the accuracy rates achieved under left or right facial region occlusion were the same. The presence of this type of occlusion did not decrease the recognition accuracy rate enough to be considered significant. Thus, each facial side is assumed to contain the same discriminatory information for facial expression recognition and is adequate for facial expression recognition. Therefore, the left/right facial region occlusion will not be studied furthermore.

5. Experiments in the presence of eyes/mouth facial region occlusion

5.1. Experiments with human observers

Human observers were employed to conduct the experiments, using non-occluded images and images under eyes or mouth region occlusion. The experiments were performed using the same methodology followed for the experiments with the proposed system. The original and occluded databases were shown to 3 expert humans (studying facial expression recognition for more than 2 years) and to 13 non-expert humans.

5.1.1. Experiments with expert human observers

Expert human observers achieved a recognition accuracy rate of 97.4%, 85.9% and 70.9% when no occlusion, eyes and mouth region is present. Therefore, overall for all facial expressions mouth occlusion seems to affect the recognition of facial expressions more than eyes region occlusion (11.5% and 26.5% decrease in accuracy rate, respectively).

The confusion matrices calculated when no occlusion, eyes and mouth occlusion is present are shown in Table 2, respectively. Regarding each facial expression separately (the lowest accuracy rates for each facial expression under eyes and mouth occlusion are displayed from now onwards in bold), the following can be observed:

- For anger, mouth occlusion decreases the recognition accuracy rate more than eyes occlusion.
- For disgust, eyes occlusion decreases the recognition accuracy rate more than mouth occlusion.
- For fear, mouth occlusion decreases the recognition accuracy rate more than eyes occlusion.
- For happiness, mouth occlusion decreases the recognition accuracy rate more than eyes occlusion.
- For sadness, mouth occlusion decreases the recognition accuracy rate more than eyes occlusion.
- For surprise, eyes occlusion decreases the recognition accuracy rate more than mouth occlusion.

Thus, the eyes region seems to play a more vital role for the recognition of disgust and surprise, while mouth region seems to influence more the recognition accuracy rate of anger, fear, happiness and sadness.

5.1.2. Experiments with non-expert human observers

Non-expert human observers achieved a recognition accuracy rate of 78.7%, 64.0% and 60.7% when no occlusion, eyes and mouth region is present. The confusion matrices calculated when no occlusion, eyes and mouth occlusion is present are shown in Table 3, respectively. Someone can see that mouth occlusion affects the recognition accuracy rate more than eyes occlusion (18.0% and 14.8% decrease in recognition accuracy, respectively). For each facial expression separately: the eyes region seems to play a more vital role for the recognition of disgust and surprise, while mouth region seems to influence more the recognition accuracy rate of anger, fear, happiness and sadness, just like the results acquired from the expert humans.

It has to be taken under consideration that the recognition rate is greatly affected by the expressiveness of the subjects building the database. If the subjects display a facial expression focusing on the mouth region (moving the mouth in a more visible way), then mouth occlusion is estimated to affect more the recognition accuracy rate. This seems to be the case for the database used for the experiments, as well as for most people in general. A visible motion in the eyes region occurs less often than a visible

Table 2
Confusion matrices acquired when expert humans participated in the experiments when no occlusion, eyes and mouth occlusion is present

	an			di			fe			ha			sa			su		
	no	ey	mo	no	ey	mo	no	ey	mo	no	ey	mo	no	ey	mo	no	ey	mo
an	92.3	84.6	**77.8**	0	0	0	0	7.3	30.8	0	0	0	7.7	23.1	38.5	0	0	0
di	0	0	0	100	**92.3**	100	0	0	0	0	0	0	0	0	0	0	0	0
fe	0	0	0	0	0	0	100	92.3	**69.2**	0	7.3	13.7	0	7.7	0	0	15.4	7.3
ha	0	0	0	0	0	0	0	0	0	100	92.3	**86.3**	0	0	0	0	0	0
sa	7.7	15.4	22.2	0	7.3	0	0	0	0	0	0	0	92.3	69.2	**61.5**	0	0	0
su	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	**84.6**	92.3

Table 3
Confusion matrices acquired when non-expert humans participated in the experiments

	an			di			fe			ha			sa			su		
	no	ey	mo	no	ey	mo	no	ey	mo	no	ey	mo	no	ey	mo	no	ey	mo
an	61.7	49.7	**46.7**	16	34.9	18.3	8.9	17.8	10.7	0.2	0.6	4.7	9.5	10.7	18.3	0	0.4	0.6
di	15.8	19.3	16.4	76.1	**43.8**	64.5	7.7	13	9.5	0.6	1.2	1.8	1.8	3	2.4	0	0.2	0.6
fe	9.5	9.9	10.7	3.6	8.3	4.1	70.4	42.6	**36.1**	4.1	5.9	6.5	6.1	7.1	17.2	3.6	10	5.9
ha	0	0	1.8	1.8	2.4	5.9	4.3	14.2	6.5	92.7	89.3	**81.7**	1.8	2.4	8.9	2.4	3	3.6
sa	11.4	16.4	22.1	2.5	3	7.1	3.6	5.9	5.9	2.4	3	4.1	79.9	75.1	**41.9**	2.6	3.6	3
su	1.6	4.7	2.4	0	0	0	5.1	6.5	6.5	0	0	1.2	9.9	1.8	4.7	91.5	**82.8**	8.6

motion in the mouth area, or is at least of a lesser extent. Therefore, someone would expect the mouth region occlusion to affect facial expression recognition the most, something that is confirmed by the results acquired from human observers.

The results acquired when non-expert and expert humans participated in the experiments are in par with the assumptions made in Section 3.1, regarding the facial region that affects the most each facial expression recognition.

5.2. Experiments using the described system

In this Section, the conducted experiments performed using the system described in Section 2 will be presented. Table 4 shows the experimental facial expression recognition accuracies for the databases used without the presence of occlusion as well as the accuracies obtained when mouth or eyes occlusion appears.

5.2.1. Non-occluded experiments

Performing facial expression recognition on unoccluded facial images from the JAFFE database, using Gabor fil-

ters and the DNMF algorithm achieved recognition rates of 88.1% and 85.2%, respectively.

Performing facial expression recognition on unoccluded facial images from the Cohn–Kanade database, using Gabor filters, the DNMF algorithm and SVMs, achieved recognition rates of 91.6%, 86.7% and 91.4%, respectively.

5.2.2. Eyes occlusion experiments

Performing facial expression recognition on facial images under eyes occlusion from the JAFFE database, using Gabor filters and the DNMF algorithm, achieved recognition rates of 83.1% (5% decrease) and 82.5% (2.7% decrease), respectively.

Performing facial expression recognition on facial images under eyes occlusion, using Gabor filters, the DNMF algorithm and SVMs, achieves recognition rates of 86.8% (4.8% decrease), 84.2% (2.5% decrease) and 88.4% (3% decrease), respectively.

5.2.3. Mouth occlusion experiments

Performing facial expression recognition on facial images under mouth occlusion from the JAFFE database, using Gabor filters and the DNMF algorithm, both

Table 4
Accuracy rates and facial expression misclassifications as derived from the experiments conducted

	No occlusion	Eyes occlusion	Mouth occlusion
JAFFE database			
Texture information classification (Gabor filters) (%)	88.1	83.1	81.5
Texture information classification (DNMF algorithm) (%)	85.2	82.5	81.5
Cohn–Kanade database			
Texture information classification (Gabor filters) (%)	91.6	86.8	84.4
Texture information classification (DNMF algorithm) (%)	86.7	84.2	82.9
Shape information classification (SVMs) (%)	91.4	88.4	86.7

achieved recognition rate of 81.5% (6.6% and 3.7% decrease, respectively).

Performing facial expression recognition on facial images under mouth occlusion from the Cohn–Kanade database, using Gabor filters, the DNMF algorithm and SVMs, achieves recognition rates of 84.4% (7.2% decrease), 82.9% (3.8% decrease) and 86.7% (4.7% decrease), respectively.

5.3. The overall effect of occlusion

Overall, for all facial expressions, mouth region occlusion decreases the recognition accuracy by more than 50% when compared to the equivalent eyes occlusion. Therefore, mouth region seems to be more important when it comes to facial expression recognition. As can be seen in Table 4, mouth occlusion affects more the anger, fear, happiness and sadness recognition rate than eyes occlusion. For disgust and surprise eyes occlusion is the one that affects the recognition rate more.

The results achieved from the experiments performed, confirm the assumptions made in Section 3.1. The results are discussed more thoroughly below.

5.4. The effect of occlusion in every facial expression

The effect of occlusion in every facial expression as indicated from our experiments is the following:

5.4.1. Anger
Anger seems to be affected most by mouth occlusion. Eyes occlusion causes a smaller decrease than the equivalent mouth one. Therefore, the mouth region is the most important region when recognizing anger.

5.4.2. Disgust
For disgust recognition, eyes occlusion is the most important one. Eyes occlusion causes a greater decrease than the equivalent mouth one. Therefore, the eyes region is the most important region when recognizing disgust.

5.4.3. Fear
For fear recognition, mouth occlusion is the most important one. Eyes occlusion causes a smaller decrease than the equivalent mouth one. Therefore, the mouth region is the most important region when recognizing fear.

5.4.4. Happiness
For happiness recognition, mouth occlusion is the most important one. Eyes occlusion causes a smaller decrease than the equivalent mouth one. Therefore, the mouth region is the most important region when recognizing happiness.

5.4.5. Sadness
For sadness recognition, mouth occlusion is the most important one. Eyes occlusion causes a smaller decrease

than the equivalent mouth one. Therefore, the mouth region is the most important region when recognizing sadness.

5.4.6. Surprise
For surprise recognition, eyes occlusion is the most important one. Eyes occlusion causes a greater decrease than the equivalent mouth one. Therefore, the eyes region is the most important region when recognizing surprise.

A study regarding the effect of partial occlusion in facial expression recognition has been conducted in [7]. Four facial expressions are examined: anger, happiness, sadness and surprise. The experiments included eyes/mouth/left/right facial region occlusion. The authors claimed that the presence of left/right facial region occlusion has the same results on facial expression recognition accuracy rate, something that was also confirmed by our experiments (Section 4). Regarding each one of the four facial expressions examined, the authors claimed that eyes region occlusion affects more the recognition of anger, happiness and surprise, while the mouth region occlusion affects more the recognition of sadness.

Someone can notice that there are some differences in the conclusions made. This was expected, as our experiments introduced the study of disgust and fear facial expressions. Thus, the recognition of the previously studied facial expressions (anger, happiness, sadness and surprise) is expected to differ now, as more misclassifications between facial expressions should occur.

More specifically, the recognition of anger seems to be affected by mouth occlusion in our experiments, contrary to the experiments performed in [7]. This, however was expected since, anger is the most misclassified facial expression, as can be seen from the estimated confusion matrices. The introduction of disgust and fear results in an increase of the misclassification errors. The misclassification percentages with those facial expressions are among the higher, if not the highest, when it comes to recognition of anger. The recognition of fear depends on mouth facial region as indicated from our experiments. Thus, mouth occlusion results in confusion between fear and anger. Disgust on the other hand is a special facial expression. As previously said (Section 3.1), disgust involves a motion emphasized on the nose. It is more visible on the upper part of the nose, thus making eyes occlusion more important for its recognition. However, some furrows appear on the upper region of the mouth as well, thus making the confusion between anger and disgust possible. Moreover, anger is expressed in many cases as a different in gaze, accompanied with a slight mouth motion. Thus, the mouth's change is more visible when compared to that of the eyes. In general, when all six facial expressions are examined, mouth occlusion plays a more vital role in anger recognition.

The experiments indicated that the recognition of disgust and fear facial expressions is affected by eyes and mouth facial region occlusion, respectively. This is in par

with the assumptions made in Section 3.1, based on FACS and visual observation.

Happiness is claimed to be mostly affected by eyes region in [7]. However, the authors mention that this is valid only when the predefined four facial expressions were examined, because of the virtual lack of motion of the eyebrows during a smile. However, the introduction of disgust and fear may be accompanied by that characteristic as well, thus making the confusion of happiness easier. When facial expressions resembling each other in the eyes region are introduced, the recognition of happiness is now greatly affected by mouth region occlusion. Therefore, in general, happiness facial expression recognition is influenced by mouth occlusion.

Our experiments indicated that the recognition of sadness is affected more from mouth region occlusion. This is in par with the assumption made in [7]. This is expected, as in most cases the specific facial expression is expressed with a subtle change in the eyes region and a more visible one in the mouth region.

The recognition of surprise is greatly affected by eyes occlusion as indicated both from our experiments and in [7]. In many cases, the mouth whose opening is characteristic when it exists, may not change significantly from the neutral state. Thus, the eyes region is the one that is more characteristic when it comes to surprise recognition.

5.5. Facial expressions that are confused when introducing occlusion

The results presented in Table 4 are averaged over all facial expressions and do not provide any information with respect to a particular facial expression. The confusion matrices are presented for the reader to be able to make any conclusions. The observations derived are based on the greatest changes appearing when each type of occlusion is introduced.

5.5.1. Facial expressions that are confused when no occlusion is present

As can be seen in Tables 5–7, the most confused pairs or facial expressions when no occlusion is present, are anger with disgust, anger with sadness, disgust with fear, disgust with sadness, fear with happiness and surprise with fear.

Someone can observe that disgust is one of the most misclassified facial expression, as mentioned above, since it appears in most of the misclassification cases. The most misclassified facial expression is anger, followed in ascending misclassification accuracy rate order, by fear, disgust, sadness and happiness. Surprise is the most correctly classified facial expression.

Anger is confused with disgust as in both of them eyes tend to move down and the mouth's width becomes smaller. Anger is confused with sadness as the eyes also move down and the lips are pressed tightly in both cases. Disgust can be confused with fear in cases when the eyebrows come closer to each other. In sadness, the mouth shape can be similar to the one in disgust forming a bow that arches. In fear and happiness, the mouth forms a specific shape, resembling a smile and in both surprise and fear, the eyes open wide. Therefore, the above mentioned misclassifications are observed.

5.5.2. Facial expressions that are affected when eyes occlusion is present

As can be seen in Tables 5–7, eyes occlusion seems to introduce more misclassifications among facial expressions. These misclassifications include happiness with disgust and surprise with disgust.

When eyes occlusion is present, the virtual mask hides the nose part that is stressed in disgust, thus making the mouth region similar to a smile when the lips are closed. Therefore, happiness can be misclassified as disgust. Surprise can be mistaken for disgust since in many cases the poser just opens his eyes without opening the mouth. In those cases, the eyes are the most discriminative region and thus their occlusion leads to misclassification.

5.5.3. Facial expressions that are affected when mouth occlusion is present

As can be seen in Tables 5–7, mouth occlusion seems to introduce more misclassifications among facial expressions. These misclassifications include anger with fear, happiness with sadness and surprise with disgust.

Anger can be misclassified as fear since in come cases the eyes tend to approach each other in both facial expressions. Happiness and sadness appear almost the same when only the mouth is moved, something that generally happens when a poser is used (someone that delib-

Table 5
Confusion matrices regarding the Cohn–Kanade database when no occlusion (91.6%), eyes (86.8%) and mouth occlusion (84.4%) is present using Gabor filters

%	an			di			fe			ha			sa			su		
	no	ey	mo	no	ey	mo	no	ey	mo	no	ey	mo	no	ey	mo	no	ey	mo
an	82	79.3	**70**	0	0	0	3.5	3.8	6.4	3.5	4	6.7	3.3	4.8	5.6	0	0.7	0.8
di	4	6	6.7	94.1	**81.5**	85.1	0	0	0	0	0	0	0	0	0	0	1.5	1.1
fe	6	6	10	3	7.4	7.4	93	92.5	**87.2**	5.9	7.2	10.1	3.7	6.7	7	3.3	4.8	4.8
ha	4	4.7	6.7	0	0	0	3.5	3.7	6.4	90.6	88.2	**83.2**	0	0	0	0	0	0
sa	4	4	6.6	2.9	11.1	7.5	0	0	0	0	0	0	93	88.5	**87.4**	0	2.1	0
su	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	96.7	**90.8**	93.3

Table 6

Confusion matrices regarding the Cohn–Kanade database when no occlusion (86.7%), eyes (84.2%) and mouth occlusion (82.9%) is present using the DNMF algorithm

%	an			di			fe			ha			sa			su		
	no	ey	mo	no	ey	mo	no	ey	mo	no	ey	mo	no	ey	mo	no	ey	mo
an	78.3	74	**69.3**	0	2.2	2.2	0	0	0	0	0	0	9.6	9.6	11	0	0	0
di	0	1.3	1.3	82	**77.8**	80	13	13	14.2	0	0.5	0	0	1.1	0.3	0	0.8	0.8
fe	17	20	24	18	20	20	76	74	**71**	3.5	4.5	5.6	0	0	0	0	0	0
ha	4.7	4.7	5.4	0	0	0	15	13	13.3	96.5	95	**93.1**	0	0	0	0	0	0
sa	0	0	0	0	0	0	0	0	1.5	0	0	1.3	90.4	89.3	**88.7**	3	4.4	4.1
su	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	97	**94.8**	95.1

Table 7

Confusion matrices regarding the Cohn–Kanade database when no occlusion (91.4%), eyes (88.4%) and mouth occlusion (86.7%) is present using SVMs

%	an			di			fe			ha			sa			su		
	no	ey	mo	no	ey	mo	no	ey	mo	no	ey	mo	no	ey	mo	no	ey	mo
an	86.9	82	**80.7**	2.5	10.6	9.6	0	0	0	0	0	0	8	9.2	11.8	0	2.4	0
di	4.8	8	7.3	86.7	**83.8**	85.2	0	0	0	0	0.5	0	0	1.1	0	0	1.6	2.2
fe	0	0	1.3	0	0	0	92.9	91.9	**87.3**	4.3	5.9	6.1	0	0	0	3.2	3.6	3
ha	0	0	0	0	0	0	3.6	4.1	4.3	95.7	93.6	**90.9**	2.5	3	5.4	0	0	0
sa	8.3	10	10.7	6	5.6	5.2	3.5	4	8.4	0	0	3	89.5	86.7	**82.8**	0	0	1.5
su	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	96.8	**92.4**	93.3

erately poses for a facial expression instead of displaying a spontaneous one). This also seems to be the case for the presence of misclassification between surprise and disgust, because obviously surprise is expressed in the misclassified cases by opening the mouth that the virtual mask covers. The virtual mask also covers the nose, thus making the discrimination of those two facial expressions more difficult.

5.5.4. Experiments regarding the JAFFE database

The above mentioned experiments were also conducted using the JAFFE database. The conclusions extracted were in par with the ones extracted using the Cohn–Kanade database. However, due to space limitations they are not presented analytically.

6. Conclusions

Facial expression recognition in the presence of mouth,eyes, left and right facial region occlusion has been investigated to determine the part of the face that contains most discriminant information for facial expression recognition. Gabor wavelets, the DNMF algorithm and shape-based SVMs have been used to achieve recognition. The experiments have been performed using human observers and the algorithms proposed, to draw any conclusions. The results showed that left/right facial region occlusion does not affect the recognition accuracy rate, indicating that both facial regions possess similar discriminate information. The results indicate that mouth occlusion, in general, causes a greater decrease in facial expression recognition than the equivalent eyes one. Mouth occlusion affects more anger, fear, happiness and sadness, while eyes occlusion the remaining disgust and

surprise. The experiments indicated that the results obtained from experiments conducted with humans and the equivalent ones conducted using the presented system are in par. Future research includes the application of the algorithms on MMI database as well as the repetition of the study of different types of occlusion on facial expression recognition.

Appendix A. Shape-based information extraction

The tracking accuracy affects the recognition performance accuracy. When the tracking result is not correct (i.e. the deformed Candide grid does not correspond to the deformation depicted on the frame), the recognition accuracy is greatly affected. However, it is out of scope of this study to examine the performance of the tracker. Thus, the acquired deformed grids were manually corrected to provide the correct information for further study of facial expression recognition.

A.1. Tracking system initialization

The initialization procedure is performed in a semi-automatic way in order to attain reliability and robustness of the initial grid displacement. In the beginning, the Candide wireframe grid is initially placed on the facial image depicted at the first frame. The grid is in its neutral state.
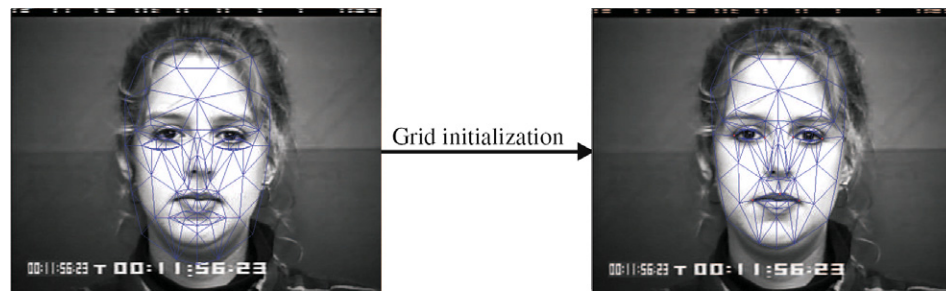
Fig. A.1. Result of initialization procedure when 7 Candide nodes are placed by the user on a facial image.

The user has to manually select a number of point correspondences that are matched against the facial features of the actual face image. The most significant nodes (around the eyes, eyebrows and mouth) should be chosen, since they are responsible for the formation of facial deformations modelled by FACS. It has been empirically determined that 5–8 node correspondences are enough for a good model fitting. These correspondences are used as the driving power which deforms the rest of the model and matches its nodes against face image points. The result of the initialization procedure, when 7 nodes (4 for the inner and outer corner of the eyes and 3 for the upper lip) are placed by the user, can be seen in Fig. A.1.

A.2. Model based tracking

The algorithm, initially fits and subsequently tracks the Candide facial wireframe model in image sequences containing the formation of a dynamic human facial expression from the neutral state to the fully expressive one. Wireframe node tracking is performed by a pyramidal variant of the well-known Kanade–Lucas–Tomasi (KLT) tracker [40]. The loss of tracked features is handled through a model deformation procedure that increases the robustness of the tracking algorithm. If needed, model deformations are performed by mesh fitting at the intermediate steps of the tracking algorithm. Such deformations provide robustness and tracking accuracy.

The facial model is assumed to be a deformable 2D mesh model. The facial model elements (springs) are assumed to have a certain stiffness. The driving forces that are needed i.e. the forces that deform the model, are determined from the point correspondences between the facial model nodes and the face image features. Each force is defined to be proportional to the difference between the model nodes and their corresponding matched feature points on the face image. If a node correspondence is lost, the new node position is the result of the grid deformation. The tracking algorithm provides a dynamic facial expression model for each image sequence, which is defined as a series of frame facial expression models, one for each image frame. An example of the deformed Candide grids produced for each facial expression is presented in Fig. A.2.
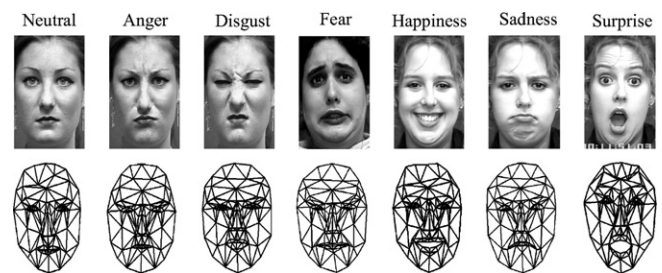

Fig. A.2. An example of the produced deformed Candide grid for each facial expression.

References

[1] A. Pentland, T. Choudhury, Face recognition for smart environments, IEEE Computer 33 (2) (2000) 50–55.

[2] M. Pantic, L. Rothkrantz, Toward an affect-sensitive multimodal human–computer interaction, Proceedings of the IEEE 91 (9) (2003) 1370–1390.

[3] P. Ekman, W. Friesen, Emotion in the Human Face, Prentice Hall, New Jersey, 1975.

[4] P. Ekman, Unmasking the Face, Cambridge University Press, Cambridge, 1982.

[5] T. Kanade, J. Cohn, Y. Tian, Comprehensive database for facial expression analysis, in: Proceedings of IEEE International Conference on Face and Gesture Recognition, 2000, pp. 46–53.

[6] M. Pantic, L. Rothkrantz, Automatic analysis of facial expressions: the state of the art, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (12) (2000) 1424–1445.

[7] F. Bourel, C. Chibelushi, A. Low, Recognition of facial expressions in the presence of occlusion, in: Proceedings of the Twelfth British Machine Vision Conference, vol. 1, Manchester, UK, 2001, pp. 213–222.

[8] M.J. Lyons, J. Budynek, S. Akamatsu, Automatic classification of single facial images, IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (12) (1999) 1357–1362.

[9] B. Duc, S. Fischer, J. Bigün, Face authentication with Gabor information on deformable graphs, IEEE Transactions on Image Processing 8 (4) (1999) 504–516.

[10] L. Chengjun, Gabor-based kernel PCA with fractional power polynomial models for face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (5) (2004) 572–581.

[11] W. Fellenz, J. Taylor, N. Tsapatsoulis, 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.

[12] M.S. Bartlett, G. Littlewort, I. Fasel, J.R. Movellan, Real time face detection and facial expression recognition: development and applications to human computer interaction, in: Workshop on Computer

Vision and Pattern Recognition for Human–Computer Interaction, 2003, pp. 1295–1302.

[13] Z. Zhang, M. Lyons, M. Schuster, S. Akamatsu, Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron, in: Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition, Nara Japan, 1998, pp. 454–459.

[14] M.J. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, Coding facial expressions with Gabor wavelets, in: Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition, 1998, pp. 200–205.

[15] G. Donato, M.S. Bartlett, J.C. Hager, P. Ekman, T.J. Sejnowski, Classifying facial actions, IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (10) (1999) 974–989.

[16] S. Zafeiriou, A. Tefas, I. Buciu, I. Pitas, Exploiting discriminant information in nonnegative matrix factorization with application to frontal face verification, IEEE Transactions on Neural Networks 17 (3) (2006) 683–695.

[17] I. Buciu, I. Pitas, A new sparse image representation algorithm applied to facial expression recognition, in: MLSP, Sao Lufs, Brazil, 2004.

[18] I. Buciu, I. Pitas, NMF, LNMF and DNMF modeling of neural receptive fields involved in human facial expression perception, Journal of Visual Communication and Image Representation 17 (5) (2006) 958–969.

[19] S.Z. Li, X.W. Hou, H.J. Zhang, Learning spatially localized, parts-based representation, in: CVPR, Kauai, HI, USA, 2001, pp. 207–212.

[20] S. Krinidis, I. Pitas, Statistical analysis of facial expressions for facial expression synthesis, IEEE Transactions on Multimedia, submitted for publication.

[21] P. Michel, R. Kaliouby, Real time facial expression recognition in video using support vector machines, in: Proceedings of Fifth International Conference on Multimodal Interfaces, Vancouver, British Columbia, Canada, 2003, pp. 258–264.

[22] I. Kotsia, I. Pitas, Facial expression recognition in image sequences using geometric deformation features and support vector machines, IEEE Transactions on Image Processing 16 (1) (2007) 172–187.

[23] I. Buciu, I. Kotsia, I. Pitas, Facial expression analysis under partial occlusion, in: IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2005, Philadelphia, 2005.

[24] L. Wiskott, J.-M. Fellous, N. Kruger, C. von der Malsburg, Face recognition by elastic bunch graph matching, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 775–779.

[25] M. Collins, R.E. Schapire, Y. Singer, Logistic regression, adaboost and bregman distances, Computational Learning Theory (2000) 158–169.

[26] D.D. Lee, H.S. Seung, Algorithms for non-negative matrix factorization, in: NIPS, 2000, pp. 556–562.

[27] D. Lee, H. Seung, Learning the parts of objects by non-negative matrix factorization, Nature 401 (1999) 788–791.

[28] D. Donoho, V. Stodden, When does non-negative matrix factorization give a correct decomposition into parts? Advances in Neural Information Processing Systems, 17 (2004).

[29] I. Kotsia, S. Zafeiriou, I. Pitas, A novel discriminant non-negative matrix factorization algorithm with applications to facial image characterization problems, IEEE Transaction on Forensics and Security. Part 2 2 (3) (2007) 588–595.

[30] I. Kotsia, S. Zafeiriou, I. Pitas, Texture and shape information fusion for facial expression and facial action unit recognition, Pattern Recognition 41 (3) (2008) 833–851.

[31] V. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.

[32] C.W. Hsu, C.J. Lin, A comparison of methods for multiclass Support Vector Machines, IEEE Transactions on Neural Networks 13 (2) (2002) 415–425.

[33] J. Weston, C. Watkins, Multi-class Support Vector Machines, in: Proceedings of ESANN99, Brussels, Belgium, 1999.

[34] C.J.C. Burges, A tutorial on Support Vector Machines for pattern recognition, Data Mining and Knowledge Discovery 2 (2) (1998) 121–167.

[35] M. Pantic, L.J.M. Rothkrantz, Expert system for automatic analysis of facial expressions, Image and Vision Computing 18 (11) (2000) 881–905.

[36] M. Pantic, M.F. Valstar, R. Rademaker, L. Maat, Web-based database for facial expression analysis, in: IEEE International Conference on Multimedia and Expo, ICME 2005, Amsterdam, The Netherlands, 2005.

[37] S. Asteriadis, N. Nikolaidis, A. Hajdu, I. Pitas, An eye detection algorithm using pixel to edge information, in: Proceedings of Second IEEE-EURASIP International Symposium on Control, Communications, and Signal Processing (ISCCSP 2006), Marrakech, Morocco, 2006.

[38] I. Cohen, N. Sebe, S. Garg, L.S. Chen, T.S. Huanga, Facial expression recognition from video sequences: temporal and static modelling, Computer Vision and Image Understanding 91 (2003) 160–187.

[39] T.M. Cover, Learning in Pattern Recognition, Academic Press, 1969.

[40] J.Y. Bouguet, Pyramidal implementation of the Lucas–Kanade feature tracker, Technical Report, Intel Corporation, Microprocessor Research Labs, 1999.