Dynamic Facial Emotion Recognition Oriented to HMI Applications

Abstract

As part of a multimodal animated interface previously presented in [38], in this paper we describe a method for dynamic recognition of displayed facial emotions on low resolution streaming images. First, we address the detection of Action Units of the Facial Action Coding System upon Active Shape Models and Gabor filters. Normalized outputs of the Action Unit recognition step are then used as inputs for a neural network which is based on real cognitive systems architecture, and consists on a habituation network plus a competitive network. Both the competitive and the habituation layer use differential equations thus taking into account the dynamic information of facial expressions through time. Experimental results carried out on live video sequences and on the Cohn-Kanade face database show that the proposed method provides high recognition hit rates.

Keywords: Active Shape Models,Cognitive Systems, Facial Emotion Recognition,Gabor filters, Human-Machine Interfaces

I. Introduction

Over the last few years, a growing interest has been observed in the development of new interfaces in Human-Machine Interaction (HMI) that allow humans to interact with machines in a natural way. Ideally, a man-machine interface should be transparent to the user, i.e. it should allow the user to interact with the machine without requiring any cognitive effort. What is sought is that anyone can use machines and computers in their daily lives, even people not very familiar with the technology, making it easier both the way people communicate with devices and the way devices present their data to the user.

Considering human to human interaction as a reference, one of the main sources of information interchange resides in the face gesture capabilities. People are extremely skilled in interpreting the behaviour of other persons according to changes in their facial characteristics. For that reason, face-to-face communication is one of the most intuitive, robust and effective ways of communication between humans. Using facial expressions as a communication channel in HMI is one way of making use of the inherent human ability to express themselves via changes in their facial appearance. Thus, the robustness and naturalness of human-machine interaction is greatly improved [1].

In the past, a lot of effort was dedicated to recognizing facial expressions and emotions in still images. Using still images has the limitation that they are usually captured in the apex of the expression, where the person is performing the indicators of emotion more markedly. In day-to-day communication, however, the apex of facial expressions are rarely shown, unless for very specific cases. Also, research has demonstrated that people perceive others' emotions in a continuous fashion, and that merely representing emotions in a binary (on/off) manner is not an accurate reflection of the real human perception of facial emotions [37]. More recently, researches have changed their efforts towards modelling dynamic aspects of facial expression. This is because differences between emotions are more powerfully modelled by studying the dynamic variations of facial appearance, rather than just as discrete states. This is especially true for natural expressions without any deliberate exaggerated posing.

It is also necessary to distinguish between two concepts which, although interrelated, should be independently considered in the field of facial recognition: facial expression recognition and facial emotion recognition. Expression recognition refers to those systems whose purpose is the detection and analysis of facial movements, and changes in facial features from visual information. On the other hand, to perform emotional expression recognition, it is also necessary to associate these facial displacements to their emotional significance, given that there will be certain movements associated with more than one emotion. This paper focuses on the latter: the automatic visual recognition of emotional expressions in dynamic sequences. However, the proposed method is divided into two different stages: the first one focuses on the detection and recognition of small groups of facial muscle actions and, based on the output, the second determines the associated emotional expression.

The proposed system is part of a multimodal animated head which was previously developed by our research group. The developed animated head has proven to be a perfect complement for a humanoid robotic construction or even a substitute for a hardware equivalent solution [38].

The rest of the paper is organized as follows. Related work is reviewed in section II. The proposed approach for facial activity detection through computer vision techniques is presented in section III. The method for emotion recognition is described in section IV. Experimental results are presented in V. Conclusions and future work are described in section VI.

II. Related work

The common procedure for the recognition of facial expressions can be divided into three steps: image segmentation, feature extraction and feature classification. Image segmentation is the process that allows an

image to be divided into groups of pixels that share common characteristics. The aim of segmentation is to simplify image content in order to analyse it easily. There are many different techniques for image segmentation that use either colour [18] or grey scale [19] information. As far as facial image segmentation is concerned, one of the most widespread techniques is the Viola and Jones algorithm [19], due to its robustness against face variations and its computational speed. Image segmentation for facial recognition can also include a normalization step, which eliminates rotations, translations and scaling of captured faces.

The goal of the feature extraction step is to obtain a set of facial features that are able to describe the face in an effective way. An optimum set of facial features should be able to minimize variations between characteristics that belong to the same class and maximize the differences between classes. There exist different approaches for feature extraction. For example, in [20] optical flow is used for modelling muscular activity through the estimation of the displacements of a group of characteristic points. The main disadvantage of optical flow is that it is highly sensitive to lighting variations and not very accurate if applied to low resolution images. For that reason, it is usually used along with other methods, such as volume local binary patterns VLBPN [21]. Analysis of facial geometry is another method used for feature extraction, where the geometric positions of certain facial features are used to represent the face. In [22] the position of manually marked facial features are used to represent the face, and then particle filters are used to study variations in those characteristics.

Changes in facial appearance are also used for facial feature extraction. More common techniques are Principal Component Analysis (PCA) [23], Linear Discriminant Analysis (LDA) [24] and many others, such as Gabor filters [25]. Gabor filters are widely used for feature extraction as they allow facial characteristics to be extracted precisely and invariantly against changes in luminosity or pose [26]. However, Gabor filters demand relatively high computational costs, which get higher as the image resolution increases. For that reason, they are usually used in small regions of the face after a previous segmentation step. Recently, Local Binary Patterns have begun to be used for image segmentation as they require lower computational costs than Gabor filters [27], although they are less precise than the latter when the number of extracted features increases [28].

Both geometry based and appearance based methods have their advantages and disadvantages. For that reason, recent proposals try to combine both techniques for feature extraction in what are called deformable models. Deformable models are built from a training database and are able to extract both geometric and appearance data from a facial image. There are many different approaches to building deformable models, but those that are more commonly used are Active Shape Models (ASMs) [31] and Active Appearance Models (AAMs) [29]. Active shape and appearance models were developed by Cootes and Taylor [30]. They are based on the statistical information of a point distribution model (shape) which is fitted to track faces in images using the statistical information of the image texture. AAMs use all texture information present in the image, whereas ASMs only use texture information around model points. As for feature tracking, ASMs have shown to be more robust than AAMs if used in real time feature extraction [6]. This is because ASM tracking is done in precise image regions which have previously been modelled during a training stage (around model points), whereas AAM tracking is performed using the information present in the whole image.

Once facial features had been extracted, there are many different approaches for their classification in order to categorize and recognize facial expressions, such as Support Vector Machines [32] or rule based classifiers [33]. For real time applications, the analysis should be done over a set of facial features obtained over time from video sequences, using neural networks [34], Hidden Markov Models [35] or Bayesian networks [36]. Among classification techniques, two big groups can also be distinguished: those that try to classify emotional expressions by analysing the face as a whole [13] [15], and those which try to detect small facial expressions separately, and then combine them for complex expression classification [32] [34].

III. Action Unit detection

There are several methods for measuring and describing facial muscular activity, and the changes which that activity produces on facial appearance. From these, the Facial Action Coding System (FACS) [9] is the most widely used in psychological research, and has been used in many approaches for automatic expression recognition [29] [34]. As a final goal, FACS seeks to recognize and describe the so-called Action Units (AUs), which represent the smallest group of muscles that can be contracted independently from the rest and produce momentary changes in facial appearance. These Action Units can, in combination, describe any global behavior of the face.

The developed facial expression recognition system first looks for the automatic recognition of FACS Action Units by using a combination of Active shape models (ASMs) [4] and Gabor filters [5]. ASMs are statistical models of the shape of objects, which are constrained to vary only in ways learned during a training stage of labelled samples. Once trained, the ASM algorithm aims to match the shape model to a new object in a new image. ASMs allow accurate face segmentation in images, in order to perform a more detailed analysis of the different facial regions. Based on the work developed by [6], a shape model of 75 points has been used to track and segment human faces from images captured with a webcam as seen in Fig.1.

The human face is extremely dynamic, and presents variations in its shape due to the contraction of facial muscles when facial expressions are performed. ASMs are able to accurately track faces, even in the presence of facial shape variations (mouth opening, eyebrows rising, etc.). However, as ASM training and tracking is based on a local profile search, it is difficult to train the model to track transient facial features such as wrinkles, or to take into account the presence of elements such as glasses, beard or scarves. The amount of variations allowed to a shape during the training stage dramatically affects tracking speed and accuracy [4]. For that reason, facial variations permitted in the images used during the training stage have been constrained to permanent features (width, height, eyebrows or nose and mouth position). However, many facial expressions involve the apparition of transient facial features that may only be visible while the expression is being performed. (E.g. the vertical wrinkle that appears between the eyebrows when frowning). Since the ASMs are not able to detect transient characteristics of facial expressions, ASM tracking has been used to track the face and detect changes in permanent facial features. It has also been used for face segmentation in smaller regions, which are processed using a bank of Gabor filters that let us look for transient facial features.

III.A. AU – related feature extraction

The following sections show the proposed method used for the detection of facial features when performing facial expressions. Four different regions of the face have been considered for feature detection: forehead and eyebrows, eyes, nose, mouth and lower and lateral sides of the mouth. For each region, the most distinctive AUs associated to each of the six universal emotional expressions, as described in FACs (joy, anger, fear, disgust, surprise and sadness) have been considered.

Forehead and eyebrow region

The action units considered for this region are AU1, AU2 and AU4 (see Table I). The contraction of the frontalis pars medialis makes the eyebrows rise and be slightly pulled together, with the formation of horizontal wrinkles in the central region of the forehead (coded in FACS as AU1), while the contraction of the frontalis pars lateralis muscle also originates horizontal wrinkles and eyebrow rising but in the outer regions of the forehead (AU2). The contraction of the corrugator supercilii muscle pulls the inner region of

the eyebrows together and downwards. Also, the eyelids are slightly pushed together. As a result, vertical wrinkles appear in the space between the eyebrows (AU4).

Variations in eyebrow position have been determined by computing the distance between the eyebrows and the distances between the inner and outer corners of the eyebrows and the line connecting the corners of the eyes. These distances are computed using shape model points' position as shown in

Fig.2. Thus, if AU1 or AU2 are active, there will be an increase in the h_1 and h_2 distances respectively, while, if AU4 is active, the distance d_1 will decrease.

However, as computed ASM points' position may suffer from noise, the proposed method complements those measures with the detection of other characteristics, such as the appearance of wrinkles in the upper region of the face. The detection of transient features, such as wrinkles, has a number of problems resulting from reflective and reflexive properties of the skin, the presence of facial hair, skin tone, etc. These problems can cause the conditions in which certain facial features appear to vary even for the same individual. To detect these features in as invariant and generic a way as possible, Gabor filters have been used. Gabor filters can exploit salient visual properties, such as spatial localization, orientation selectivity, and spatial frequency characteristics, and are quite robust against illumination changes [7]. A 2D Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave. The considered filter equations in spatial ($g_{\sigma,F,\theta}$) and frequency ($G_{\sigma,F,\theta}$) domains are shown in equations (1) and (4):

$$g_{\sigma,F,\theta}(x,y) = {g'}_{\sigma}(x,y)e^{j2\pi F x'} \qquad (1)$$

where:

$${g'}_{\sigma}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} exp\left[-\frac{1}{2}\left(\left(\frac{x'}{\sigma_x}\right)^2 + \left(\frac{y'}{\sigma_y}\right)^2\right)\right] \qquad (2)$$

$$x' = x\cos\theta + y\sin\theta \; y' = x\sin\theta + y\cos\theta \qquad (3)$$

and

$$G_{\sigma,F,\theta}(x,y) = exp\left[-\frac{1}{2}\left(\frac{(u'-F)^2}{\sigma_u^2} + \frac{{v'}^2}{\sigma_v^2}\right)\right] \qquad (4)$$

where:

$$\sigma_u = \frac{1}{2\pi\sigma_x} \quad , \quad \sigma_v = \frac{1}{2\pi\sigma_y} \qquad (5)$$

$$u' = u\cos\theta + v\sin\theta \quad , \quad v' = -u\sin\theta + v\cos\theta \qquad (6)$$

F is the modulated frequency, which defines the frequency of the complex sinusoid, θ *is* the orientation of the Gaussian function and σ_x and σ_y define the Gaussian function scale on both axes.

Gabor filters are located both in the spatial and frequency domains, which makes them suitable for relating frequency characteristics to their position in the image space. One advantage of using Gabor filters resides in the possibility of orienting them so that facial characteristics with a known orientation can be highlighted. For the detection of wrinkles in the forehead region, shape points location is used to extract the face region between the eyebrows and the base of the hair on the forehead. The extracted region is resized

to 32x32 pixels size and gray-scaled. That region size coincides with Gabor filter sizes, which are computed off-line in order to minimize filtering processing time. The extracted region is then convolved with the corresponding filters in frequency domain. To achieve more robustness against lighting conditions, all filters are turned to zero DC:

$$f(x,y)\otimes g(x,y) = T^{-1}\big(F(u,v)G(u,v)\big) \quad with \;\; G(0,0)=0 \tag{7}$$

After applying the filter, the resulting phase component of the convolution is thresholded. Whenever possible, we have worked with the phase component, since it is the one that suffers from less variation due to different lighting conditions or the presence of facial hair or shiny skin. Since Gabor filtering causes the appearance of artefacts at the corners of the images, a region larger than the interest area has been considered. The corner pixels of the filtered image pixels are then masked out, given that they do not provide information. Masking also allows image regions to be easily segmented with irregular shapes.

In the case of the forehead, three regions are considered in the resulting filtered and masked image: one central and two lateral. If the average image intensity in the central region exceeds a threshold, it is encoded as the presence of wrinkles, contributing to the activation of AU1. Similarly, the mean intensity in the lateral regions allows wrinkles that are considered to contribute to the activation of AU2, for right and left eyebrows respectively, to be detected. Fig.3 shows the detection of wrinkles for AU1, AU2 and the combination AU1+AU2. It has to be noted that filtering and thresholding calculations are done with floating point precision, although the resulting image is shown as an 8 bit grey image for visualization purposes.

The same procedure has been applied for the AU4, where the region of interest is located between the two eyebrows and the top of the nose (Fig.3). In this case a Gabor filter with horizontal orientation has been used, as the wrinkles generated by AU4 activation have a vertical orientation (filter parameters: F = 0.07 pixels^{-1}, θ = 0 rad, σ_x= 7.9, σ_y = 4 and size 32x32).

Eyes region

Action Units considered in this region are AU5, AU6, AU7, AU43, AU45 and AU46 (Table II).

In the proposed approach, AU detection in the region of the eyes is done according to three characteristics: two transitional, the position of the eyelids and the presence or absence of wrinkles on the outer edges of the eyes, and one permanent, the amount of visible sclera.

The opening and closing of the eyelids is determined through the distance between eyelids computed from the shape model points. The amount of visible sclera is also used for eye aperture detection. To compute this, a filter of parameters F = 0.12 pixels^{-1}, θ = 0 rad, σ_x = 3.2 pixels and σ_y = 2 pixels and size 32x32 pixels has been applied to each extracted eye region. The resultant phase image is masked using an ellipsoidal mask, and thresholded to extract the two biggest contours of visible sclera. The size of the biggest contour contributes to determining whether the eyes are open, partially open or completely closed, as shown in Fig.4.

A partially visible sclera (Fig.4, middle column) could indicate that either AU6 or AU7 are active. To distinguish those AUs, an area near the outer edges of the eyes is extracted and filtered with a Gabor filter of parameters F = 0.12pixels^{-1}, θ = 1.5 rad, σ_x = 4.2 pixels and σ_y= 1pixels. If the resulting image intensity is high enough to consider the presence of wrinkles, AU6 is considered active, otherwise AU7 is active.

Nose region

The region of the nose has distinctive transient features related to the activation of AU9 and AU10 (Table III).

Wrinkle detection in the nose region follows a similar schema to the ones described in previous sections. In this case, three regions are considered: the region of the nose between the eyes (Fig.5 region A), and on both sides of the nose base (Fig.5 regions B and C). Wrinkles in both regions indicate the activation of AU9, while, if they only appear near the base of the nose, they indicate the activation of AU10.

Mouth region

The main action units whose activation produces visual changes in the region of the mouth are AU10, AU12, AU15, AU16, AU17, AU25 and AU26 (Table VI). The main facial characteristics in the mouth region are the position of the lips and the teeth.

Mouth opening and closing is determined based on the area of the ellipse defined by the distances between shape points in the upper and lower lips and on the mouth corners. These points are also used to detect the mouth corners being pulled down. The rising and falling of the lip corners is also determined by shape model points located in the mouth corners. Along with ASM points' position, two Gabor filters are used in the mouth region to complement AU detection with parameters F = 0.9 pixels^{-1}, θ = 1.57 rad, σ_x = 4 pixels,σ_y = 1.3 pixels and F = 0.21 pixels^{-1}, θ = 2.74, σ_x = 1.9 pixels, σ_y = 0.7 pixels, both 32x32 pixels size. Nine different regions are considered in the resulting image, as shown in Fig.6. The combination of the thresholded intensities on those sub-regions contributes to the activation of the different AUs. For example, if the intensity in sub-region 1 in Fig.6 exceeds a threshold, then AU12 (left) is considered to be active, whereas, if the intensity in sub-regions 7, 8 and 9 exceed a threshold, then AU17 is considered active. The result of filtering in the mouth region when performing other Action Units are also shown in Fig.6.

AU16 and AU10 make the upper and lower teeth, respectively, become visible when activated in conjunction with AU25; so only the combination of AU25 with those action units has been considered. For the detection of the teeth, a filter similar to that used for the sclera has been used, with parameters F = 0.07 pixels^{-1}, θ = 0 rad, σ_x = 7.9 pixels, σ_y = 4 and 32x32 pixels size.

Lower and lateral sides of the mouth region

In this region, distinctive wrinkles appear as the result of the activation of AU12, AU15 and AU17 (Table V). Activation of AU12 causes a wrinkle to appear from the nostrils to the corners of the mouth. Both its angle with respect to the sagittal plane of the face and its depth determine the intensity of AU12. To detect this feature, a face region bounded by the cheeks, nostrils and the angle of the mouth is extracted and filtered with the parameters F = 0.21 pixel^{-1}, θ = 2.3 rad, σ_x = 1.9 and σ_y = 0.7 and size 32x32 pixels for the right wrinkle, as shown in Fig.7, and θ = 0.84 rad for the left one.

As described before, along with making the lips adopt a convex arc form, both AU 15 and 17 produce wrinkles in the chin region when active. To complement the detection of both AUs, the chin region has been filtered with parameters F = 0.52 pixels^{-1}, θ = 0 rad, σ_x= 0.9 and σ_y = 1.5 and size 32x32, following a similar process to the one described for the rest of the wrinkle detection.

III.B. Survey of considered parameters for AU detection

Table I to Table V show the considered parameters for the detection of each of the Action Units in the five facial regions.

III.C. Computation of AU activation

The activation value for each Action Unit is then computed as the weighted sum of all normalized features it depends on (see Table I to Table V).

Equation (8) shows the calculation of AU activation where c_{ni} are the normalized features value and P_{ni} are the weights for each feature. Weights have been manually adjusted so they give more importance to those features obtained from the filtering process, rather than from the ASM points' position, as the latter are more noise sensitive.

$$AU_i = P_{1i}c_{1i} + P_{2i}c_{2i} + .. + P_{ni}c_{ni} \quad (8)$$

where

$$\sum_{j=1}^{n} P_{ji} = 1 \quad (9)$$

Lastly, outputs obtained for each Action Unit are low-pass filtered in order to avoid fluctuations in their values due to noise, as shown in

$$c_{ji}[n] = c_{ji}[n-2] + A(c_{ji}[n-1] - c_{ji}[n-2]) \quad (10)$$

IV. Emotional expression recognition

The previously described Action Unit recognition method allows different facial movements that happen when performing facial expressions to be determined and categorized. In human-computer and human-robot interaction, gesture recognition could be useful *per se*, as it provides humans natural ways to communicate with devices in terms of simple non-verbal communication. However, when developing more complex cognitive systems intended for natural interaction with humans, facial expressions should be endowed with a meaning.

Considering Ekman's work on facial emotions [9], emotions are displayed in the face as movements which could be visually categorized in terms of Action Units. For each universal emotional expression, the Facial Action Coding System is able to parameterize it with the corresponding set of AUs that are active. Based on Ekman's work, normalized outputs of the Action Unit recognition stage are used as inputs of a neural network which performs emotion recognition. The architecture is based on real cognitive systems and consists of a habituation based network plus a competitive based network. Fig.8 shows an overview of the network architecture, while the following subsections describe the functionality and purpose of each layer in detail.

IV.A. Input filter

Activation levels of the detected Action Units serve as stimuli for the input buffer in the *F0* level (Fig.8). Each stimulus is normalized into the [0,1] range and its salience is modulated with a sensitivity gain K_i. Those gains allow certain inputs to be inhibited in case there is a facial feature that causes errors during recognition (e.g. sunglasses, scarfs). In those cases, inputs that are known to be erroneous can be deactivated, and the network could still work using the rest of the detected facial features. Timing is one of the main aspects to be taken into account when performing facial expressions, given that not all expressions

are performed at the same speed. Some facial expressions could be performed quicker than the network is able to process them, thus a filtering stage is also applied to the inputs. Thus, the input's effective salience will decay more slowly over time.

Equation (11) shows how sensitivity gain and filtering are applied to the inputs, where I_i are the input stimuli, S_i is the filtered input stimuli, K_i the sensitivity gain and a the attenuation rate. Thus, the next level's neuron activity decays with a rate a when there is no input.

$$\frac{dS_i}{dt} = -aS_i + aK_iI_i \qquad (11)$$

IV.B. Habituation layer

The proposed network has habituation capabilities, that is, it loses interest in permanent stimuli over time. Stimuli decay due to habituation allows the network to dynamically adapt itself against permanent inputs not caused by facial expressions, but by facial features present in some users even in rest position. It could be possible, for example, that a user's nasolabial wrinkle is pronounced even in rest position because of his/her physiognomy, or that permanent aging wrinkles are present in the user's face. By using a habituation layer, those continuous stimuli will lose preponderance over time, allowing the rest of the facial features to acquire importance. Habituation is performed by multiplying the input stimuli with a gain that is actualized over time. The gain computation is based on Grossberg's Slow Transmitter Habituation and Recovery Model [10]:

$$\frac{dg_i}{dt} = E(1 - g_i) - FS_ig_i \qquad (12)$$

where S_i is the filtered stimulus and g_i is the habituation gain for that stimulus. When a stimulus is active, habituation gain decreases from the maximum value of 1 to a minimum value given by $E/(E + FS_i)$, which is proportional to the stimulus value S_i. This gain is recharged to its initial unity value when the stimulus ends. Charge and discharge rates are determined by the parameters E and F.

IV.C. AUs competitive layer

Habituated stimuli are the inputs of the F1 layer (Fig.8), that is, a competitive neuron layer. The competitive model used is on-centre off-surround [11], so each neuron is reinforced with its own activity, but is attenuated by the activity of the neurons it is connected to, which is known as lateral inhibition. Lateral inhibition is used for those AU which are complementary, such as AU5 (eyes wide open) and AU43 (eyes closed).

Neuron activity is computed using equations (13) and (14), where x_i is the activity of neuron i and A_i is the decay rate. The second term is the auto-reinforcement (on-centre), which makes neuron activity tend to its saturation value B. The last term in the equation represents lateral inhibition (off-surround). This equation is based on the Hodking-Huxley model [12]:

$$\frac{dx_i}{dt} = -Ax_i + (B - x_i)[S_ig_i + f(x_i)] - x_i\sum_{i\neq j} f(x_j) \qquad (13)$$

where

$$f(x_i) = Dx^2 \quad (14)$$

Different results can be obtained in terms of neuron competition by varying the function $f(x_i)$. We have used a parabolic function so that the winner neuron is reinforced against the rest. The competition schema is a winner-take-all, as it is desirable that only one AU prevails over the complementary ones.

IV.D. Emotional expressions layer

After competition, the outputs of the previous layer are the inputs of the emotional expression layer. Those values follow equation (15):

$$y_j = \sum_{j=1}^{N} x_i z_{ij} \quad (15)$$

z_{ij} is the weight link between neuron i of the AU competitive layer and neuron j of the emotional expression layer. That weight is adjusted following an outstar learning law given by equation (16):

$$\frac{dz_{ij}}{dt} = -\gamma z_{ij} + \beta(y_i - z_{ij})x_j \quad (16)$$

where γ is the forgetting factor, $\beta$is the learning rate and y_j is the activity of neuron j in the emotional expression layer. Finally, the neurons' activity also follows a similar competitive architecture to those in the AU layer, but in this case, all neurons are interconnected. Competition is shown in equations (17) and (18):

$$\frac{dy_j}{dt} = -Ay_j + (B - y_j)[T_j + f(y_j)] - y_j \sum_{i \neq j} f(x_j) \quad (17)$$

$$f(y_j) = Dy_j^2 \quad (18)$$

V. Experimental results

V.A. Emotional recognition on video sequences

The described method was tested with a set of 210 video sequences of 35 Spanish people aged from 24 to 50 years. 24% were woman. Half the video sequences where used for network training. Each participant was asked to perform each of the 6 universal emotional expressions with their face (anger, fear, disgust, joy, surprise and sadness), and to hold the expression for a few seconds. Video capture was carried out using a *PS3 Eyetoy* webcam, whose image resolution was set to 320x240 pixels. As the processing rate was 50 frames per second, sequences where captured at that frame rate. A trapezoidal approximation was used to implement the differential equations of the neural network

$$\frac{dx}{dt} = g(x)$$
$$x(kh) = x\big((k-1)h\big) + \frac{\big[g(kh) + g\big((k-1)h\big)\big]h}{2} \quad (19)$$

where h is the integration step and k is a positive integer, so that $0 < kh < t$. The h value was set to 0.02 seconds.

Fig.9 shows the neuron activity on the top layer of the emotional recognition stage for three different expressions: surprise, sadness and fear. It can be seen that, thanks to the proposed design, it is able to capture the development of facial expressions through time.

The first expression on the top left of Fig.9 is surprise. The user has quite a pronounced nasolabial wrinkle, even in rest position, so the algorithm detects a slight activation of joy expression. However, as the rest of the facial features become more noticeable (eyelids opening, eyebrows rising or mouth opening), neuron competition makes the output change clearly to the correct surprise expression. The second expression on the top is sadness. It can be seen that the presence of glasses does not affect the method performance. The third expression shown is fear. Here, the algorithm starts detecting a surprise expression, due to the fact that, at the beginning of the sequence, facial features which are characteristic for this emotion are active: whole eyebrow rising (AU1+AU2), eyelids opening (AU5) and mouth opening (AU26). As the expression evolves, fear expression features become more intense: inner part of the eyebrows rising (AU1), mouth corners pulled down (AU15), and nostrils opening (AU10). Also, the position of the lips in the second half of the sequence causes a little rise in the output that corresponds to the sadness expression. However, neuron competition makes the fear expression prevail over the rest of the expressions.

Table VI shows the results obtained with the proposed method for the 210 captured sequences. Results are given as a confusion matrix expressed as a percentage of the total number of sequences for each expression. The criteria used to determine that an emotional expression is active is that its corresponding output neuron activity is over a certain threshold (0.4 for the current experiment) and its value is significantly higher compared to the rest of expressions (four times higher for the current experiment). Those decision parameters could be different, depending on the implementation to make the system more or less sensitive to transitions between facial expressions.

V.B. Experiments with the Cohn-Kanade database

The proposed emotion recognition approach has also been tested using the Cohn-Kanade expression database [8]. This database is widely used in both Action Unit recognition and emotional expression recognition experiments. However, image sequences are only labelled in terms of Action Units but not emotional expressions. It is therefore necessary to translate the Action Unit labelling for each sequence into its corresponding emotional expression. For our experiments, we have made use of the emotional labelling developed by Buenaposada [13], who selected a subset of 333 sequences of 92 people from the Cohn-Kanade database and labelled them with their corresponding emotional expression. The following paragraphs describe some examples of the outputs obtained with our method. For these experiments, we have used the same conditions as the ones used in previous experiments with our own video sequences. As images in the Cohn-Kanade database have a resolution of 640x480 pixels, they were rescaled to 320x240 pixels. Sequence images were also sent to the system at a rate of 30 images per second, the same frame rate in which they were captured, so h = 0.03 has been chosen as the integration step. The criteria chosen to determine whether an expression is correctly recognized was the same as in the previous experiment, that is, the winning expression should have at least a four times higher value than the rest and a minimum value of 0.4 in its [0,1] range.

Fig.10 shows an example of an anger expression sequence and the outputs obtained. It can be seen that the obtained output value for the anger expression at the end of the sequence is five times higher than the rest, as AU4 (frowning) and AU7 (eyelids closing) are intensively activated along with AUs 23 and 24 (lip tightening and pressing). On the other hand, AU17 is also active (mouth corners fall and wrinkles appear in the chin region), which is also indicative of fear and sadness, which have the highest output values after the anger expression. Joy also has a small output value, due to a displacement of the cheeks that cause the nasolabial wrinkle to appear slightly (AU12), which is characteristic of the joy expression.

It has to be noted that the aim of the Cohn-Kanade database is to study facial changes due to the activation of Action Units, so sequences only last the amount of time needed for the AUs to be displayed. For this reason, the structure of the sequences is such that most of the discriminative information is stored in the last frames of the sequences near the expression apex. Moreover, emotional expressions last longer. In a real context, it is uncommon for an emotional expression to be displayed during a fraction of a second as most of the Cohn-Kanade sequences do, but take place over a longer period of time. AU timing is related to isolated facial micro-expressions, and their timing is related to the amount of time needed for each associated muscle to contract. Emotional expression timing is context dependent, but it is unlikely, in a real scenario, to have an expression of emotion that only lasts a fraction of a second. For that reason neural network parameters have been chosen so that the expression apex needs to be shown to the system for a small amount of time, just as in natural human communication.

Fig.11 shows the effect of expression timing over the obtained output. The image sequence is displayed, along with two different outputs: one for the original sequence (bottom left) and one for a sequence in which the first frames coincide with the original, but where the last frame has been maintained for more than a second (bottom right). It can be seen how we are able to detect the displayed disgust expression in both conditions, but it is not until frame 20 that the output begins to reach a stationary state. It can also be seen that, if an expression is maintained over time, our method is able to differentiate the expression displayed more clearly, just as the human cognitive system does [14].

TABLE VII shows the results obtained for the 333 sequences in a confusion matrix expressed as a percentage of the total number of sequences for each expression. Comparing these results with the ones obtained for the video sequence experiment (Table VI), it can be seen that a good recognition rate is maintained. However, a decrease in the hit rate can be appreciated for the anger and fear expressions, which are commonly confused with disgust and sadness respectively. This could be associated with the characteristics of the Cohn-Kanade database, which is intended for AU recognition, but not for expression recognition, so there are sequences in which only a few sets of Action Units are active and with a low intensity. This means that the neural network does not have enough information to correctly discriminate the winning expression.

Table VIII shows a comparison of the obtained results against those of other authors. It can be observed that the results obtained with our approach are comparable to those obtained by Buenaposada [13] and Vretos [16], in spite of our lower requirements (v.i.). Our results also outperform those obtained by Wimmer [17], although Xiao's results [15] are better.

Establishing fair comparisons between authors turns out to be a difficult task, due to different factors. First, from the totality of sequences of the database, authors use different subsets for their experiments. Also, even if the sequences used were the same, emotional expression labelling would not be the same, given that the translation of Action Units into emotional expressions is author dependent. Using a common set of sequences with associated labels would be required to make fair comparisons. As stated before, Buenaposada's [13] emotion labelling has been used in our experiments, so more equitable comparisons can be established between the results obtained with our experiment and his.

It has to be pointed out that, although Buenaposada's method can be used in real time, the analysis of the Cohn-Kanade sequences is done over static images; while our experiment treats them as live streaming video, which implies that recognition is done in real time with its inherent computational restrictions. Also, his approach is based on a manifold of facial characteristics previously obtained from the database. This makes his method highly dependent on the subset of AUs associated to each expression during training. As an example, most of the sequences labelled by Buenaposada as fear have AU25 (showing the teeth) as the common primary facial feature, which is present in all of them. Even some of the considered fear expressions only have AU25 active along with AU17 or AU15 (pulling lip corners outwards and down). This means that the performance of his method could decrease, having used a set of fear expressions that included a wider range of facial features, or features that change over time. Moreover, our method expression recognition is not bond to a particular set of facial features in a particular frame, but to a competition of different facial features over time, making it more reliable for real time applications and new users. As in [13], a subspace of facial features obtained from the database is used in [15] for expression recognition on static images, so the same previously stated restrictions can be considered for this method.

In [16], a Principal Component Analysis is carried out on a facial grid, and the obtained principal components are combined with information of the barycentre of grid vertices to recognize facial expressions using Support Vector Machines. Although the approach of using a grid to study facial geometry is similar to ours, their analysis is based on manual initialization of the deformable grid used to track the face, which restricts their approach for real time applications. In [17], a deformable point model and its variations over time are used for expression recognition. The approach is intended for a robotic application, so real time computational restrictions, such as image resolution, are taken into account during the experiment. It should also be noted that, in order to improve computational performance for real time applications, the proposed approach is intended to work at 320x240 image resolution. Given that the Cohn-Kanade sequences have a resolution of 640x480, we have resized all sequences to half-size in our experiment, in order to maintain real time specifications. However, the rest of the authors, except [17], use the original database resolution. It can be seen that, in similar conditions, our method outperforms the one presented in [17] for all emotional expressions.

V.C. Qualitative real time experiments

Fig.12 shows the performance of the proposed approach when used with a 40 second video sequence captured at 50 frames per seconds. In this sequence, all six emotional expressions are performed one at a time, and are maintained in their append position between one and two seconds. The transition between expressions was done in a continuous and natural way, and no rest position was adopted in the whole sequence, except at the beginning. Before the last surprise expression (frames 1300 to 1500 approximately), some quick and non-expressive facial movements were performed, like the ones that occur when talking (i.e. mouth movements or eyebrows rising).

In order to show the effect of modifying network habituation parameters, the F value was increased (equation (12)), so the network quickly habituates to input stimuli. It can be seen that once the maximum output value for each expression is reached, a decrease in the output value occurs even when the expression is at its apex position, due to the habituation to active stimuli (see for example frames 450 to 600, where the anger expression is taking place).

It can be observed that the system is able to correctly detect all performed expressions except fear, which is confused with sadness and surprise. It can also be seen how, thanks to network dynamic characteristics, the transition between expressions takes place smoothly (see for example the transition between disgust and sadness during frames 800 to 875), while some time is needed for the system to swap between one emotional expression and the next. This is of great importance for HCI real time applications, in order to obtain a correct interpretation of human expressiveness, given that emotional expressions do not occur punctually, but are a consequence of a complex cognitive process which takes place over time.

VI. Conclusions

The complexity of facial movement, and the large set of different nuances that the human face can achieve, makes it difficult to automatically recognize the emotional state of a person from their facial expressions. In this paper, a set of minimal face displacements have been considered along with the characteristic changes in facial appearance that those displacements produce. The combination of those minimal features allows the six universal emotional expressions to be recognized.

The recognition method presented is divided into two independent stages. The first one looks for the recognition of Action Units described in the Facial Action Coding System. Automatic Recognition of FACS Action Units in live streaming images is a challenging problem, since they have no quantitative definitions, can vary from one person to another and appear in combinations. Also, the appearance of transient features while performing expressions makes the recognition task even more difficult. Our approach, based on Active Shape Models and Gabor filters, is able to recognize some of the most significant action units that contribute to facial expression. Moreover, recognition is carried out on low resolution images, so the system is suitable for work in real time using a webcam with limited resolution.

The outputs of the Action Unit recognition module feed an emotional recognition system based on a layered hybrid neural network that looks for similarities in a real cognitive process. The network is based on a habituation layer plus a competitive layer. The habituation layer allows the system to adapt to users that have marked facial characteristics even in rest position, due to their particular anatomy (i.e. scarves or age wrinkles). Both the competitive and the habituation layer are based on differential equations, so they take into account the dynamic information of facial expressions over time. Also, the proposed network is highly parameterizable, so that different characteristics of the recognition process, such as the sensitivity to input stimuli, the degree of competition between neurons or the habituation to persistent stimuli, can be modified.

Experimental results have proven that the proposed method can achieve a high hit rate in real time video sequences. When compared to other approaches using the Cohn-Kanade database, our method has provided

similar results to methods based on higher resolution and static images, and better results than methods based on low resolution images and video sequences.

VII. Acknowledgements

This work has been partly supported by the Spanish Ministry of Economy and Competitiveness (No. DPI2008-06738-C02-01) and the Junta de Castilla y León (Agencia de Inversiones y Servicios, y Prog. de Apoyo a Proys. de Investigación VA013A12-2 and VA036U14).

References

[1] Edlund, J. &Beskow, J. 2007, 'Pushy versus meek - using avatars to influence turn-taking behaviour', *INTERSPEECH-2007*, pp. 682-685.

[2] Marcos-Pablos S., Gómez-García-Bermejo, J., Zalama, E. 2008. 'A realistic facial animation suitable for human-robot interfacing'.*Proceedings of the 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS2008*, pp. 3810-3815, ISBN 978-1-4244-2058-2.

[3] Marcos-Pablos S., Gómez-García-Bermejo, J., Zalama, E. 2010. 'A realistic, virtual head for human-computer interaction'.*Interacting with Computers*, Vol.22 (3), pp. 176-192, May 2010, ISSN 0953-5438.

[4] Cootes,T.F., Taylor,C.J. and Cooper D.H. , Graham, J., 1995. Active shape models - their training and application. Computer Vision and Image Understanding (61): pp. 38-59.

[5] Daugman, J.G., 1985. Uncertainty relations for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. Journal of the Optical Society of America A, vol. 2, pp. 1160-1169.

[6] Wei, Y., 2009. Research on Facial Expression Recognition and Synthesis.Master Thesis, Department of Computer Science and Technology, Nanjing University.

[7] Daugman, J.G., 1985. Uncertainty relations for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. Journal of the Optical Society of America A, vol. 2, pp. 1160-1169.

[8] Kanade, T., Cohn, J.F., Tian. Y., 2000.Comprehensive database for facial expression analysis. Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00). Grenoble, France, pp. 46-53.

[9] Ekman, P., Friesen, W.V., Hager, J.C., 2002. The Facial Action Coding System - Second edition.Weidenfeld& Nicolson, London, UK.

[10] Grossberg, S, 1968. Some Nonlinear Networks capable of Learning a Spatial Pattern of Arbitrary Complexity. Proceedings of the National Academy of Sciences, USA, vol. 59, pp. 368 – 372.

[11] Grossberg, S., 1973. Contour enhancement, short-term memory and constancies in reverberating neural networks. Studies in Applied Mathematics, vol. 52, pp. 217-257.

[12] Hodgkin, A. L., Huxley, A. F., 1952. A quantitative description of ion currents and its applications to conduction and excitation in nerve membranes.J. Physiol. (Lond.), vol. 117, pp. 500-544.

[13] Buenaposada, J. M., Muñoz, E., Baumela, L., 2008. Recognising facial expressions in video sequences. Pattern Analysis and Applications, vol. 11, pp. 101 – 116.

[14] McAndrew, F. T., 1986. A Cross-Cultural Study of Recognition Thresholds for Facial Expressions of Emotion. Journal of CrossCultural Psychology, vol. 17(2), pp. 211-224.

[15] Xiao, R., Zhao, Q., Zhang, D., Shi, P., 2011. Facial expression recognition on multiple manifolds.Pattern Recognition, vol.11, pp. 107-116.

[16] Vretos, N., Nikolaidis, N., Pitas, I., 2009. A model – based facial expression recognition algorithm using Principal Components Analysis. IEEE International Conference on Image Processing, pp. 3301 – 3304.

[17] Wimmer, M., MacDonald, B. A., Jayamuni, D., Yaday, A, 2008. Facial Expression Recognition for Human – Robot Interaction – A prototype. Lecture Notes in Computer Science, vol. 4931, pp. 139-152.

[18] Kakumanu, P., Makrogiannis , S., Bourbakis, N., 2007. A survey of skin-color modeling and detection methods. Pattern Recognition, vol. 40, pp. 1106-1122.

[19] Viola, P., Jones, M. J., 2004. Robust Real-Time Face Detection. International Journal on Computer Vision, vol. 57, pp. 137-154.

[20] Yeasin, M., Bullot, B., Sharma, R., 2004. From facial expression to level of interests: a spatio-temporal approach. En IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[21] Kong, J., Zhan, Y., Chen, Y., 2009. Expression Recognition Based on VLBP and Optical Flow Mixed Features. Fifth International Conference on Image and Graphics, pp.933-937.

[22] Tripathi, R., Aravind, R., 2007. Recognizing facial expression using particle filter based feature points tracker. Proceedings of the 2nd international conference on Pattern recognition and machine intelligence, pp. 584-591.

[23] Wang, F., Huang, C., Liu, X., 2009. A Fusion of Face Symmetry of Two-Dimensional Principal Component Analysis and Face Recognition. International Conference on Computational Intelligence and Security, vol. 1, pp.368-371.

[24] Zhao, T., Liang, Z., Zhang, D., Zou, Q., 2008. Interest filter vs. interest operator: Face recognition using Fisher linear discriminant based on interest filter representation. Pattern Recognition Letters, vol. 29 (1), pp. 1849-1857.

[25] Li, J.B., Pan, J. S., Lu, Z. M., 2009. Face recognition using Gabor-based complete Kernel Fisher Discriminant analysis with fractional power polynomial models. Neural Computing & Applications, Vol. 8 (6), pp. 613-621.

[26] Ou, J., Bai, X. B., Pei, Y., Ma, L., Liu, W., 2010. Automatic Facial Expression Recognition Using Gabor Filter and Expression Analysis. Second International Conference on Computer Modeling and Simulation, vol. 2, pp.215-218.

[27] Lajevardi, S. M., Hussain, Z. M., 2009. Facial expression recognition using log-Gabor filters and local binary pattern operators.International Conference on Communication, Computer and Power (ICCCP'08), pp. 349-353, Oman, 2009.

[28] Shan, C., Gong, S., McOwan., P. W., 2009. Facial expression recognition based on Local Binary Patterns: A comprehensive study. Image and Vision Computing, vol. 27 (6), pp. 803-816.

[29] Lucey, P., Cohn, J., Lucey, S., Sridharan, S., Prkachin, K.M., 2009. Automatically detecting action units from faces of pain: Comparing shape and appearance features, 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp.12-18.

[30] Cootes, T. F., Cooper, D. H., Taylor, C. J., Graham, J., 1991. A Trainable Method of Parametric Shape Description. 2nd British Machine Vision Conference pp. 54-61.

[31] Milborrow, S., Nicolls, F., 2008. Locating Facial Features with an Extended Active Shape Model. Proceedings of the 10th European Conference on Computer Vision: Part IV, pp. 504 – 513.

[32] Surendran, N., Xie, S., 2009. Automated facial expression recognition – an integrated approach with optical flow analysis and Support Vector Machines.International Journal of Intelligent Systems Technologies and Applications, vol 7 (3), pp. 316 – 346.

[33] Pantic, M., Patras, I., 2006. Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences. IEEE Transactions on Systems, Man, and Cybernetics vol. 36 (2) pp. 433–449.

[34] Tian, Y., Kanade, T., Cohn, J. Recognizing action units for facial expression analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence vol. 23 (2) , pp. 97–115.

[35] Schmidt, M., Schels, M., Schwenker, F., 2010. A Hidden Markov Model Based Approach for Facial Expression Recognition in Image Sequences. Artificial Neural Networks in Pattern Recognition pp. 149-160.

[36] Valenti, R., Sebe, N., Gevers, T., 2007. Facial Expression Recognition: A Fully Integrated Approach. 14th International Conference of Image Analysis and Processing - Workshops (ICIAPW 2007), pp.125-130.

[37] Schiano, D.J., Ehrlich, S.M., Sheridan, K., 2004. Categorical Imperative NOT: Facial Affect is Perceived Continuously. CHI '04 Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 49–56.

[38] Samuel, M., Jaime, G., Eduardo, Z. "A realistic, virtual head forhuman-computer interaction". Interacting with Computers, Vol.22 (3),pp. 176-192, May 2010, ISSN 0953-5438.

Figures and Tables

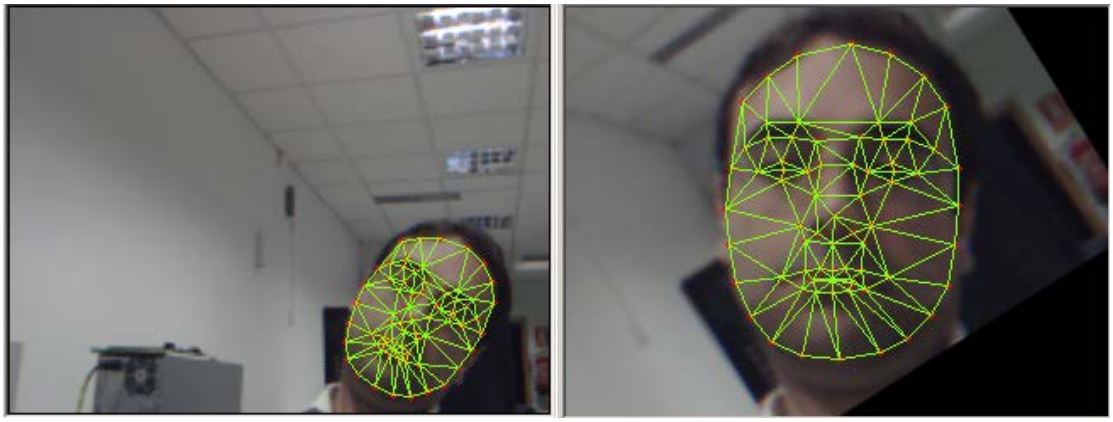

Fig.1.Triangulated 75 points shape used for feature extraction and segmentation, and result of image normalization.

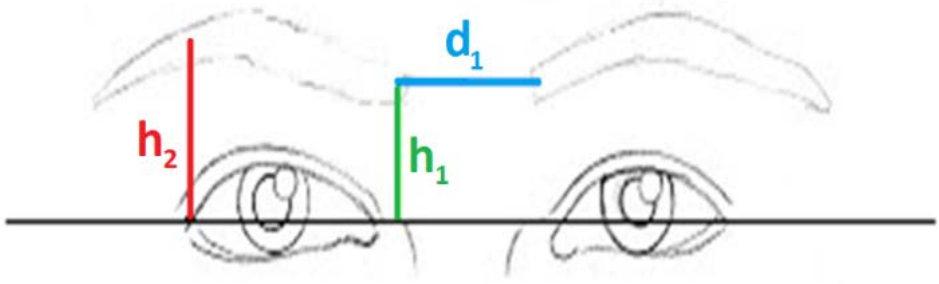

Fig.2. Distances computed from the shape model for eyebrow displacement. h2: outer eyebrow corner distance, h1: inner eyebrow corner distance. d1: distance between eyebrows.

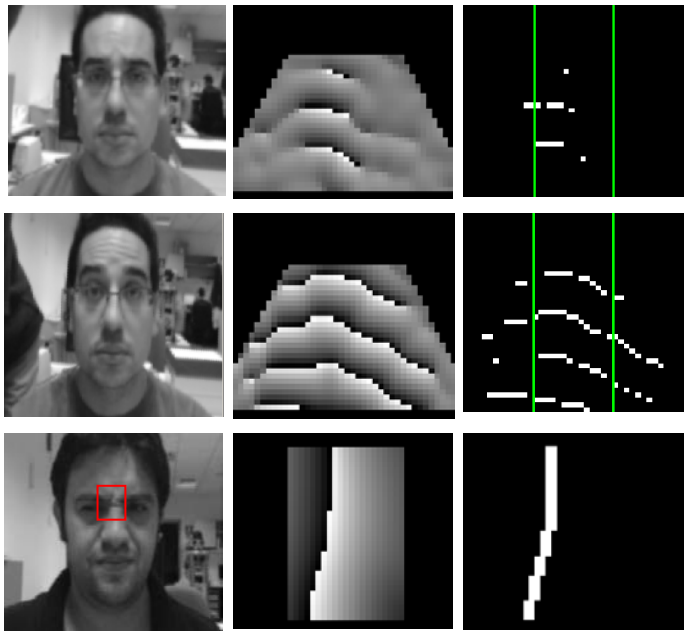

Fig.3.Detection of forehead wrinkles. Rows from top to bottom: AU1, AU1+AU2 and AU4. Columns from left to right: grey-scaled image, filtered and masked image and thresholded image.

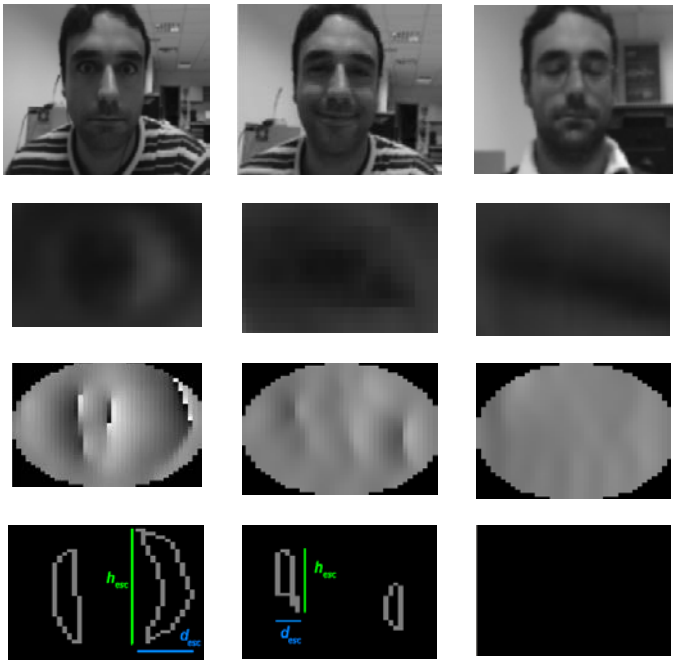

Fig.4. From top to bottom: original image, extracted region, filtered+masked region and thresholded region result of the process of visible sclera extraction in the eye region. Left column images correspond to AU5, middle column to AU6+12 and right column to AU43. The area of biggest contour (hsc * dsc) is computed to determine the amount of visible sclera.

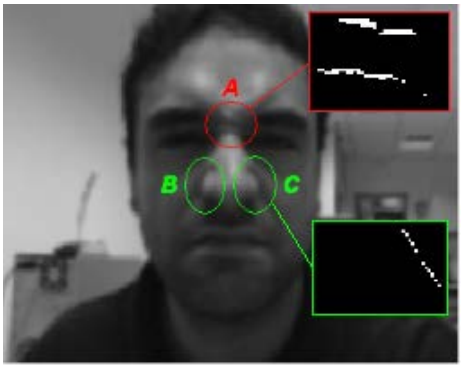

Fig.5. Wrinkle detection for AU9 and AU10 in the nose region. Gabor filter parameters: F = 0.12 pixels-1, θ = 1.5 rad, σx = 4.2 pixels and σy = 1 pixels and size 32x32 pixels for region A. Gabor filter parameters: F = 0.21 pixels-1, θ = 1.9 rad, σx = 1.9 pixels, σy = 0.7 pixels and size 32x32 pixels for region C.

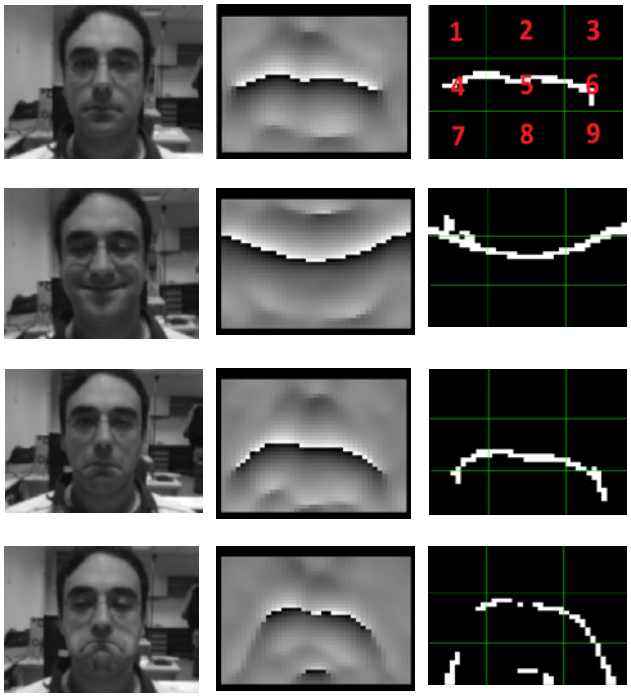

Fig.6. Filtering result in the mouth region when performing different AUs. Top right image displays the labelling of the nine subregions considered. From top to bottom: rest position, AU12 left and right active, AU15 active, AU17 active.

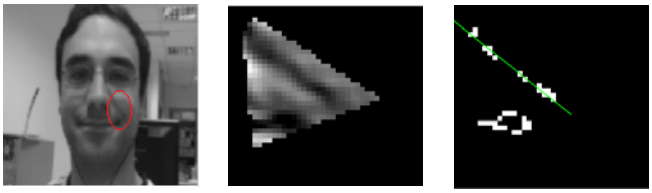

Fig.7. Original, filtered and thresholded image in the proccess of extraction of the nasolabial wrinkle. Angle and intensity are computed from the thresholded image by pixel pattern searching.

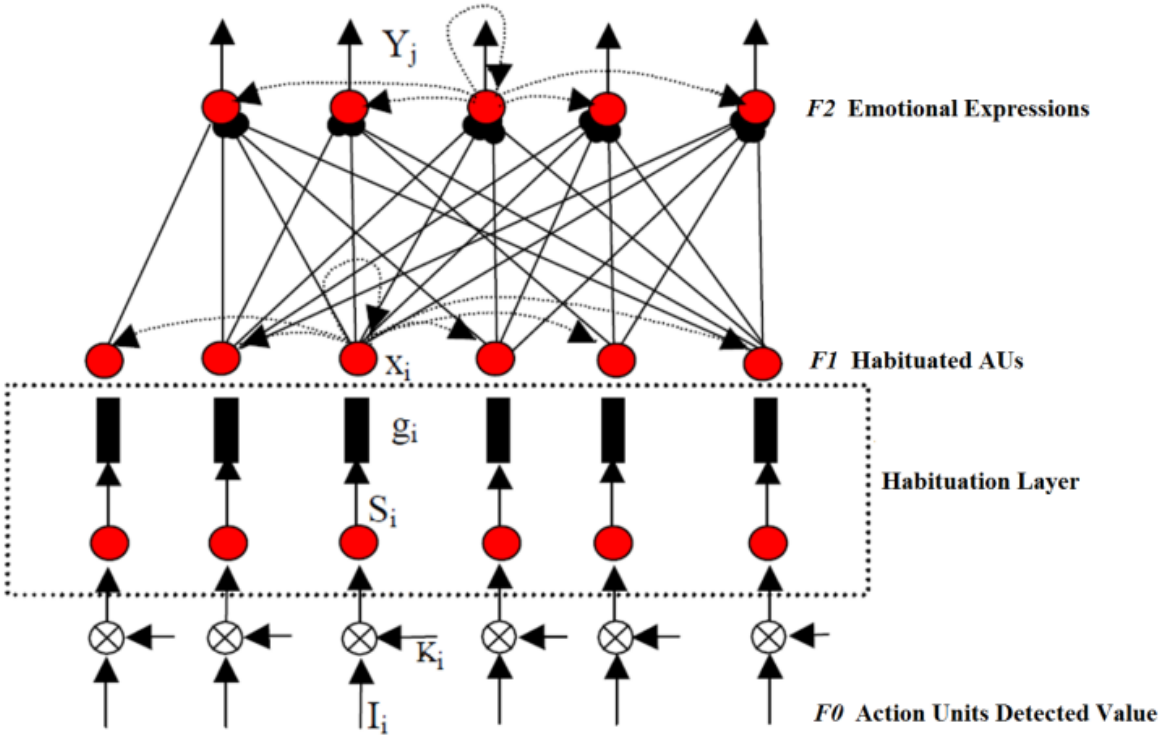

Fig.8. Competitive + Habituation based network schema used for facial emotion recognition.

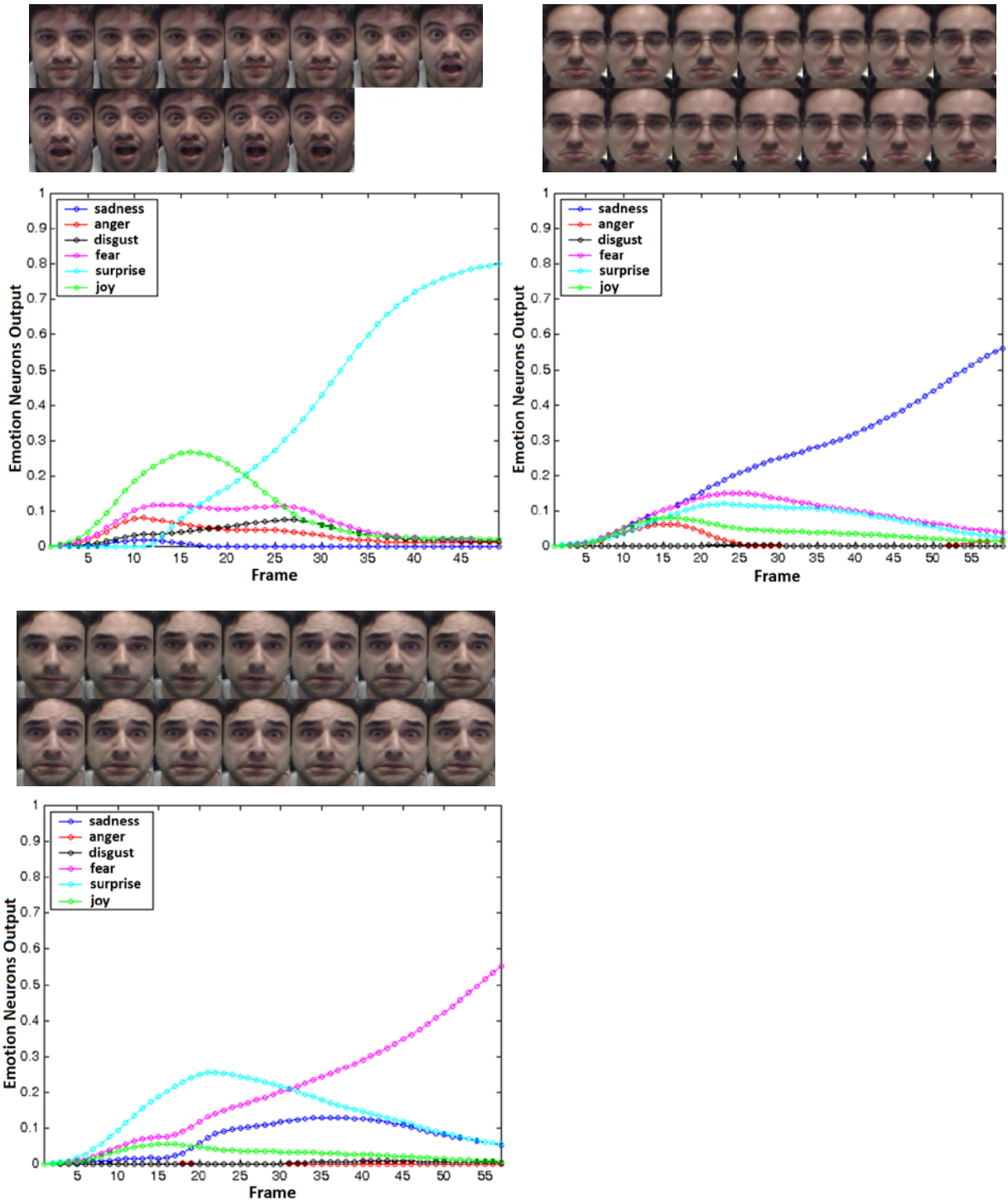

Fig.9. Image sequences corresponding to surprise, sadness and fear along with their corresponding system outputs. Images correspond to one of each 4 frames of the original sequence. Neural Network parameters: h = 0.02, A = 0.3, B = 1, D = 5, E = 0.04, F = 0.08.

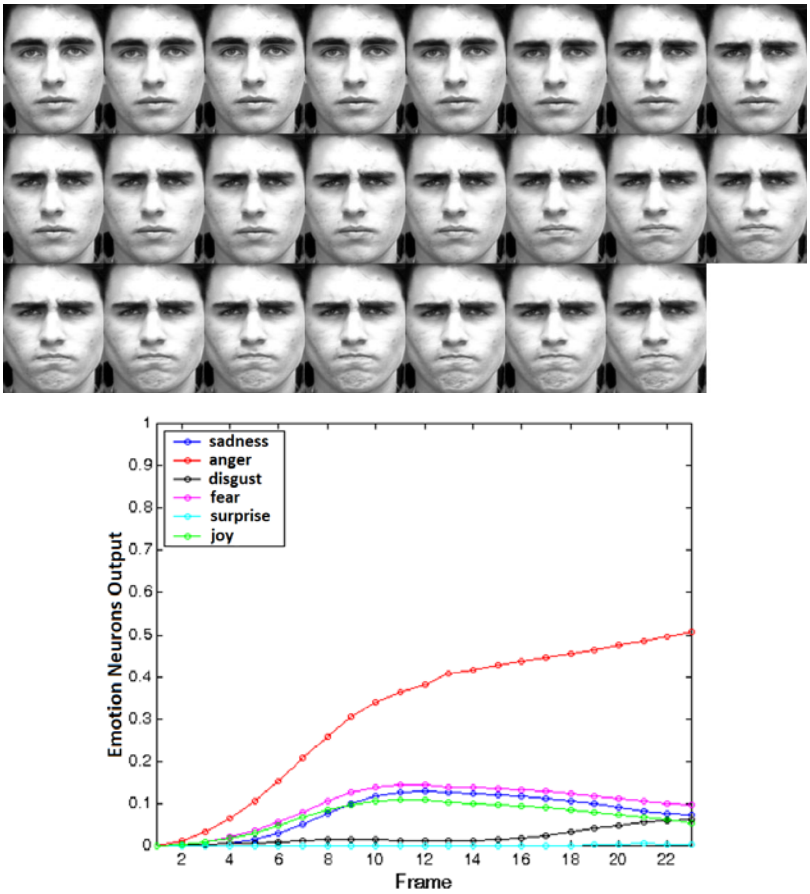

Fig.10. Image sequences corresponding to person 113, sequence 8 of the Cohn-Kanade database [8] with its corresponding system outputs. Neural Network parameters: h = 0.03, A = 0.3, B = 1, D = 5, E = 0.04, F = 0.08.

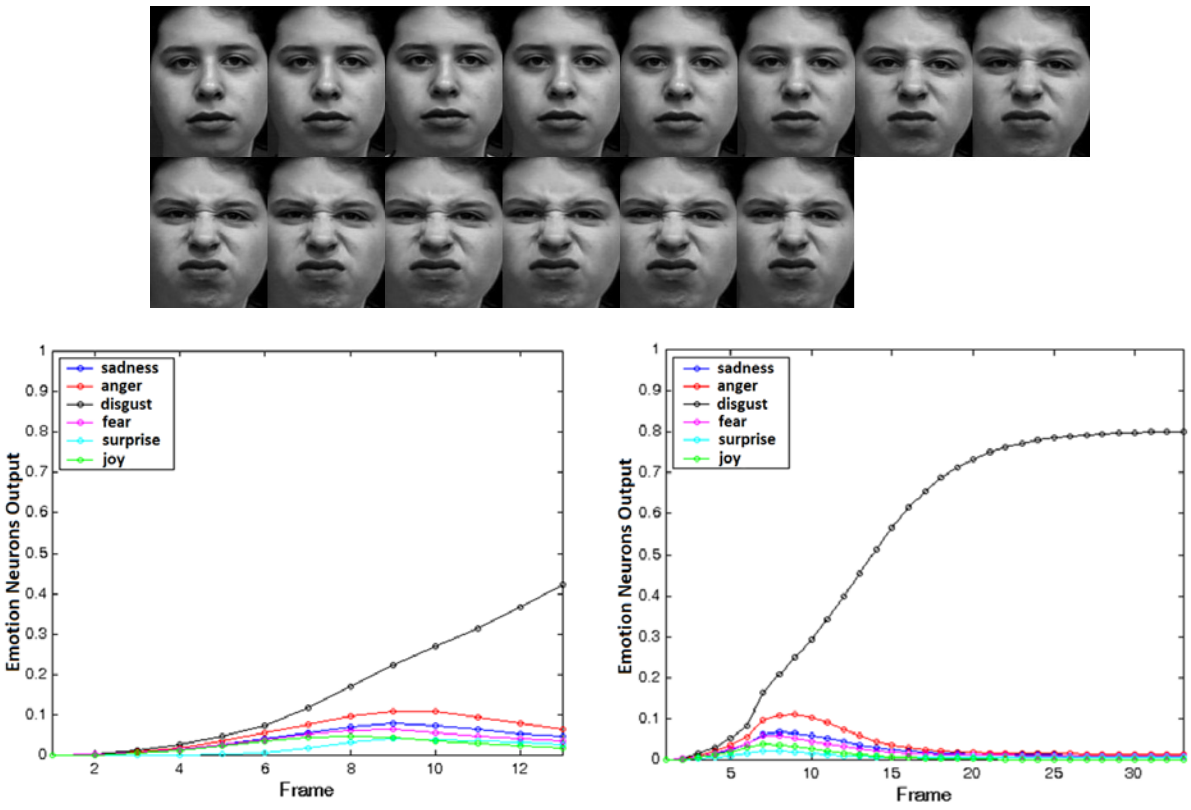

Fig.11: Image sequences corresponding to person 111, sequence 7 of the Cohn-Kanade database [8] with its corresponding system outputs for the original sequence (bottom left) and a longer sequence built from repeating the last frame (bottom right). Neural Network parameters: h = 0.03, A = 0.3, B = 1, D = 5, E = 0.04, F = 0.08.

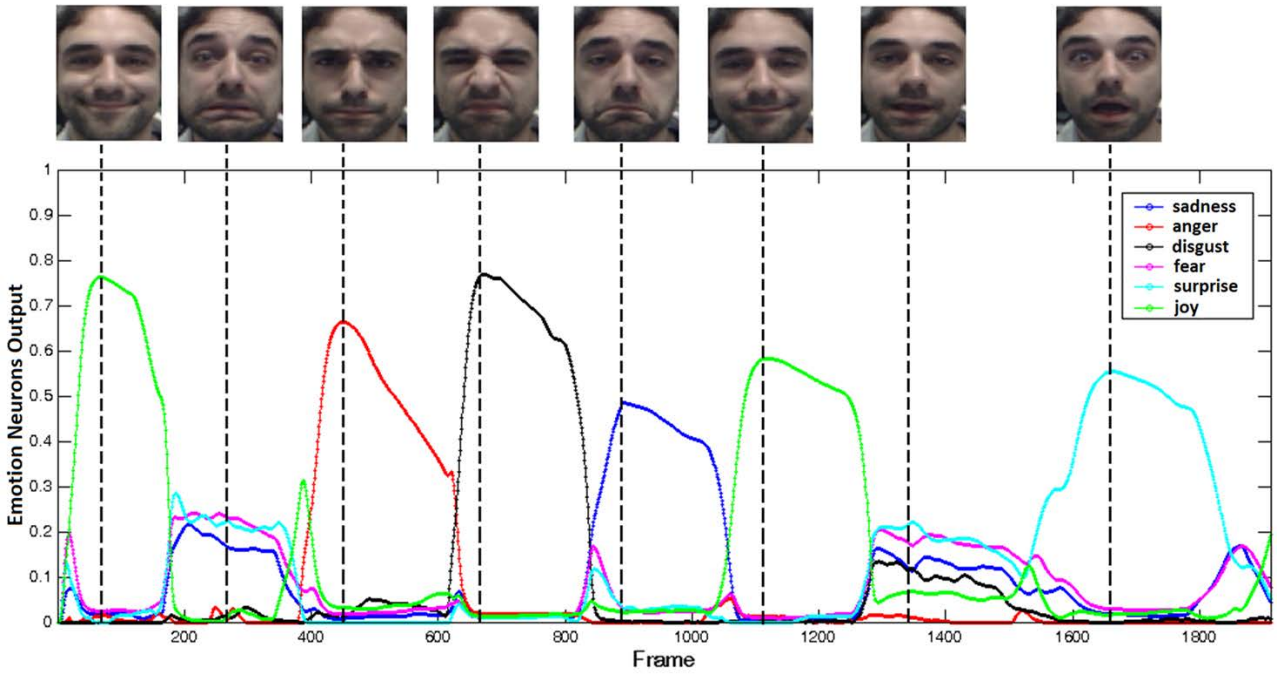

Fig.12. System outputs for a 40 second sequence where the six universal emotion expressions are performed. From left to right performed expressions were: joy, fear, anger, disgust, sadness, joy, talking, surprise. Neural Network parameters: h = 0.02, A = 0.3, B = 1, D = 5, E = 0.04, F = 0.30.

Table I
AUs considered in forehead region and searched characteristics

Action Unit	Searched characteristics
AU1	Wrinkle intensity in the central part of the forehead. Distance between the lower part of the eyebrows and the line that crosses the outer edges of the eyes.
AU2	Wrinkle intensity in the lateral sides of the forehead. Distance between the outer part of the eyebrows and the line that crosses the outer edges of the eyes.
AU4	Wrinkle intensity between the eyebrows. Distance between the inner and outer parts of the eyebrows and the line that crosses the outer edges of the eyes. Distance between the eyebrows.

Table II
AUs considered in eyes region and searched characteristics

Action Unit	Searched characteristics
AU5	Eyelids distance. Amount of visible sclera.
AU6	Eyelids distance. Amount of visible sclera. Wrinkle intensity in the lateral sides of the eyes.
AU7	Eyelids distance. Amount of visible sclera. Absence of wrinkles in the lateral sides of the eyes.
AU43/ AU45	Eyelids distance. Amount of visible sclera. Amount of time eyes remain closed.
AU46	Eyelids distance. Amount of visible sclera. Only one eye is closed.

Table III.
AUs considered in eyes region and searched characteristics

Action Unit	Searched characteristics
AU9	Wrinkle intensity in the lower part of the nose. Wrinkle intensity in the nose region located between the eyes.
AU10/ AU11	Wrinkle intensity in the lower part of the nose.

Table IV
AUs considered in eyes region and searched characteristics

Action Unit	Searched characteristics
AU10+25	Distance between the lips. Upper teeth are visible.
AU16+25	Distance between the lips. Lower teeth are visible.
AU15	Corners of the lips fall down.
AU17	Corners of the lips fall down.
AU18	Corners of the lips come together.
AU26	Mouth is open.

Table V
AUs considered in eyes region and searched characteristic.

Action Unit	Searched characteristics
AU12	Nasolabial wrinkle intensity. Nasolabial wrinkle angle Lip corners are risen
AU15	No wrinkles in the chin.
AU17	Wrinkles appear in the chin.

Table VI
Confusion matrix expressed as percentage of the total number of sequences for each expression. 210 sequences were used in the experiment, 35 per each emotional expression.

	Joy	Sadness	Anger	Disgust	Surprise	Fear
Joy	100	0	0	0	0	0
Sadness	0	88.57	0	0	0	5.72
Anger	0	8.57	91.43	2.86	0	0
Disgust	0	0	8.57	97.14	0	0
Surprise	0	0	0	0	100	11.43
Fear	0	2.86	0	0	0	82.85

Table VII
Confusion matrix expressed as percentage of the total number of sequences for each expression. 330 sequences from the Cohn-Kanade database were USed in the experiment.

	Joy	**Sadness**	**Anger**	**Disgust**	**Surprise**	**Fear**
Joy	91.4	0	0	0	0	0
Sadness	0	83.0	5.3	0	0	18.6
Anger	0	3.8	81.6	4.8	0	4.7
Disgust	7.5	0	13.1	95.2	0	0
Surprise	1.1	0	0	0	95.8	6.9
Fear	0	13.2	0	0	4.2	69.8

Table VIII
Comparison of the performance of our system against other authors.

	Joy	**Sadness**	**Anger**	**Disgust**	**Surprise**	**Fear**
Our method	91.4	83.0	81.6	95.2	95.8	69.8
Buenaposada [13]	98.8	82.0	73.9	87.9	100	73.9
Wimmer[17]	70.27	73.33	66.67	58.82	93.33	59.09
Bretos[16]	86.8	81.0	84.8	93.3	97.1	89.1
Xiao [15]	97.5	83.2	85.8	97.2	98.4	78.6