# Automatic emotion recognition through facial expression analysis in merged images based on an Artificial Neural Network

Javier G. Rázuri\*, David Sundgren\*, Rahim Rahmani\*, Antonio Moran Cardenas†

\*Dept. of Computer and Systems Sciences, Stockholm University, Stockholm, Sweden

\*{javier, dsn, rahim} @dsv.su.se

†Pontifical Catholic University of Perú PUCP, Lima, Perú

†amoran@pucp.edu.pe

Abstract—This paper focuses on a system of recognizing human's emotion from a detected human's face. The analyzed information is conveyed by the regions of the eye and the mouth into a merged new image in various facial expressions pertaining to six universal basic facial emotions. The output information obtained could be fed as an input to a machine capable to interact with social skills, in the context of building socially intelligent systems. The methodology uses a classification technique of information into a new fused image which is composed of two blocks integrated by the area of the eves and mouth, very sensitive areas to changes human's expression and that are particularly relevant for the decoding of emotional expressions. Finally we use the merged image as an input to a feed-forward neural network trained by back-propagation. Such analysis of merged images makes it possible, obtain relevant information through the combination of proper data in the same image and reduce the training set time while preserved classification rate. It is shown by experimental results that the proposed algorithm can detect emotion with good accuracy.

keywords—Artificial Neural Network, Merged Images, Facial Expression Recognition, Emotions, Detection of Emotional Information.

### I. INTRODUCTION

We are approaching a probable future where machines could exceed human performance, but it is more accurate to think of humans with enhanced capabilities with the help from machines that understand their emotions [1]. Nowadays, Artificial Intelligence with its major contributions seems to show us that there is no doubt that sooner or later the machines will be become more and more clever, e.g., intelligent machines are balanced to fill a growing number of roles in today's society, and its influence is entering directly virtually every domain of our lives, including surgical assistants [2], fighting on battlefields like autonomous fighter machines [3], and assisting in classrooms in educational contexts [4], nursing homes [5], and offices [6].

In humans, emotions play an extremely important role in their lives. They are the most reliable indicators of capacity in human socialization. They determine how we think, how we behave and how we communicate with others. With this information, it is possible to think, that the new future generation of machines must have some skills to understand human emotions and generate synthetic emotions similar to humans. The new machines could have some kind of functionalities to respond more flexibly, foretell and adjusting to what humans want. Depending on the complex process of emotion internalized in the body, the meaning of the postures, facial expression, eye gaze, gestures, and tone of voice, etc., of the individuals differs a lot.

Therefore, we can underline into the group of humans emotional and attentional expressions "the emotion recognition through facial expression", as well as automatic emotion recognition, necessary to generate a communication bridge between humans and machines. In order to facilitate the human-machine interrelation, the machines must possess some type of system that could allow measuring human emotions and provide its corresponding interpretation. Like in humans, we could be think that the machines could have some rationality and it cannot be understood in a separate manner from emotion, "Particularly for mobile robots, we need to have something similar to emotions in order to know -at every moment and at least- what to do next" [7].

The facial features and expressions are one of the groups of windows through which humans express emotions, and are critical to daily communication. The human face assists a number of cognitive tasks, e.g., in a noisy environment the lips from the mouth can contribute greatly to speech comprehension or a simple action of "knit one's brows" makes clear our disagreement about something. Research in social psychology has shown that conveying messages in meaningful conversations can be dominated by facial expressions, and not by spoken words [8]. The complex world of face emotional expression contains important



<sup>\*</sup>This work is supported by the project FORMAS: Multimodal Communication for Participatory Planning and Decision Analysis: Tools and Process Models.

information, and it can then be said to play a communicative role; it can be controlled to some extent and be intentionally used to acquire knowledge from the others and know their emotional state. A greater understanding on the part of the machines about the human emotional expressions could generate in them expectations about our own behavior and adapt theirs accordingly.

In the present article, we addressed the problem of emotion recognition through facial expression and, in the same time, we introduce a new method to detect six basic facial emotional expressions by using an ANN (Artificial neural network) in merged images. The output of this system is a recognized emotion, that it will impact in a future model that supports the decision making process of a survey machine, see [9], that issue is very related with the emotional behavior in user's faces when they confront a set of questions. The advantage of using merged images is the selection of relevant information only in implied areas affected by facial changes. The original image of the face is represented by another reduced-size image that keeps the basic features, where facial expression shows more changes, these advantages in merging images are show in a reduced training time, performance in classification rate and the fast convergence of the back-propagation error.

This paper is structured as follows: Section II presents a literature review related with emotions in machines, emotion recognition in facial expressions and Facial Emotion Recognition through Neural-network-based methods, while section III describes the implementation of facial emotion detection model. Experimental results are presented in Section IV, while Section V presents the discussion and future work.

### II. LITERATURE REVIEW

### A. Emotions in Machines

Not long ago, the idea of integrating computational systems with emotions was formulated [10]. Rosalind Picard [11], developed the term "Affective Computing" where the computing relates to, arises from, or deliberately influences emotion or other affective phenomena and it has a main objective: to recognize and generate synthetic emotions affecting artificial agents. In relation to this topic, there are important works which try to look for the emotional interrelation between humans and machines, specifically simulating and interpreting emotions [12], [13].

Affective computing combines the research to produce machines which are able to recognize, model and communicate emotions to enhance human computer interaction and aid related research in surprising ways [14]. As an example, Decision Affect Theory [15] provided empirical evidence of the effect of expectations on emotion

generation. Extensive models such as TAME [16] create demand for research connecting different theories and their effects on emotions and decision making, as we shall do. For several examples, see FLAME [17], EMA [18], Decision Field Theory [19] or AiSoy 1 [9].

One of two main areas of the affective computing is: *To detect emotional information*, which is the problem addressed in this study. Affective computing tries to capture the information of the signs, which are related to the expression of emotions, and their interpretation. The data of users and environment can be captured by a set of sensors, after obtaining the information, this is classified in order to select just the relevant data for the analysis. Some of these types of detection are grouped in: voice recognition, natural language processing, face recognition (object of our study), etc.

### B. Emotion Recognition in Facial Expressions

The concept of automatic face recognition closer to our days was developed in [20], with an early attempt to automatically analyze facial expressions by tracking the motion spots on an image sequence. Recent work on construction of intelligent machines have attempted to replicate identically human beings [21]. But there is something lacking in the interaction between human beingmachine; "The emotional factor" remains very diffuse. Machines are still this cold component that precludes an understanding the humans' emotional state. Physically the ability to display emotions on a human like face is both an important and necessary step in making machines more accessible to the general public, we see clearly that the face is the window where the emotions are showed. Exists a correlation between all the emotions that are experienced and expressed in a spontaneous manner [22], e.g., the expressions of the face are a clear manifestation of the intensity of an involuntarily emotion, without previous planning or intention.

Then the key will be to explore new ways of humanmachine interaction by enabling the computer to be more aware of the human user's emotional and attentional expressions, in this road, several approaches to recognize emotions from face have been reported, consistent with this view, this group of studies also analyzes facial emotional expression, diverse physical zones in tune with mechanical movements of the face muscles. A comprehensive review related with the analysis of facial expressions can be found in [23]. We could underline some within this group of approaches; the parametric models in order to extract the shape and movements of the mouth, eye and eyebrows was used in works of [24], the major directions of specific facial muscles were the input in an emotion recognition system treated in [25], the permanent and transient facial features such as lip, nasolabial furrow and wrinkles are

good recurrent indicators over the emotions [26], but in this technique, it is important to use geometrical models that with great accuracy located the shapes and appearances of these features, the presence of the features and their geometrical relationship with each other appears to be more important than the details of the features [27].

Nonetheless, the common denominator in the facial emotion detection always be initiated with a detection of face zone, extraction and tracking of relevant facial information, and finally the facial expression classification, then with all of this information shall proceed to analyze the facial expressions to estimate emotion-related activities.

## C. Facial Emotion Recognition through Neural-networkbased methods

The extraction of emotion from the static image allows the recognition of several physical features such as: eyes, wrinkles on the forehead, size of eyebrows, color of the skin, etc., and the corresponding sizing and location. In this case, the neural network is accurate for the acquisition of nonlinear mapping between different sets of data, this analysis allows decode the relationship between the physical features of a face and its impression. The potential of the Neural-network-based methods is the performance of facial expression classification into a single basic emotion category. The sort of selection in six basic emotions using Neural-network-based methods was proposed in the work of [28], where, the units of the input to the ANN correspond to the brightness distribution data extracted from an input static image. The average recognition rate was 85 percent in a group of 90 tested images.

Furthermore, the ANN can be reinforced with the use of hybrid approach, in the work of [29], the ANN are often combined with Hidden Markov Models (HMMs) and is employed in facial emotion classification. This analysis used an ANN to estimate the posterior for the discriminant HMM, and it achieved positive results on the recognition of emotion in the upper and lower parts of the static image separately. The focus research of [30] used the analysis of principal facial regions using principal component analysis and neural networks, the classifiers were built in a group of fifteen neural networks. Only one ANN into this group was used for region detection and the other fourteen were used to learned to recognize seven universal emotions over eyes and mouth regions. The conducted experiments showed a 46 percent to 80 percent rate of successful recognition that was reduced to the average precision of 70 percent.

Positive results in emotional classification of an input static facial image can be found too in [31], this study showed outputs from six different classes of neutral

emotions. In the construction of the ANN, the output layer had contained seven units, each of which corresponds to one category of emotion; the average of the correct recognition rate achieved was 86 percent. In works of [32], the neural network performs a nonlinear reduction of the dimensionality in the input image, because the data of interest lies on an embedded non-linear manifold of the higher-dimensional space. In this step the algorithm makes a statistical decision about the category of the observed expression. The set of outputs gives an estimation of the probability of the examined expression belonging to the associated category. The power of this classification raised the 90.1 percent.

Others types of approaches like in the work of [33] propose the use of Multilayer Feed-forward Neural Networks and a Radial Basis Function Networks, commonly used in nonlinear mapping approximation and pattern recognition. The experiment shows the classification of seven basic types of emotions: Neutral, Happiness, Sadness, Anger, Fear, Surprise and Disgust. The Euclidean distances from the contour point in the static image and the geometric coordinates from facial characteristics points represent the set of data input into the neural network. This approach was tested with the set of images from the JAFFE database and it reached 73 percent of accuracy. Projections of feature regions on a fixed filter set of images was proposed in [34], this model used a feedforward neural network, in which the inputs are the group of features based representation of the face, considering the observations found in the study of human expressions. The network model consists of ensembles of 11 feed-forward, fully connected by vanilla neural networks. This architecture has 105 inputs per network and each network include hidden layers with 10 nodes. The training of each network works with online back propagation. The outputs of each network are combined to produce a percentage value for the classification of each emotion. Experiments conducted in [35], showed the applicability of neuro-fuzzy networks to extract emotion in facial motion. This approach have attempted at classifying primary and intermediate emotions using the definition of FAP intensities and the definition of emotions in terms of FAPs. The facial animation parameters (FAPs) specified in MPEG-4 compose a very rich set of parameters that allowing a wide range of facial motion. The challenge of this classification is performed by translating the Feature Point (FP) movements to the Facial Animation Parameters (FAPs).

# III. IMPLEMENTATION OF FACIAL EMOTION DETECTION MODEL

An overview of the proposed work contains two major modules: Facial Image Treatment and an ANN propagation algorithm to recognize facial expressions. Basically, this facial emotion detection loop starts with the input of a new image, it passes through a series of phases, turning it into a new image (merged image) prepared for analysis in the ANN. The ANN has trained previously with a series of images that form the training set related with six different face emotions (six for each individual), with the same treatment that the input image. Finally, once the group to which the image belongs is found, the system reports about the emotional state of the face. The developed facial emotion detection loop can be seen in Fig. 1.

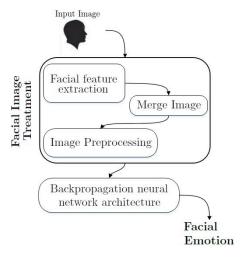


Fig. 1. Facial emotion detection Loop

A system such as an automatic interviewer constantly analyzes the human face and will involve the task of categorizing active and spontaneous facial expressions using only the inputs of the eyes and mouth zone so as to extract information about the underlying emotional states. The extracted features are either geometric features such as the shape of the facial components (eyes, mouth) and the locations of facial ducial points (corners of the eyes, mouth), or appearance features representing the texture of the facial skin including wrinkles, bulges, and furrows, in our approach we work with the complete zone of the eyes and the mouth, for the sole purpose of merging the two extractions into a single new image. In order to resize the merged image we use the Nearest Neighbor Interpolation method, because it is very simple and requires less computation, as it use nearest neighbor's pixel to fill interpolated point. For the new matrix, a new value is calculated from a neighborhood of samples and replaces these values in the minimized image; this technique is applied to all the images in the training set of the Artificial Neural Network. For the object of our study is important apply some type of binarization over the images. We apply a simple formula related to the threshold from the pixel,

$$V_{n_{pix}} = \begin{cases} 0, & \text{if } V_{c_{pix}} < 230\\ \frac{V_{c_{pix}} - 255}{255}, & \text{otherwise,} \end{cases}$$
 (1)

The value 230 was determined through trial and error depending on the luminous intensity over the set of images. This value could change accord to others luminous intensities. Our research takes in account, find in the future an efficient adaptive threshold rate. The threshold rate must have a value autofit, affected by the level of brightness in the images. An example of the total process of the facial image treatment can be seen in Fig. 2.

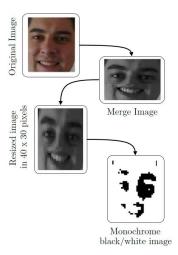


Fig. 2. Facial Image Treatment

Once we have the group of images merged, the next step is to provide the ANN the current input data, what in this case are the images, here we fixed the first stage for the emotion recognition through facial expression. We use a backpropagation algorithm to recognize facial expressions with feed-forward architecture, here the algorithm uses complex relationships between inputs and outputs or to find patterns in data. The back-propagation of feed-forward architecture is designed based on facial features that merged on new image as illustrated in Fig. 3.

The algorithm consists of (1) an input layer containing 1200 neurons that represent the input variables in form of pixels from the image of 40 x 30 size, that is extracted data from the part of the forehead and the mouth zone, that form the new merged image. The input variables  $x_i$  (each pixel) is multiplied by a weight  $w_i$  and are added after. We used a neuron bias b in the initialization process. Now we have the neuron's activation z. So,

$$z = \sum_{i=1}^{p} w_i x_i + b = \begin{bmatrix} W^T & b \end{bmatrix} \begin{bmatrix} X \\ 1 \end{bmatrix}$$
 (2)

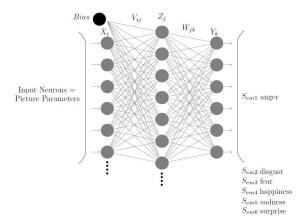


Fig. 3. Backpropagation Neural Network architecture

The neuron bias b works the same as adding an additional input with weight b and value 1, so we have,

$$z = \sum_{i=1}^{p} w_i x_i = W^T X \tag{3}$$

Having the neuron's activation z, it is fed into the activation function  $\sigma(z)$ . This function returns a value on the interval [0,1]. The activation function that we used is the sigmoidal function, Thus, we shall assume that

$$\sigma(z) = \begin{cases} 0, & \text{for } z < 0 \\ 1, & \text{for } z \ge 0 \end{cases} \quad \text{and} \quad \sigma(z) = \frac{1}{1 + e^{-sz}} \quad (4)$$

The parameter z determines the steep grade of the sigmoid function. We choose s=1 and the output of the activation is then the output of the neuron. (2) one hidden layer containing one or more neurons to help capture the nonlinearity in the data, the number of neurons in the hidden layer depends on the number of training samples. The hidden layer is capable of approximating any continuous function. That is, as long as there is a sufficient number of hidden neurons in it. In order to fix the number of hidden neurons in the one hidden layer, we can use various criteria developed by researchers during the last couple of decades. We will use the approach developed in [36],  $N_h = \frac{(N_{in}) + \sqrt{N_p}}{L}$  where L is the number of hidden layer,  $N_{in}$  is the number of input neuron and  $N_p$  is the number of input sample. The optimum number of hidden layers and hidden units depends on the complexity of network architecture, the number of input and output units, the number of training samples, the degree of the noise in the sample data set, and the training algorithm; and (3) an output layer containing six nodes representing the output variable to the problem, the number of neurons that has to be in the output layer is fixed as we know the number of emotions that we are going to consider, and in our case that is a set of facial

expressions: anger, disgust, surprise, happiness, sadness and fear, into a set of six combinations per individual. Fig. 4 shows a part of the images that belong to the training data set.

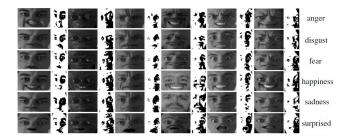


Fig. 4. Facial expressions per individual

Our structure have one hidden layer and we start with input  $X_i$ . The neurons between layers are fully interconnected with weight  $V_{ij}$  and  $W_{jk}$ . The activation of one hidden neuron j can then be found using  $z_j = W_i^{h^T} X$ , So, we can define

$$X = \begin{bmatrix} X_1^T \\ X_2^T \\ \vdots \\ \vdots \\ X_N^T \end{bmatrix}, \quad W^h = \begin{bmatrix} W_1^{h^T} \\ W_2^{h^T} \\ \vdots \\ \vdots \\ W_p^{h^T} \end{bmatrix}$$
 (5)
and 
$$Z = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1p} \\ z_{21} & z_{22} & \cdots & z_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ z_{N1} & z_{N2} & \cdots & z_{Np} \end{bmatrix}$$

The elements of the matriz Z define the activation of the hidden neuron j to input sample  $x_i$ . We can find the hidden layer activation Z, the hidden layer output V and the system output Y, in such case,

$$Z = XW^h$$
,  $V = \sigma(z)$  and  $Y = \sigma(VW^0)$  (6)

Where  $Y = \begin{bmatrix} y_1 & y_2 \dots y_N \end{bmatrix}^T$ , and the output layer weights  $W^0$  are defined equal that the hidden layer weights  $W^h$ .

Now we need to train the ANN by back-propagation and this involves three stages: the feed-forward of the input training pattern, the calculation and back-propagation of associated error, and the adjustment of the weights. The data are fed forward from the input layer, through hidden layer, to output layer without feedback. The ANN is initialized randomly, the input x is taken and the next will be find the resulting output y, the desired output d serves to calculated the back-propagation of associated error e = d - y, the goal now is to minimize the cost function,

$$J = \frac{1}{2} \sum_{l} e_l^2 \tag{7}$$

We assume that the output layer has no activation function. So we have  $y_l = \sum_j w_{jl}^0 v_j$ . Now, we need adjust the weights of the output layer using the update rule,

$$w_{jl}^{0}(n+1) = w_{jl}^{0}(n) - \alpha(n) \frac{\partial J}{\partial w_{jl}^{0}}$$
(8)

In this equation the learning rate is  $\alpha(n)$ , and affects the speed at which the ANN arrives at the minimum . To find the Jacobian  $\frac{\partial J}{\partial w^0_{il}}$  we need use the chain rule, that is,

$$\frac{\partial J}{\partial w_{il}^0} = \frac{\partial J}{\partial e_l} \frac{\partial e_l}{\partial y_l} \frac{\partial y_l}{\partial w_{il}^0} \tag{9}$$

These three partial derivatives are all relatively easy to find. We have,

$$\frac{\partial J}{\partial e_l} = e_l, \ \frac{\partial e_l}{\partial y_l} = -1, \ \frac{\partial y_l}{\partial w_{il}^0} = v_j \tag{10}$$

The output layer weights are updating through,

$$w_{il}^{0}(n+1) = w_{il}^{0}(n) + \alpha(n)v_{j}e_{l}$$
(11)

In the hidden layer we apply a similar principle, the Jacobian is obtained through,

$$\frac{\partial J}{\partial w_{ij}^h} = \frac{\partial J}{\partial v_j} \frac{\partial v_j}{\partial z_j} \frac{\partial z_j}{\partial w_{ij}^h} \tag{12}$$

After some computation, we obtain,

$$\frac{\partial J}{\partial v_i} = \sum_{l} \left( \frac{\partial J}{\partial e_l} \frac{\partial e_l}{\partial y_l} \frac{\partial y_l}{\partial v_i} \right) = \sum_{l} -e_l w_{jl}^0 \tag{13}$$

with,

$$\frac{\partial v_j}{\partial z_j} = \sigma'_j(z_j)$$
 and  $\frac{\partial z_j}{\partial w_{ij}^h} = x_i$  (14)

The data is actualized through law for hidden neuron weights, so that,

$$w_{ij}^{h}(n+1) = w_{ij}^{h}(n) + \alpha(n)x_{i}\sigma_{j}'(z_{j})\sum_{l}e_{l}w_{jl}^{0}$$
 (15)

Then, based on the feed-forward error back-propagation learning algorithm, back-propagation will search the error surface using gradient descent. Based on the error, the portion of error correction is computed, and then the weights for all layers are adjusted simultaneously, it is necessary use of N test samples  $x_i, d_i$  in order to adjust the weights. The training data set are each of six output expressions per individual in a merged image. We have decided to have 3 layers in order to train the ANN and the "Bias Neurons" option and

"Sigmoid" for transfer function is important because our data set is normalized. Once the weights are calculated the ANN is ready to select the emotion in a merged image output, the algorithm will provide six outputs related with each set of emotions. Fig. 5 shows the complete figure of the Facial emotion detection process.

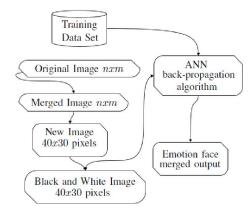


Fig. 5. Flow chart of Facial emotion detection algorithm

### IV. TESTS AND RESULTS

We have evaluated our algorithm in two different data sets, using our own data base of images and the images from the Cohn-Kanade (CK) database. We will show the results of four experiments that are divided in: the facial emotion detection using merged images in our own data base, the isolated analysis in the mouth and eye zones using our own data base, and finally, the facial emotion detection using merged images from the Cohn-Kanade (CK) database.

A. Facial emotion detection system using the complete merged in our own data base

We developed a controlled study to evaluate whether our methodology recognizes the emotions in static facial expressions, for this purpose, we use 72 images from 12 individuals randomly chosen, with different ethnic groups, sex (8 male and 4 female) and different lighting conditions at the time of image capture. A set of questions has been executed at the time to the image capture in order to trigger the six facial emotions and make more natural the experiment. We have another group used for the training set containing 36 images of six individuals, each of which consists of six images of emotions. The other group is used for test set that have 72 new images of 12 individuals, all the images that serve like a input to the ANN were subjected to a previous preprocessing step that cover resizing and merging, the final image composes the union of the areas from the eyes and mouth, the selection of these two regions as the basis for emotion recognition is predictable because the most visual indication of emotions is visible in those areas [37]. Table I shows the confusion matrix of the emotion recognition system based facial emotion detection which gives details of the strengths and weaknesses of this system. The overall performance of this classifier was 83.3 percent. The diagonal components reveal that all the emotions can be recognized with more than 75.00 percent of accuracy.

TABLE I
CONFUSION MATRIX OF THE FACIAL EMOTION DETECTION SYSTEM
USING THE COMPLETE MERGED IMAGE

	Anger	Disgust	Fear	Нарру	Sad	Surprise	
Anger	75.00	0.00	0.00	16.67	0.00	0.00	
Disgust	0.00	83.33	8.33	0.00	16.67	0.00	
Fear	0.00	0.00	75.00	0.00	0.00	0.00	
Happy	0.00	0.00	0.00	83.33	0.00	0.00	
Sad	25.00	16.67	0.00	0.00	83.33	0.00	
Surprise	0.00	0.00	16.67	0.00	0.00	100.00	

The data is represented as a percentage of successful emotion recognition per emotion type. A graph of the same data can be seen in Fig. 6

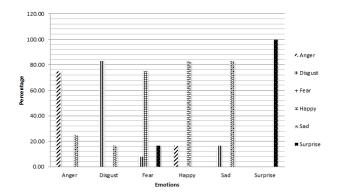


Fig. 6. Percentage of successful emotion recognition in merged image

Fig. 7 shows the groups of desired outputs contrasted with the real inputs, this graph provides an idea of the accuracy of our classifier, we can observe the six output emotions codified in six binary data. In the same graph we can observe the error calculated from backpropagation algorithm for 2000 iterations, it is clearly observable that this error decreases to zero and verifies that the ANN produces an output that exactly matches the expected output.

As can be seen, out of 12 images of anger, 9 of them are successfully classified, while the remaining 3 were recognized as sadness, it could be because the facial expressions like sadness, anger, and fear occurring in response to frustrating and threatening situations, e.g., in faces related with anger and sadness the common denominator are the

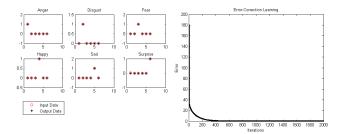


Fig. 7. I/O and BP error for the system using a merged image

lips, sometimes, they are pressed tightly in both cases [38]. From 12 images of disgust, 10 of them were successfully recognized, while the remaining 2 were classified as sadness, in this case we could think that the confusion is produced by the similarity of the mouth, in these two face emotions, the mouth shape forms a bow that arches. Studies that use Positron Emission Tomography (PET) to examine brain function reveal that the activation zone for the disgusted and sadness faces is located in the left lingual gyrus from brain, hence the similarity of their expressions [39]. Out of 12 images of fear, 9 of them were successfully classified, while the rest was mostly attributed to surprise and a little less to disgust. In the case of disgust, we highlight some commonalities [40], e.g., in both cases the eyebrows come closer to each other causing confusion. Is very common to feel surprise and fear sensations at the same time. When we face to fearful events, the human face shows mixed emotions, because some areas of the face are connected with surprise and others with fear. For this reason, many times the face keeps the dominant impression that could be fear, e.g., when we feel surprise and fear the eyes have mostly the same opening. The surprise can be presented sometimes only in the mouth and with a mixture with fear in the eyes and brow [41]. From 12 images with expressions of happiness, 10 were successfully classified, while the remaining 2 were classified as anger, but the similarity scale between happiness and anger becomes clear, when we note the visibility of teeth exposed by the smile and angry face, this characteristic makes stronger the visual percept, the great percentage of the test images related with faces from happiness and anger have showed some areas of teeth. Out of 12 images of sadness, 10 were successfully classified and the remaining 2 were classified as disgust, here, we have the same feature perceived, that when we carried out the analysis between disgust and sadness. Finally we can observe that surprise is recognized with very high accuracy. Some group of misclassifications could attributable to the diverse individuals, each individual displays emotions differently and many emotions are entangled or overlap, the system made some mistakes easily related to human emotion recognition, which is fully understandable if one takes into account the universe from the nature of human

emotions and various ways of expressing them.

# B. Facial emotion detection using the area around eyes and mouth separately

We conducted the new experiment with the same own data test set that have 72 new images of 12 individuals and the training set containing 36 images. Here, the detection algorithm is focused in the eyes zone like a permanent facial feature. The new image includes the brows, eyes and the forehead, as illustrated in Fig.8. The new image will feed the feed-forward algorithm of the neural network trained by back-propagation.

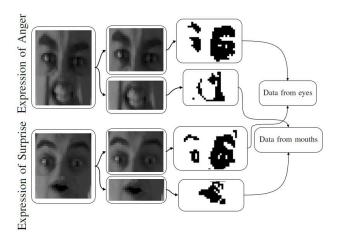


Fig. 8. Image data from the eyes and mouth separately

The average successful emotion recognition was 80.6 percent, which means that 58 out of 72 test images were successfully classified. As shown in Table II, the emotions using the eyes zone can be recognized with more than 66.67 percent of accuracy. Emotions like anger, sadness and surprise are affected in the percentage of accuracy in the recognition, as can be seen in Fig.9. In this case, the experiment shows that the recognition rates can be improved if the network has trained with more information.

TABLE II

CONFUSION MATRIX OF THE FACIAL EMOTION DETECTION SYSTEM
USING THE EYES ZONE

	Anger	Disgust	Fear	Нарру	Sad	Surprise	
Anger	66.67	0.00	0.00	16.67	0.00	0.00	
Disgust	0.00	91.67	8.33	0.00	16.67	0.00	
Fear	0.00	0.00	91.67	0.00	8.33	0.00	
Happy	0.00	0.00	0.00	83.33	0.00	25.00	
Sad	33.33	0.00	0.00	0.00	75.00	0.00	
Surprise	0.00	8.33	0.00	0.00	0.00	75.00	

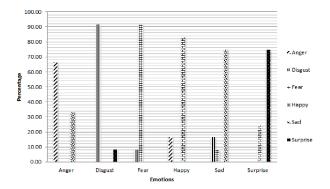


Fig. 9. Percentage of successful emotion recognition in eyes zone

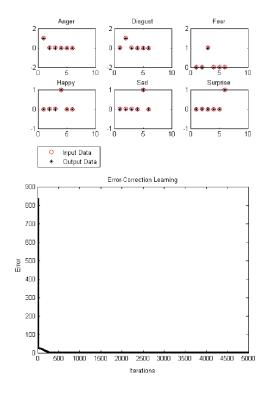


Fig. 10. I/O and BP error for the system using the eyes zone

The specific regions of the face like the mouth and eyes zone are particularly relevant for the decoding of emotional expressions, because of this, an isolated classification could not provide more details for the improvement in the identification. The simulations in MATLAB have increased a range of 5000 iterations as can be seen in Fig.10. We increased the iterations because achieving a good learning performance means a good recognition over the outputs. Compared with the recognition in merged images, the learning process has required more time in order to found the desired outputs that contrast with the real inputs.

In the case of mouth region, the average successful emotion recognition was 68 percent, which means that 49 out of 72 test images were successfully classified. As shown in Table III the emotions using the eyes zone can be recognized with more than 58.33 percent of accuracy. Emotions like anger, disgust, fear, happiness and surprise are affected in the accuracy percentage of emotion recognition, as can be seen in Fig.11. Fig.12 shows the increment of iterations number that the system used, in order to reach a good learning performance. In the set of test images, the whole training set to the system is widely known as an epoch. The system usually use many epochs before it converges in the training process. During training process, the network has settled in a function that could be the best predictor from the input data. In areas of mouth and eyes areas, a higher value for iterations was necessary, in order to achieve a greater magnitude of change in the weights. In contrast with the number of iterations used in merged images, in the areas of mouth and eyes, we used successive iterations. Increase the number of iterations was favorable to reduce error only slightly and reach a good classification.

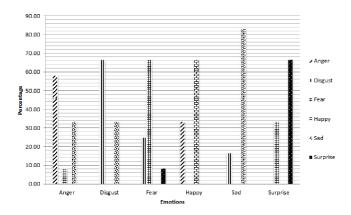


Fig. 11. Percentage of successful emotion recognition in mouth zone

TABLE III
CONFUSION MATRIX OF THE FACIAL EMOTION DETECTION SYSTEM
USING THE MOUTH ZONE

	Anger	Disgust	Fear	Нарру	Sad	Surprise	
Anger	58.33	0.00	0.00	33.33	0.00	0.00	
Disgust	0.00	66.67	25.00	0.00	16.67	0.00	
Fear	8.33	0.00	66.67	0.00	0.00	33.33	
Happy	0.00	0.00	0.00	66.67	0.00	0.00	
Sad	33.33	33.33	0.00	0.00	83.33	0.00	
Surprise	0.00	0.00	8.33	0.00	0.00	66.67	

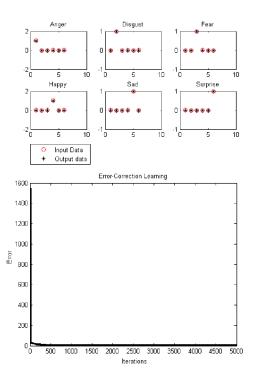


Fig. 12. I/O and BP error for the system using the mouth zone

C. Facial emotion detection using the complete merged image in the Cohn-Kanade (CK) database

We used the Cohn-Kanade (CK) database [42] to construct the training set and test set as can be seen in Fig.13. This database contains 97 individuals showing different expressions. We selected only 40 individuals from the totally group of 97 in order to capture the set of more pronounced facial emotions like anger, disgust, fear, happiness, sadness, and surprised. We collected 6 images per individual coming to have a totally of 240 images. We trained the ANN with the images from 15 individuals (90 images). The rest is comprised of images of 25 individuals (150) and has been used as test set. The emotion recognition in 126 facial images was correctly determined. The system was unable to recognize facial emotions from another 24 facial images, even so, the performance of the emotion recognition algorithm amounts 84 percent. As shown in Table IV, the emotions using merged images from the Cohn-Kanade (CK) database can be recognized with more than 76 percent of accuracy. As can be observed in Fig.14, the percentage of successful emotion recognition is high in emotions like surprise, disgust, happiness and sadness, this group of results are very similar to the first experiment. We increase the number of iteration to 4000 as seen in Fig.15, nonetheless, in front of the isolated analysis in different parts of the face, the back-propagation error had a fast convergence without many iterations.

TABLE IV
CONFUSION MATRIX USING THE COHN-KANADE (CK) DATABASE

	Anger	Disgust	Fear	Нарру	Sad	Surprise	
Anger	80.00	0.00	0.00	16.00	0.00	0.00	
Disgust	0.00	88.00	24.00	0.00	8.00	0.00	
Fear	16.00	0.00	76.00	0.00	0.00	16.00	
Нарру	0.00	0.00	0.00	84.00	0.00	0.00	
Sad	4.00	12.00	0.00	0.00	92.00	0.00	
Surprise	0.00	0.00	0.00	0.00	0.00	84.00	

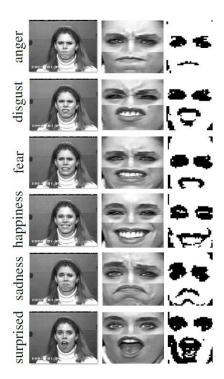


Fig. 13. Some samples from the Cohn-Kanade (CK) database

### V. DISCUSSION AND FUTURE WORK

The results presented above show the great potential of artificial neural networks to recognize human emotions based on face images. In order to design a neural network applicable in practical implementations, the face image should be reduced. To reduce the size of the images a pixeling technique was proposed where the original large-size face image is represented by another reduced-size image, keeping the basic features of the face so that the person could even be recognized by this image. This technique is different from other methods based on face descriptors, where the face is represented by a set of coefficients from which the face can be reconstructed. Also, considering that human emotions are mostly represented through eyes and mouth, only these portions of the face have been considered in the emotion

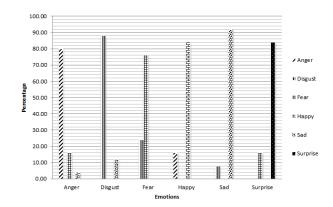


Fig. 14. Percentage of successful emotion recognition using the Cohn-Kanade (CK) database

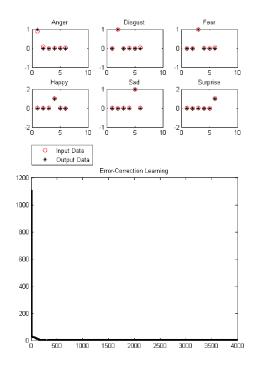


Fig. 15. I/O and BP error using the Cohn-Kanade (CK) database

recognition process. Other parts of the face containing unnecessary information have been eliminated. This method not only allows for further reduction of the size of the image input to the neural networks, but it only considers information useful for emotion recognition discarding unnecessary information which could slow the learning process and negatively affect the recognition performance of the neural network. The neural networks used in this study are static networks where there is not a temporal relationship between inputs and outputs. As future work, it is suggested to use dynamic neural networks for emotion and pattern recognition.

In dynamic networks, the state of the network evolves in time from an initial state to a final state corresponding to the face with the desired emotion. With the information provided by the dynamic neural network we could attempt manage the relevant information into a model that supports the decision-making process, of an autonomous interviewer machine which interacts with several users. However, we shall focus on collecting and analyzing peoples explicit emotions shown on the face, using this information as an emotional variable in a decision making model.

#### ACKNOWLEDGMENT

The authors greatly appreciate the financial support provided by the project FORMAS: Multimodal Communication for Participatory Planning and Decision Analysis: Tools and Process Models.

#### REFERENCES

- R.W. Picard, E. Vyzas, J. Healey, "Toward Machine Emotional Intelligence: Analysis of Affective Physiological State," in IEEE Transactions Pattern Analysis and Machine Intelligence, vol. 23, no. 10, pp. 1175-1191, 2001
- [2] J.L. Fernández, J.M. Gómez, V.F. Muñoz, I. García, D. Melgar, C. Vara and A. García, "Human-Machine Interface Evaluation in a Computer Assisted Surgical System," in IEEE International Conference on Robotics and Automation, pp. 231-236, New Orlean, 2004
- [3] P. Lichocki, P. Kahn Jr, A. Billard, "The ethical landscape of robotics," in IEEE Robotics and Automation Magazine, vol. 18, no. 1, pp. 39-50, 2011
- [4] I. Schulte, "Robots @ School, Project by Latitute Research, Lego Learning Institute and Project Synthesis," Some findings can be accessed at: http://latd.com/2012/01/16/robots-at-school-findings/, 2012
- [5] K.H. Park, H.E. Lee, Y. Kim, Z.Z. Bien, "A Steward Robot for Human-Friendly Human-Machine Interaction in a Smart House Environment," in IEEE Transactions on Automation Science and Engineering, vol. 5, no. 1, pp. 21-25, 2008
- [6] C. Lisetti, S. Brown, K. Alvarez, A. Marpaung, "A Social Informatics Approach to Human-Robot Interaction with a Service Social Robot," in IEEE Systems, Men, and Cybernetics. Special Edition on Human-Robot Interaction, vol. 34 no. 2, 2004
- [7] S. Pinker, How the Mind Works, New York, NY: HarperCollins Publishers, 1997
- [8] A. Mehrabian, Silent messages, Belmont, CA: Wadsworth, 1971
- [9] J.G. Rázuri, P.G. Esteban, D.R. Insua, "An adversarial risk analysis model for an autonomous imperfect decision agent," in T.V. Guy, M. Kárný and D.H. Wolpert, Eds. Decision Making and Imperfection. SCI, vol. 474, pp. 165-190. Springer, Heidelberg, 2013
- [10] E. Mueller, M. Dyer, "Day dreaming in humans and computers," in Proceedings of the Ninth International Joint Conference on Artificial Intelligence, CA: Los Angeles, 1985
- [11] R. Picard, Affective Computing, United States: the MIT Press, 1998
- [12] C. Breazeal, "Emotion and sociable humanoid robots," in International Journal of Human-Computer Studies, vol. 59, no. 1-2, pp. 119-155, 2003
- [13] R.W. Picard, "Toward computers that recognize and respond to user emotion," in IBM Systems Journal, vol. 39, no. 3-4, pp. 705-719, 2000
- [14] R. Picard, "Affective computing: challenges," in International Journal of Human-Computer Studies, vol. 59, no. 1, pp. 55-64, 2003
- [15] B.A. Mellers, A. Schwatz, K. Ho, and I. Ritov, "Decision affect theory: Emotional reactions to the outcomes of risky options," in Psychological Science, vol. 8, no. 6, pp. 423-42, 1997
- [16] L. Moshkina, "An integrative framework for affective agent behavior," in Proceedings of the International Conference on Intelligent Systems and Control, IASTED, 2006

- [17] M.S. El-Nasr, J. Yen and T.R. Ioerger, "FLAME Fuzzy logic adaptive model of emotions," in Autonomous Agents and Multi-Agent Systems, vol. 3, no. 3, pp. 219-257, 2000
- [18] J. Gratch and S. Marsella, "A Domain-independent Framework for Modeling Emotion," in Cognitive Systems Research, vol. 5, no. 4, pp. 269-306, 2004
- [19] J.R. Busemeyer, E. Dimperio and R.K. Jessup, "Integrating emotional processes into decision making models," in W. Gray, Eds. Integrated models of cognitive systems. Oxford University Press, 2007
- [20] M. Suwa, N. Sugie and K. Fujimora, "A preliminary note on pattern recognition of human emotional expression," in International Joint Conference on Pattern Recognition, pp. 408-410, 1978
- [21] K.F. MacDorman and H. Ishiguro, "Toward Social Mechanisms of Android Science: A CogSci2005 Workshop," in Interaction Studies vol. 7, pp. 361-368, 2006
- [22] D. Krstić, Psiholoski reénik, Beograd: Savremena administracija, 1991
- [23] M. Pantic and L.J. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," in IEEE Trans. Pattern Anal. Mach. Intell. vol. 22, no. 12, pp. 1424-1445, 2000
- [24] M.J. Black and Y. Yacoob, "Tracking and recognizing rigid and nonrigid facial motions using local parametric model of image motion," in Proceedings of the International Conference on Computer Vision, pp. 374-381. IEEE Computer Society, Cambridge, MA, 1995
- [25] K. Mase, "Recognition of facial expression from optical flow," in IEICE Transc., E. vol. 74, no. 10, pp. 3474-3483, 1991
- [26] P. Ekman and W. Friesen, "Unmasking the face," Prentice-Hall, 1975
- [27] V. Bruce, Recognizing Faces, Hove, East Sussex: Lawrence Erlbaum Assoc, 1986
- [28] H. Kobayashi and F. Hara, "Facial Interaction between Animated 3D Face Robot and Human Beings," in Proceedings of the International Conference of Systems, Man and Cybernetics, pp. 3732-3737, NY: IEEE New York, 1997
- [29] T. Hu, L.C. De Silva and K. Sengupta, "A hybrid approach of NN and HMM for facial emotion classification," in Pattern Recognition Letters, vol. 23, no. 11, pp. 1303-1310, 2002
- [30] D. Filko and G. Martinović, "Emotion Recognition System by a Neural Network Based Facial Expression Analysis," in Automatika - Journal for Control, Measurement, Electronics, Computing and Communications. ATKAFF. vol. 54, no. 2, pp. 263-272, 2013
- [31] C. Padgett and G.W. Cottrell, "Representing Face Images for Emotion Classification," in Proceedings of Advances in Neural Information Processing Systems, pp. 894-900, 1996
- [32] Z. Zhang, M. Lyons, M. Schuster and S. Akamatsu, "Comparison between Geometry-Based and Gabor Wavelets-Based Facial Expression Recognition Using Multi-Layer Perceptron," in Proceedings of Third IEEE Conf. Face and Gesture Recognition, pp. 454-459. Nara, Japan, 1998
- [33] M. Gargesha, P. Kuchi and K. Torkkola, "Facial Expression Recognition," in EEE 511: Artificial Neural Computation Systems, 2002
- [34] C. Padgett, G.W. Cottrell and R. Adolphs, "Categorical perception in facial emotion classification," in Proceedings of the 18th Annual Conference of the Cognitive Science Society. Erlbaum, pp. 249-253, 1996
- [35] A. Raouzaiou, S. Ioannou, K. Karpouzis, N. Tsapatsoulis, S. Kollias and R. Cowie, "An intelligent scheme for facial expression recognition," in Kaynak, O, Eds. Artificial Neural Networks and Neural Information Processing, Lecture notes in Computer Science 2714. pp. 1109-1116, Springer, 2003
- [36] K. Jinchuan and L. Xinzhe, "Empirical analysis of optimal hidden neurons in neural network modeling for stock prediction," in Proceedings of the Pacific-Asia Workshop on Computational Intelligence and Industrial Application, vol. 2, pp. 828-832, 2008
- [37] P. Ekman, "Strong evidence for universals in facial expressions," in Psychological Bulletin, vol. 115, no. 2, pp. 268-287, 1994
- [38] I. Kotsia, I. Buciu and I. Pitas, "An analysis of facial expression recognition under partial facial image occlusion," in Image and Vision Computing, vol. 26, no. 7, pp. 1052-1067, 2008
- [39] S.F. Taylor, I. Liberzon, L.M. Fig, L.R. Decker, S. Minoshima and R.A. Koeppe, "The effect of emotional content on visual recognition

- memory: a PET activation study," in Neuroimage, vol. 8, pp. 188-197, 1998
- [40] J.M. Susskind, D.H. Lee, A. Cusi, R. Feiman, W. Grabski and A.K. Anderson, "Expressing fear enhances sensory acquisition," in Nature Neuroscience, vol. 11, no. 7, pp. 843-850, 2008
- [41] P. Ekman and W.V. Friesen, "Unmasking the face: A guide to recognizing emotions from facial clues," NJ: Prentice-Hall, Englewood Cliffs, 1978
- [42] T. Kanade, J.F. Cohn and T. Yingli, "Comprehensive database for facial expression analysis," in Proceedings of Fourth IEEE Int. Conf. Automatic Face and Gesture Recognition, pp. 46-53, 2000