# Proceedings of the UEC International Mini-Conference for Exchange Students on Informatics & Engineering and Information Systems

# No.35

The University of Electro-Communications
Center for International Programs and Exchange
March 3-4, 2016

国立大学法人
電気通信大学

# Facial expression recognition in the wild

*HERNANDEZ MATAMOROS Andres Gerardo* *
UEC Student No. 1595002
*PEREZ MEANA Hector*
National Polytechnic Institute
Mexico City,Mexico

*NAGAI Takayuki*
Department of Mechanical Engineering and Intelligent Systems
The University of Electro-Communications
Tokyo, Japan

**Abstract**

In the last years the study of facial expression has grown up but it remains limited to narrow small vocabularies of emotion into videos. In this study, we tackle the challenge of recognizing emotions through the facial expression into activities "in-the-wild" adding the accuracy rate for each expression. We propose a solution that takes a short video clip along with description sentences, it describes the main activity in the video. The action, as we have a brief sentence that describe the action of the video, it is possible that we will set a relation between the actions and emotions into the video. To recognize the facial expression we modify our previous work,it detects and segments automatically the regions of interest (ROI) adding the detection of different profile of face(left, straigth and rigth).We then combine a classifier based on clustering, it has the advantage that if a new class (emotion) is added, it is not necessary to train this completely.The obtained results is so interesting, we can recognize the emotion principal in a video using 7 facial expressions, for other hand, we can assign the accuracy rate of facial expression for daily actions for example: answer phone, get our car, sit up .

**Keywords:** Projectives Integrals,Modal Value, Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA)

## 1 Introduction

The use of smart devices in the solution of several problems has increased recently given as a result the development of very efficient systems that can be used in many practical applications. Among them the facial expression recognition (FER) systems have been used to recognize the mood. This is because several problems can be avoided if it is possible to accurate detect the mood of a person, i.e., if a given person has a nervous breakdown, if he is tired, angry or happiness, etc. For this reason, during the last several years the interest for developing these kinds of systems has increased. A very important part of such systems is the detection of face regions because an accurate detection of such regions may improve the performance of FER. Currently, in the literature exists some algorithms able to detect faces in an image and even smiles must of them based on Viola-Jones algorithm. Unfortunately these schemes are not enough accurate to detect the facial expression and thereby to achieve accurate mood detection. This happens because when someone does a mood a expression, it could be strong or not, some movement of the face muscles is done involuntarily. This movement is, in general, different in each facial expression doing it possible to determine the regions of interest of the face in each case. Several problems are present in facial expression recognition; some of them are related with the face orientation related to the camera, because if the person isn't looking straightforward to the camera partial occlusion of the face may occur, or the presence of shadows due to poor illumination conditions. For this reason, we propose a FER algorithm that is able to detect when the face is straigthforward to the camera, after segment the face ROI under different illumination conditions, after the ROI estimation, each region is segmented in a set of NxM blocks to get the characteristic vector using the modal value. The resulting features matrix is then applied to a PCA and LDA for dimensionality reduction. We use a classifier with low computational cost which provides recognition rates similar to those provided by other high performance classifiers such as the SVM and ANN. The proposed algorithm was trained with KDEF data base which consists of 490 images which are divided into seven facial expressions of 70 people and it was tested with HOHA database which consists of 32 videos which are divided in 8 actions. Evaluation results show that using the proposed system, we

*the author is supported by JASSO Scholarship.

able to recognize the facial expression in a video. The rest of the paper is organized as follows: Section 2 describes the system framework, the experimental results are shown in Section 3 and Section 4 provides the conclusion of this work.

## 2 Proposed System

The System framework of proposed facial expression recognition (FER) system is shown in 1. Here, firstly the received face image is fed into the face extraction stage, which extracts the face image using the Viola-Jones algorithm, in the next stage we determine the profile (half left, straight, half rigth). Next the straight profile is inserted into the region of interest (ROI) detection stage, which firstly estimates the face dimensions. Then using this information the ROIs are automatically segmented to get the mouth and Forehead/ Eyes using the image moments and projective integrals. These ROIs are then segmented in 35x40 no overlapping blocks,The modal value is calculated for each block, this gets as result a vector with 1400 dimensions for each ROI .Next the feature vectors are independently process by a PCA and LDA for dimensionality reduction. Finally, the resulting vectors are fed in to the classifier stage to take the final decision. Next sections provide a description of all stages of proposed system.

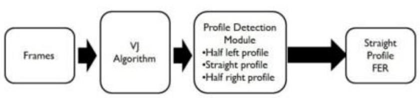



Figure 1: Proposed System

### 2.1 Profile Detection

Firstly,we use the Viola-Jones Algorithm to detect the face in a frame on a video, we implemented this algorithm using minimum window size at a third of its width and height respectively.The image received by the Viola-Jones algorithm is redimensioned at size 350x350, the next step we equalize the image [1], after we apply the YIQ color conversion using the equations 1-3 on the original image and the equalized image.We can see the system in 2

$$Y = 0.2989 * R + 0.5870 * G + 0.1140 * B \quad (1)$$

$$I = 0.5960 * R - 0.2740 * G - 0.3220 * B \quad (2)$$

$$Q = 0.2110 * R - 0.5230 * G + 0.3210 * B \quad (3)$$

In order to detect the face parts in an image, the pixels of both images are thresholded [1], the threshold value is given by the following equation.

$$(60 < Y < 200) AND (20 < I < 50) \quad (4)$$

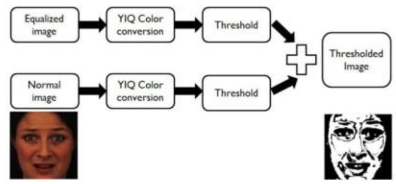



Figure 2: Profile Detection Diagram

The next step is add both umbralized images,so we have only one image, this image is divided widthwise in 7 parts and for each part we determine the percentage of information, with this we can determine the profile of the face namely if the percentage of information is higher in the center than at the sides the profile is straight, for other hand if the percentage of information is higher in the rigth side than at the left sides and center sides the profile is half left profile (that is because we see the left side of the face), by last if the percentage of information is higher in the left side than at the rigth sides and center sides the profile is half rigth profile.In this paper only uses the straigth profile because it have more information than the other profiles.

### 2.2 ROI Segmentation

The detected face image may contain noise, such as the hair or ears, which does not contain relevant information for the facial expressions; or the background where the photograph was taken. In order to eliminate this problem that may decrease the recognition rate of the proposed face expression recognition system, a more accurate estimation for face dimension is carried out.

### 2.3 Adjusment of face dimension

To adjust the face dimension parameters, firstly the color face image is divided into its three color components: Red, Green and Blue channels. Next the red and green channels are subtracted from them to highlight the skin as shown in .Finally, the resulting image is binarized using 6-6

$$I(x, y) = 0; I(x, y) < 1 \quad (5)$$

$$I(x, y) = 255; I(x, y) \geq 1 \quad (6)$$

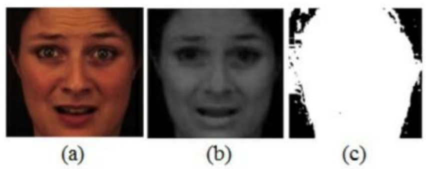



Figure 3: (a) Original image, (b) Image obtained by the subtraction, (c) binarized image

After the image binarization, shown in /refda1(c), the moments of the resulting image are estimated as follows [1]

$$M_{pq} = \sum_{x=1}^{N}\sum_{y=1}^{M} x^p y^q I(x,y) \tag{7}$$

where I(x,y) is the image intensity at position (x,y), N is the number of columns and M is the number of rows in the image; while p and q define the moment of the image. Next using 7 the centriod can be estimated as follows:

$$x_c = \frac{M_{1,0}}{M_{0,0}} \tag{8}$$

$$y_c = \frac{M_{0,1}}{M_{0,0}} \tag{9}$$

Next, using 8 and 9 the following variables are defined

$$a = \frac{M_{2,0}}{M_{0,0}} - x_c^2 \tag{10}$$

$$b = 2(\frac{M_{0,1}}{M_{0,0}} - x_c y_c) \tag{11}$$

$$c = \frac{M_{0,2}}{M_{0,0}} - y_c^2 \tag{12}$$

Next using 10-12 the face image width can be estimated as follows:

$$W = \sqrt{\frac{(a+b) - \sqrt{b^2 + (a-c)^2}}{2}} \tag{13}$$

Using 13, the left, Xl, and right, Xr, edges of the face image can be estimated as

$$Xl = x_c - \frac{W}{2} \tag{14}$$

$$Xr = Xl + W \tag{15}$$

Next, using W the upper edge of the face image can be estimates as follows

$$Yu = y_c - 0.84\frac{W}{2} \tag{16}$$

From 14-16 the face image can be segmented as follows

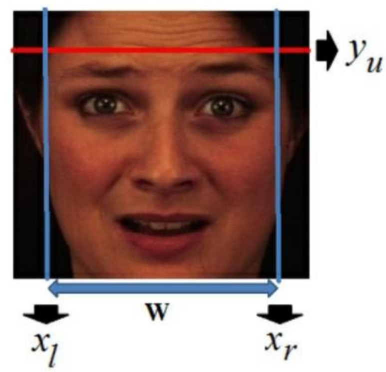



Figure 4: Segmented face region

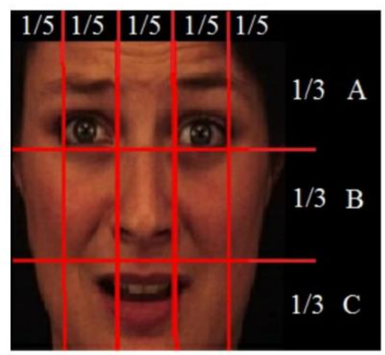



Figure 5: Symmetrical relationship of the face

## 2.4 Forehead/ eye segmentation

A very important part of proposed face expression recognition system is the fore-head/face segmentation. To this end, the segmented face region is divided into three regions from the top (A, B and C) as shown in 5, where the region A is the ROI in the case of the Forehead/eyes segmentation task.

## 2.5 Mouth Segmentation

To perform the segmentation of the mouth region, consider the segmented face re-gion which is divided into three regions, of same high, and take the C region of 5 as our ROI, but unlike the Forehead/eyes region, in this case it is necessary to segment only the mouth region. To this end, the Red and Green image's channels are sub-tracted among them, then a histogram equalization was performed of image ob-tained above, obtaining an image as shown in 6.

Figure 6: Equalized version of the image obtained from the subtraction of red and green planes

The next step for the automatic segmentation of the mouth region is the estima-tion of the horizontal projective integral which is the average of the pixel values of each column. This is a vector containing the average value of the pixels in each column of the image inside the ROI. 7 shows the horizontal projective integral estimated using the equalized image shown in 6. Next we obtain the maximum value of the projective error, which will be denoted as "D". Then using the value "D" the left border of the ROI containing the mouth is estimating by subtracting D from xc, i. e. while the right border is obtained adding it to xc, keeping the original image height, as shown in refda6, with this procedure the region of interest is extracted automatically.

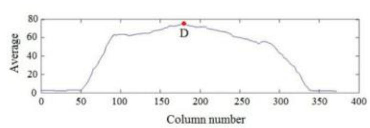



Figure 7: Horizontal projective integral of the mouth ROI

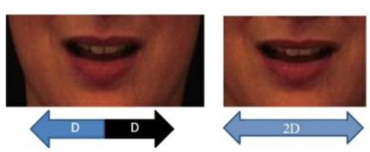



Figure 8: Detection of mouth ROI

# 3 Feature extraction

To perform the feature extraction, each one of the detected ROIs is divided in 35x40 blocks which are characterized by the modal value of the pixels for each block. Next the resulting features vector of each training ROI, with 35x40 elements, are arranged in a matrix form and applied to a PCA and LDa stage for dimensionality reduction. Next sections provide a brief description of these stages.

## 3.1 Principal component analysis

Principal component analysis (PCA) is a standard tool in modern data analysis in diverse elds from computer science because it is a simple, non-parametric method for extracting relevant information from confusing data sets. That uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables, for this paper the number of principal components is N-1 original variables.

## 3.2 Linear Discriminant analysis

Linear discriminant analysis (LDA) is a method used in pattern recognition to find a linear combination of features that characterizes or separates two or more classes of objects or events. The resulting combination may be used as a linear classifier, or, more commonly, for dimensionality reduction before later classification.LDA is also closely related to principal component analysis (PCA) and factor analysis in that they both look for linear combinations of variables which best explain the data. LDA explicitly attempts to model the difference between the classes of data. PCA on the other hand does not take into account any difference in class, and factor analysis builds the feature combinations based on differences rather than similarities.

## 3.3 Classification stage

A low computational complexity classification method is used, which uses a supervised training approach, like the ANN or SVM approaches, with the characteristic that if a new class must be added, it is not necessary to train the system with all patterns again but only with the patterns belonging to the new class.

# 4 Results

The HOHA database consists in 150 videos, these was analyzed between Figure /refre1- /refre4 , we can see the result to 4 frames in a video As Good As It Gets - 01766.avi with action is Kiss, the results to the figure /refre1- /refre3 are similar, that is because the facial expression is similar but to figure /refre4 the result is so different, The goal is recognize the facial expression ina video, to get this we apply the modal value to all frames in a video to get the result that figure /refre5 shows to this video, In /refre6 we have to the video Butterfly Effect, The - 02093.avi where the action is Stand Up.

## 5 Conclusions

This paper presents an algorithm for recognizing facial expressions, performing an automatic segmentation of facial regions of interest to achieve this, first what is pro-posed is a segmentation of the face image obtained through the algorithm Viola Jones, then based on the symmetry of the face and using the integral projective automatically remove the 2 regions of interest, the first region is the Forehead / Eye and the second the Mouth, here is important that adequate extraction of regions even with different luminescence is achieved, this one of the main problems that present for facial expression recognition, moreover a classifier is proposed with low computa-tional cost which performs better than an ANN, in both the percentage of recognition, as in training time. When making a comparison with the literature we can conclude that the proposed system performs better than [6] and [2], because a higher percentage of recognition in all possible cases was obtained and also a facial expression is recognized more accurately. Proposed system provides similar performance that [13] which use the whole image. Thus, it is possible to conclude that our system is able to recognize adequately the facial expressions with a percentage higher than 97, either taking the whole face, which in our case consists of regions of interest concatenated, or with partial occlusion, that is only considering one of the regions of interest proposals.

## 6 Acknowledgments

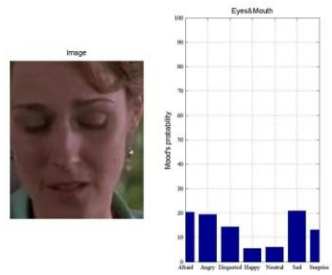
We thank the JASSO for the financial support during the realiza-tion of this research.

## References

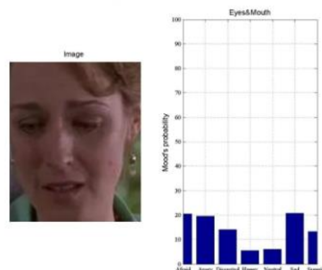
[1] 1. Y. Tian, T. Kanade, and J.F. Cohn, "Facial Expressions Analysis, Handbook of Face Recognition' Springer-Verlag, Dec, 2004.

[2] 2. L. Zhang, D. Tjondronegoro, V. Chandran: Random Gabor based templates for facial expression recognition in images with facial occlusion, Neurocomputing 145 451464,(2014).

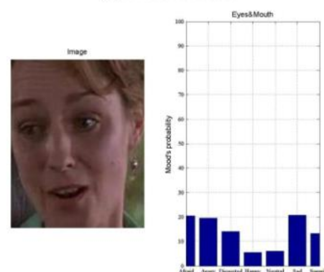
[3] K. Buciu, and I. Pitas: An analysis of facial expression recognition under partial face image occlusion, Image Visi. Comput, 26 (7),1052-1067 (2008).



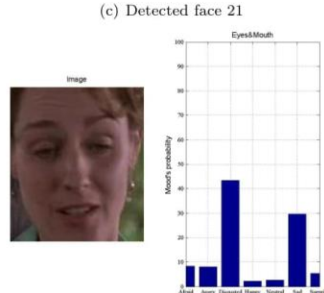

(a) Detected face 3




(b) Detected face 9

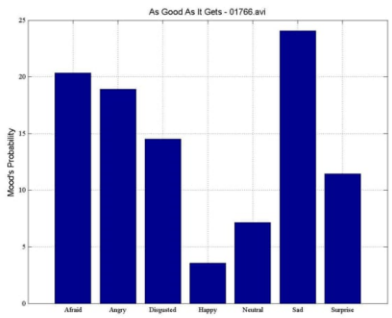



(c) Detected face 21

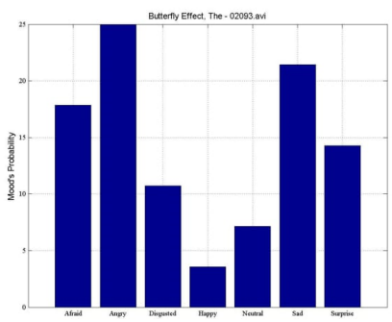



(d) Detected face 26

Figure 9: FER, As Good As It Gets - 01766.avi, Action: Kiss




(a) FER, As Good As It Gets - 01766.avi, Action: Kiss

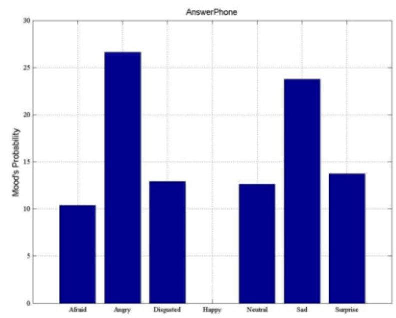



(b) FER, Butterfly Effect, The - 02093.avi, Action: Stand Up

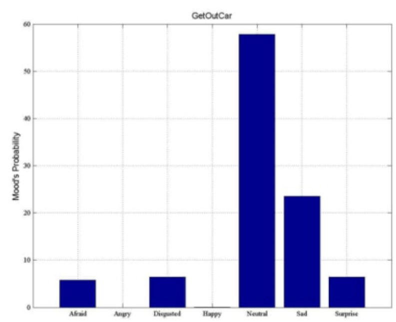
Figure 10: FER, As Good As It Gets - 01766.avi, Action: Kiss


[4] Y. Miyakoshi and S. Kato: Facial emotion detection considering partial occlusion of face using Bayesian Network", 2011 IEEE Symposium on Computers and Informatics (ISCI), 96-101 (2011).

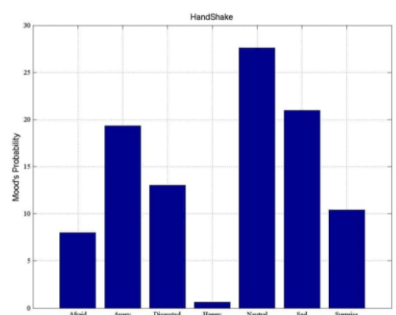
[5] P. Viola and M. Jones: Rapid object detection using a boosted cascade of sim-ple features, Computer Vision and Pattern Recognition, 511-518 (2001).


(a) Answer Phone




(b) Get Out Car




(c) Hand Shake