

Identification of Facial Gestures using Principal Component Analysis and Minimum Distance Classifier

Sandipan Banerjee
Undergraduate Student, Dept. of CSE
National Institute of Technology,
Durgapur, India

G. Sanyal, PhD
Professor, Dept. of CSE
National Institute of Technology,
Durgapur, India

ABSTRACT

Face recognition has been one of the most active fields of research for the last couple of decades and vision enthusiasts have proposed many ways to identify and recognize correctly a person from the image of his face. Here we propose an algorithm which not only detects the person's face but also identifies the expression or gesture on his face. The method at first detects the face region from a 2-dimensional image of a given person and then are projected on the "Face Space" which is the feature space containing the Eigen Vectors of the face images. The face template that has the minimum distance from the test image is evaluated as the best match and the expression label it possesses is therefore taken as the required expression of the test image.

General Terms

Image Processing, Pattern Recognition.

Keywords

Face Detection, Gesture Recognition, Minimum Distance Classifier, Principal Component Analysis (PCA).

1. INTRODUCTION

Faces are one of the most complex entities that can be found in a human being. Because of its intricate structure, faces differ from a person to person by a huge margin. Even it can be found out that the faces of a pair of identical twins vary a lot when the structural components are properly analyzed. So developing a computational model for detecting and identifying the face of a person perfectly in the first place, let alone identifying its expression, is an extremely difficult task owing to the minute structural features in hand. Therefore, it is better if a set of properly pre-processed faces are kept in hand and we can train the classifier using this set of pre-processed faces. Once the training is done, the test images are fed to the classifier which in turn gives us the result produced due to this projection.

Most of the earlier works on facial gesture recognition have been carried out with the configuration of the different regions of the face and their related features in mind. Basically some nodal points which defined a face like the eyes, the nose, the ears, the mouth etc. were chosen as the active regions for classification and the immediate relation between them gave us the idea about the expression on the face of a person. But because of this configuration based system the algorithm cannot be extended to cover for multiple views. Although

these type of methods are more popular among vision enthusiasts, their action steps lack a holistic approach and therefore can never match up to the cerebral capacity of human beings – which basically is the sole aim of machine learning. So, we have proposed here a gestalt – like system which takes the whole face as a single unit and saves the important features from it with the help of Principal Component Analysis. Then the test image is projected on to these training faces with their individual features and the nearest match gives us the correct expression.

The organization of the paper is as follows – Section 2 deals with our proposed method for recognition of human face expressions, Section 3 covers the algorithm, in Section 4 we discuss our experiments and the generated results while Section 5 concludes it.

2. THE PROPOSED METHOD

The algorithm comprises of two stages, one for getting the "face space" by applying PCA on the training images and the next to project the test image onto the face space and finding out the best match from the lot in order to extract the required facial expression. Here are the two stages in details:-

2.1. Stage 1

Suppose, there are a total of M different images of a face with different levels of a fixed set of emotions attached to them. Let S denote this whole set of initial training images. So, S can be represented as,

$$S = \{t_1, t_2, t_3, \dots, t_M\}$$

where t_i represents each face in the training set and $i = 1, 2, 3, \dots, M$.

Given the set of initial images we can easily find out the mean image from the lot and denote it as t' . So,

$$t' = \frac{1}{M} \sum_{i=1}^M t_i$$

and so we get the mean average image of the distribution t' .

Now, for carrying out PCA on the set of images we need to get the covariance between the images from the mean image. Therefore we calculate the variance for each image as,

$$V_i = t_i - t' \text{ for all } i = 1, 2, 3, \dots, M.$$

So now we have a set of M vectors which measure the variance of each of the M images from the mean. These vectors are now ready for being subjected to Principal Component Analysis.

By the PCA method, we find out a set of M Eigen vectors and their corresponding Eigen values u_k and v_k respectively where the former is a vector and the latter is a scalar quantity and $k = 1, 2, 3, \dots, M$. These two quantities are found out from the covariance matrix C of the set of images where

$$C = \frac{1}{M} \sum_{i=1}^M V_i V_i^T = A \cdot A^T$$

where A represents the initial matrix of the vectors V_k and A^T represents the transpose matrix of A .

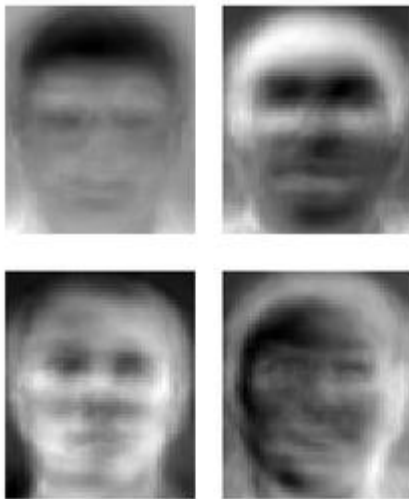


Fig 1. An example displaying a set of Eigen Image

For an initial number of M images which were trained into the system a total of M Eigen vectors and their corresponding Eigen values are found out. But as mentioned earlier, not all the Eigen values contribute to the same extent to the distribution and therefore the ones having a contribution below a pre-fixed threshold can be neglected all together. Suppose we skim through our set of M Eigen values and get a set of M' new meaningful Eigen values for the distribution. This set of M' trained faces is known as the Face Space and each of the faces are known as Eigen faces in the Face Space.

2.2. Stage 2

Now that we have the Face Space for the given set of images and the respective M' Eigen faces for the different emotions so now we can test any new image on to this space by projecting it directly to this set in an effort to match itself to any of the face templates. The one expression and hence Eigen face which matches to the test image perfectly will have the least Euclidean Distance from it in the whole set.

Suppose, there is a new image in the test set denoted by λ then we project it to the Face Space as the following:

$$w_k = u_k^T (\lambda - t')$$

where w_k is the weight for that component, u_k is the Eigen vector and t' is the mean Eigen face in the Face Space and $k = 1, 2, 3, \dots, M'$.

So for an input image we get a set of weights $\Omega^T = \{w_1, w_2, w_3, \dots, w_{M'}\}$. Now the face components in the Eigen Space have been pre-defined to be part of an expression class before hand and has been labeled so in the training process. So now when we compute the Euclidian Distance between the each of the contribution of the M' Eigen faces to the new image and the mean, we can compare and get the one with the minimum distance attached to it such that $\epsilon_k = \|\Omega - \Omega_k\|$ for the k 'th class. So the class with the least value of ϵ is the class to which the test image matches the best and therefore it is the class to which the image belongs to.

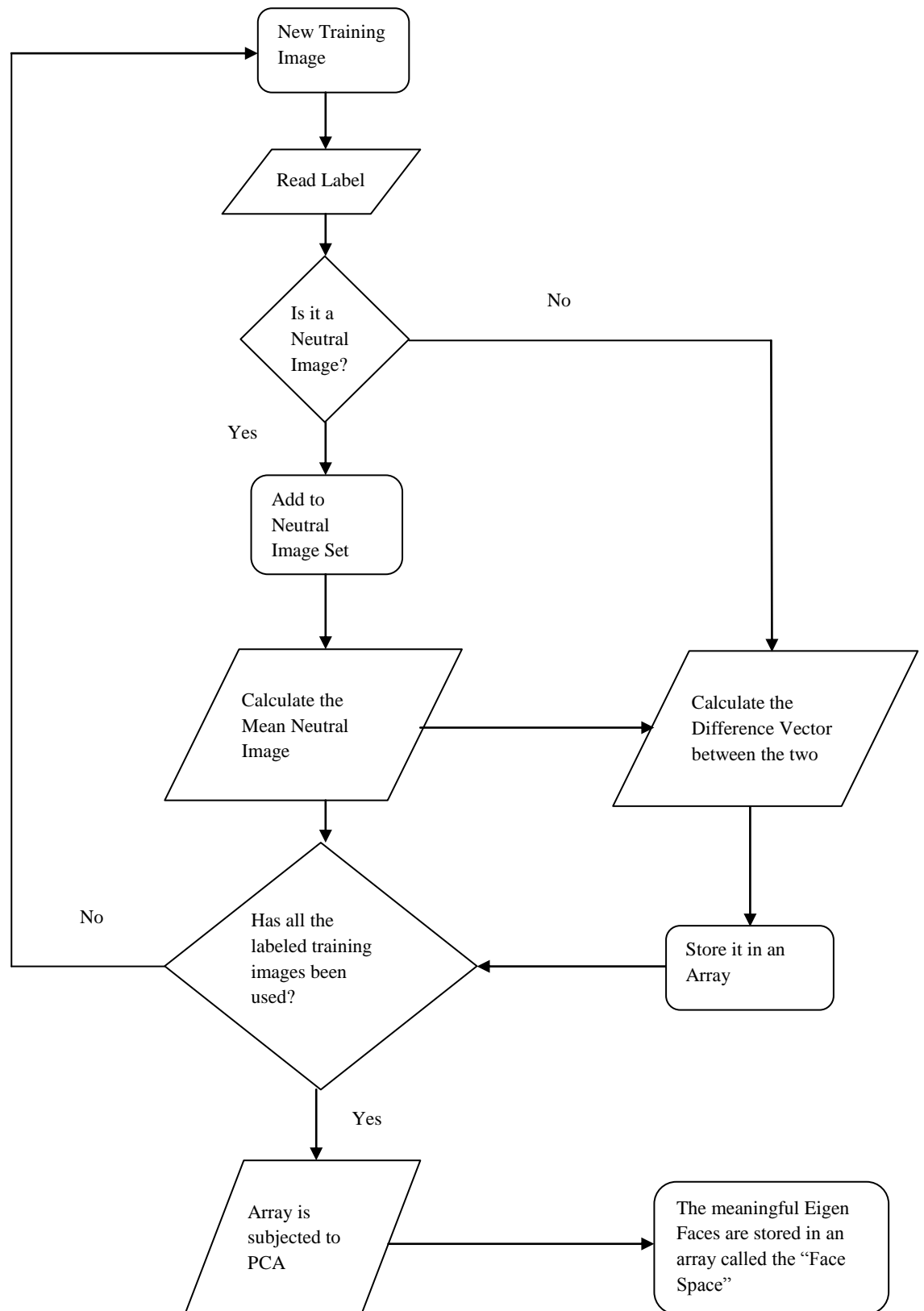
In our experiment specifically, each of the Eigen faces in the Face Space represent any of the basic facial expressions and therefore when a new image is projected on to it and matches with any given vector k the best in the set, then the image can be said to comprise of the same facial expression depicted in the image k . The intensity of the expression can be computed by calculating its distance from the mean neutral image and the further away it is from the mean, the more intense the expression in magnitude.

3. THE COMPUTER ALGORITHM

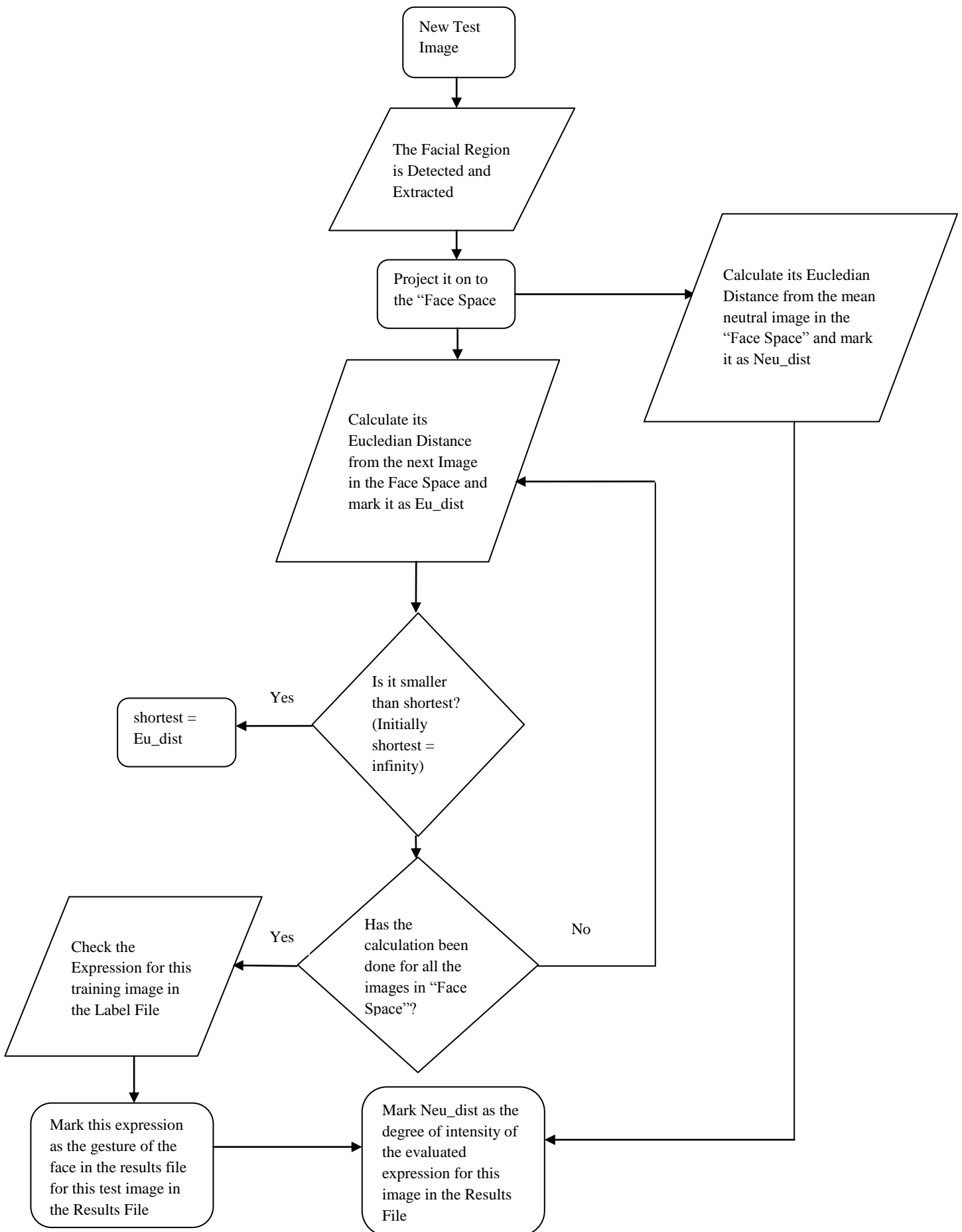
For the computation purpose, the sample set of facial images were divided into two different sets: the training image set and the testing image set. The training set contained images which were used to train the classifier and each individual image had a particular number assigned to it. This number corresponded to a particular expression class in the Label File which was also fed to the classifier. Thus when training was done, each of the individual training images pointed towards a expression class in the label file which correctly indexed the facial gesture of the image at hand.

The results thus evaluated are stored in the Results File with the facial expression being stored as the required gesture of the facial test image at hand and the shortest Euclidean Distance of the image from its closest match giving the degree of intensity of the expression from the neutral state. These fields are stored directly into the Results File.

3.1. Training



3.2. Testing



4. EXPERIMENTAL RESULTS

For the experiment we compiled a set of 60 training images which had face images having any one of the following five basic expressions – “neutral”, “happy”, “sad”, “disgust” and “anger” of the same individual with a number of images having the different intensities of the same expression like in Figure 3. At first the face region in each of the images was extracted out with the help of smoothing of pixel values and subsequently connected component labeling which gave the face region as the Bounding Box content as a result. Because of the initial 60 training images with each contributing its own to the Face Space, a total of 60 Eigen vectors were found out in the beginning. But the meaningful Eigen vectors were calculated and the first 30 Eigen vectors were taken into considered only and put into the Face Space.



Fig 2. The different expressions – “neutral”, “happy”, “sad”, “anger” and “disgust” respectively

A total of 45 test images were fed into the system (of the same individual) for testing the algorithm after training was finished. Once the face region was detected in each of the training face images, it was projected on to the Face Space and was matched with each of the 30 Eigen Faces. The one which gave us the minimum distance of the lot was taken as the best matching destination image and the test image was included in its expression class. From the set of neutral images, the mean neutral image was computed and the

Euclidian Distance between this mean and the test image gave us the intensity of the expression of the individual in question. Because the process revolves around template-matching therefore adding more and more images in the training set would ensure a better performance as then more templates would there to match to for an image in the test set.



Fig 3. The different intensities for the expression “sad”

The algorithm was implemented using the MATLAB programming language and a comprehensive use of its huge collection of in-built image processing functions was made. The 60 training images had a Label File with them to be fed into the system such that the expression in each of the 60 images was mentioned in the file beside the corresponding image name. After the classifier processed the training images and got hold of the facial regions in each of the 60 images, PCA was carried out in each of them for getting the orthogonal components in the distribution. The 60 coefficients thus found out were then narrowed down to 30 by the Proportionality test such that the meaningful contributing components only remain in the Face Space. And 5 different expression classes were constructed from these components with each belonging to any one of them. When any new test image was input into this system, it projected it straight on to this set and calculated its distance from each of the 30 Eigen faces to get the one with the minimum value. This extracted image matched best to the test image and its corresponding face class was the expression of the test subject in question. The distance of the image from the mean neutral image was also computed to get an idea about the magnitude of the facial gesture of the test individual. The output result file contained the test image name, its distance from the mean neutral (and thus its intensity), the facial expression class it belongs to and the image name it matches to best in the training set. Thus the facial gesture recognition system was found to be simple, fast and efficient.

Most of the test images were properly classified and were matched to their closest counterpart in the training image set. But a couple of test images were classified in an improper manner due to any of the two cases – i) A few of the test images which were actually neutral were marked as happy with a low intensity quotient because of the close boundary between the two classes in the training set like shown in Figure 4. Varying the training images however with only intense expressive faces resolved this issue.



Fig 4. The “happy” and “neutral” expressions which are almost similar

ii) The method was tested on facial images with different expressions of the same person and therefore might not do well when images of a number of persons are present in the training and test images.

Training Images	Test Images	Test Images Successfully Classified	Efficiency
60	45	42	93.33%

5. CONCLUSION

The proposed method deals with facial images in a 2-D surrounding and therefore is appropriate when the images of the individuals are already in possession a priori. But the algorithm can be extended in such a way that it tracks a surrounding and from the video feed extracts frame by frame images of a person to get the different emotions on his/her face at different intervals.

The algorithm can also be extended to classify different expressions in the images of faces of different individuals rather than doing the same on one person. If this can be done then a room full of people can be classified according to their emotions and in that case we would get a robust vision system suitable for real time purposes.

6. REFERENCES

- [1] Weiming Hu and Tieniu Tan, “A Survey on Visual Surveillance of Object Motion and Behaviours”, IEEE Transactions on Systems, Man and Cybernetics-Part C, Application and Reviews, Vol. 34 No. 3, August 2004.
- [2] M. Turk and A. Pentland, “Eigenfaces for Recognition”, Journal of Cognitive Neuroscience, 1991.
- [3] M. Turk and A. Pentland, “Face Recognition using Eigenfaces”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1991.
- [4] Giovanni Garibotto and Carlo Cibei, “3D Scene Analysis by Real-Time Stereovision”, Proceedings of the IEEE Conference on Image Processing, 2005.
- [5] T. Tian and C. Tomasi, “Comparison of Approaches in Egomotion Computation”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 315-320, 1996.
- [6] Chung-Lin Huang and Yu-Ming Huang, “Facial Expression Recognition using Model Based Feature Extraction and Action Parameter Classification”, Journal of Visual Communication and Image Representation, Vol. 8 No. 3, pp. 278-290, 1997.
- [7] A.J. Calder, A.M. Burton, P. Miller Brown, A.W. Young, S. Akamatsu, “A Principal Component Analysis of Facial Expressions”, Elsevier Vision Research 41, pp. 1179-1208, 2001.
- [8] L. Di Stefano and A. Bulgarelli, “A Simple and Efficient Connected Component Labeling Algorithm”, Proceedings of the International Conference on Image Analysis and Processing, pp. 322-327, 1999.
- [9] B.K. Gunturk, A.U. Batur, Y. Altunbasak, M.H. Hayes III, R.M. Marsereau, “Eigenface based super-resolution for face recognition”, Proceedings of International Conference on Image Processing, 2002.
- [10] O.D. Trier and A.K. Jain, “Goal-directed Evaluation of Binarization Methods”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 17 No.12, pp. 1191-1201, 1995.