Face Emotion Analysis Using Gabor Features In Image Database for Crime Investigation

Lt. Dr. S. Santhosh Baboo, Reader santhos2001@sify.com
P.G & Research, Dept. of Computer Application,
D.G. Vaishnav College,
Chennai-106. INDIA

V.S. Manjula manjusunil.vs@gmail.com
Research Scholar, Dept. of Computer Science & Engineering,
Bharathiar University,
Coimbatore-641 046. INDIA

Abstract

The face is the most extraordinary communicator, which plays an important role in interpersonal relations and Human Machine Interaction. Facial expressions play an important role wherever humans interact with computers and human beings to communicate their emotions and intentions. Facial expressions, and other gestures, convey non-verbal communication cues in face-to-face interactions. In this paper we have developed an algorithm which is capable of identifying a person's facial expression and categorize them as happiness, sadness, surprise and neutral. Our approach is based on local binary patterns for representing face images. In our project we use training sets for faces and non faces to train the machine in identifying the face images exactly. Facial expression classification is based on Principle Component Analysis. In our project, we have developed methods for face tracking and expression identification from the face image input. Applying the facial expression recognition algorithm, the developed software is capable of processing faces and recognizing the person's facial expression. The system analyses the face and determines the expression by comparing the image with the training sets in the database. We have followed PCA and neural networks in analyzing and identifying the facial expressions.

Keywords: Facial Expressions, Human Machine Interaction, Training Sets, Faces and Non Faces, Principal Component Analysis, Expression Recognition, Neural networks.

1. INTRODUCTION

Face expressions play a communicative role in the interpersonal relationships. Computer recognition of human face identity is the most fundamental problem in the field of pattern analysis. Emotion analysis in man-machine interaction system is designed to detect human face in an image and analyze the facial emotion or expression of the face. This helps in improving the interaction between the human and the machine. The machines can thereby understand the man's reaction and act accordingly. This reduces the human work hours. For example, robots can be used as a class tutor, pet robots, CU animators and so on.. We identify facial expressions not only to express our emotions, but also to provide important communicative cues during social interaction, such as our level of interest, our desire to take a speaking turn and continuous feedback signaling or understanding of the information conveyed. Support Vector Algorithm is well suited for this task as high dimensionality does not affect the Gabor Representations. The main disadvantage of the system is that it is very expensive to implement and maintain. Any changes to be upgraded in the system needs a change in the algorithm which is very sensitive and difficult; hence our developed system will be the best solution to overcome the above mentioned disadvantages.

In this paper, we propose a complete face expression recognition system by combining the face tracking and face expression identifying algorithm. The system automatically detects and extracts the human face from the background based on a combination of a retrainable neural network structure. In this system, the computer is trained with the various emotions of the face and when given an input, the computer detects the emotion by comparing the co-ordinates of the expression with that of the training examples and produces the output. Principle Component Analysis algorithm is the one being used in this system to detect various emotions based on the coordinates of the training sample given to the system.

2. RELATED WORK

Pantic & Rothkrantz [4] identify three basic problems a facial expression analysis approach needs to deal with: face detection in a facial image or image sequence, facial expression data extraction and facial expression classification. Most previous systems assume presence of a full frontal face view in the image or the image sequence being analyzed, yielding some knowledge of the global face location. To give the exact location of the face, Viola & Jones [5] use the Adaboost algorithm to exhaustively pass a search sub-window over the image at multiple scales for rapid face detection.

Essa & Pentland [6] perform spatial and temporal filtering together with thresholding to extract motion blobs from image sequences. To detect presence of a face, these blobs are then evaluated using the eigenfaces method [7] via principal component analysis (PCA) to calculate the distance of the observed region from a face space of 128 sample images. To perform data extraction, Littlewort et al. [8] use a bank of 40 Gabor wavelet filters at different scales and orientations to perform convolution. They thus extract a “jet” of magnitudes of complex valued responses at different locations in a lattice imposed on an image, as proposed in [9]. Essa & Pentland [6] extend their face detection approach to extract the positions of prominent facial features using eigenfeatures and PCA by calculating the distance of an image from a feature space given a set of sample images via FFT and a local energy computation.

Cohn et al. [10] first manually localize feature points in the first frame of an image sequence and then use hierarchical optical flow to track the motion of small windows surrounding these points across frames. The displacement vectors for each landmark between the initial and the peak frame represent the extracted expression information. In the final step of expression analysis, expressions are classified according to some scheme. The most prevalent approaches are based on the existence of six basic emotions (anger, disgust, fear, joy, sorrow and surprise) as argued by Ekman [11] and the Facial Action Coding System (FACS), developed by Ekman and Friesen [12], which codes expressions as a combination of 44 facial movements called Action Units. While much progress has been made in automatically classifying according to FACS [13], a fully automated FACS based approach for video has yet to be developed.

Dailey et al. [14] use a six unit, single layer neural network to classify into the six basic emotion categories given Gabor jets extracted from static images. Essa & Pentland [6] calculate ideal motion energy templates for each expression category and take the euclidean norm of the difference between the observed motion energy in a sequence of images and each motion energy template as a similarity metric. Littlewort et al. [8] preprocess image sequences image-by-image to train two stages of support vector machines from Gabor filter jets. Cohn et al. [10] apply separate discriminant functions and variance-covariance matrices to different facial regions and use feature displacements as predictors for classification.

2.1. Principle Component Analysis

Facial expression classification was based on Principle Component Analysis. The Principle Component Analysis (PCA) is one of the most successful techniques that have been used in image recognition and compression. PCA is a statistical method under the broad title of factor analysis. The purpose of PCA is to reduce the large dimensionality of the data space (observed variables) to the smaller intrinsic dimensionality of feature space (independent variables), which are needed to describe the data economically. [3] The main trend in feature extraction has been representing the data in a lower dimensional space computed through a linear or non-linear transformation satisfying certain properties.

The jobs which PCA can do are prediction, redundancy removal, feature extraction, data compression, etc. Because PCA is a classical technique which can do something in the linear domain, applications having linear models are suitable, such as signal processing, image processing, system and control theory, communications, etc. [3]

Given an s-dimensional vector representation of each face in a training set of images, Principal Component Analysis (PCA) tends to find a t-dimensional subspace whose basis vectors correspond to the maximum variance direction in the original image space. This new subspace is normally lower dimensional. If the image elements are considered as random variables, the PCA basis vectors are defined as eigenvectors of the scatter matrix. PCA selects features important for class representation.

The main idea of using PCA for face recognition is to express the large 1-D vector of pixels constructed from 2-D facial image into the compact principle components of the feature space. This can be called eigenspace projection. Eigenspace is calculated by identifying the eigenvectors of the covariance matrix derived from a set of facial images(vectors). [3] we implement a neural network to classify face images based on its computed PCA features.

Methodology

Let $\{X_1, X_2, \ldots X_n\}$, $\mathbf{x} \in \Re^n$ be *N* samples from *L* classes $\{\omega_1, \omega_2, \ldots\ldots \omega_L\}$, and $p(\mathbf{x})$ their mixture distribution. In a sequel, it is assumed that a priori probabilities $P(\omega\ i)$ *, I = 1,2,...,L*, are known.
Consider **m** and **Σ** denote mean vector and covariance matrix of samples, respectively. PCA algorithm can be used to find a subspace whose basis vectors correspond to the maximum variance directions in the original *n* dimensional space. PCA subspace can be used for presentation of data with minimum error in reconstruction of original data. Let $\mathbf{\Phi}^p PCA$ denote a linear $n\times p$ transformation matrix that maps the original *n* dimensional space onto a *p* dimensional feature subspace where $p < n$. The new feature vectors,

$$Y_i = (\mathbf{\Phi}^p PCA)^t * i,\ I = 1,2,\ldots N \quad --- \quad (1)$$

It is easily proved that if the columns of $\mathbf{\Phi}^p PCA$ are the
eigenvectors of the covariance matrix corresponding to its *p* largest eigenvalues in decreasing order, the optimum feature space for the representation of data is achieved. The covariance matrix can be estimated by:

$$\Sigma = (\sum_{i=1}^{N} (x_i - m)(x_i - m)^t) / (N - 1) \quad ------ \quad (2)$$

Where *m* in (2) can be estimated by:

$$M_k = m_{k-1} + \eta (x_k - m_{k-1}) \quad ------- \quad (3)$$

Where *mk* is estimation of mean value at *k-th* iteration and *xk* is a the *k-th* input image. PCA is a technique to extract features effective for representing data such that the average reconstruction error is minimized. In the other word, PCA algorithm can be used to find a subspace whose basis vectors correspond to the maximum variance directions in the original *n* dimensional space. PCA transfer function is composed of significant eigenvectors of covariance matrix. The following equation can be used for incremental estimation of covariance matrix :

$$\Sigma_k = \Sigma_{k-1} + \eta_k (X_k X_k^t - E_{k-1}) \quad ------- \quad (4)$$

Where $\mathbf{\Sigma}k$ is the estimation of the covariance matrix at k-th iteration, $\mathbf{x_k}$ is the incoming input vector and η_k is the learning rate.

The emotion categories are : Happiness, sadness, surprise, disgust, fear, anger, neutral. In this project the emotions such as Happiness, sadness, Surprise and neutral is taken into account for human emotion reorganization.

2.2 Training Sets
The ensemble of input-desired response pairs used to train the system. The system is provided with various examples and is trained to give a response to each of the provided examples. The input given to the system is compared with the examples provided. If it finds a similar example then the output is produced based on the example's response. This method is a kind of learning by examples.

Our system will perform according to the number of training sets we provide. The accuracy of our system depends on the number of training sets we provide. If the training samples are higher the performance of the system is accurate. Sample training sets are shown in the fig[1] below,

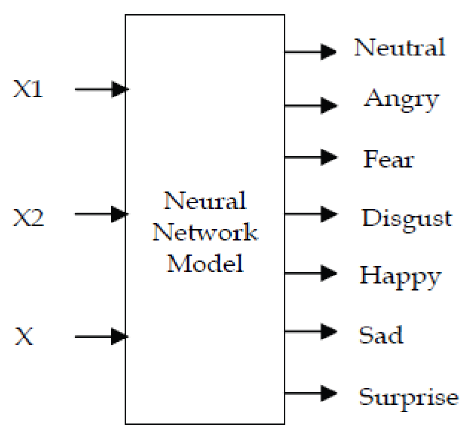
FIGURE 1: Samples of training set images

2.3. Neural Networks
Recently, there has been a high level of interest in applying artificial neural network for solving many problems**.** The application of neural network gives easier solution to complex problems such as in determining the facial expression**.** Each emotion has its own range of optimized values for lip and eye. In some cases an emotion range can overlap with other emotion range. This is experienced due to the closeness of the optimized feature values. For example, in Table, X1 of Sad and Dislike are close to each other. These values are the mean values computed from a range of values. It has been found that the ranges of feature values of X1 for Sad and dislike overlap with each other. Such overlap is also found in X for Angry and Happy. A level of intelligence has to be used to identify and classify emotions even when such overlaps occur.

Emotions	Manually Computed Mean Value (in pixels)			Optimized Mean Value by GA (in pixels)		
	b1	b2	b	X1	X2	X
Neutral	38	41	21	34.2644	35.2531	19.6188
Fear	25	41	16	23.0287	36.9529	14.7024
Happy	25	48	16	21.5929	43.4742	15.0393
Sad	33	34	19	30.9104	28.5235	16.9633
Angry	25	34	16	24.2781	30.8381	15.4120
Dislike	35	29	13	31.3409	21.6276	12.8353
Surprise	43	57	17	42.6892	55.5180	16.0701

FIGURE 2: Optimized Value of three features

A feed forward neural network is proposed to classify the emotions based on optimized ranges of 3-D data of top lip, bottom lip and eye. The optimized values of the 3-D data are given as inputs to the network as shown in Figure. The network is considered to be of two different models where the first model comes with a structure of 3 input neurons, 1 hidden layer of 20 neurons and 3 output neurons (denoted by (3x20x3)) and the other model with a structure of (3x20x7). The output of (3x20x3) is a 3-bit binary word indicating the seven emotional states. The output (Oi, i=1, 2,.., 7) of (3x20x7) is of mutually exclusive binary bit representing an emotion. The networks with each of the above listed input sizes are trained using a back-propagation training algorithm. A set of suitably chosen learning parameters is indicated in Table. A typical “cumulative error versus epoch” characteristic of the training of NN models as in Figure 10 ensures the convergence of the network performances. The training is carried out for 10 trials in each case by reshuffling input data within the same network model. The time and epoch details are given in Table 3 which also indicates the maximum and minimum epoch required for converging to the test-tolerance.



(b) NN of (3x20x7) structure

Neural Network Structures – (a & b)

Hidden neurons: 20			Learning rate: 0.0001			Activation function: $(1/(1+e^{-x}))$		
Momentum factor: 0.9			No. of samples: 70			Maximum no. of epoch: 1000		
Testing tolerance: 0.1			Training tolerance: 0.0001			No. of trained samples: 50		
NN structure	Epoch (in 10 trials)			Training Time (sec) (in 10 trials)			Classification % (in 10 trials)	
	Min	Max	Mean	Min	Max	Mean	Range	Mean
3x20x3	105	323	225	2.18	7.32	5.02	75.71 - 90.00	83.57
3x20x7	71	811	294	3.43	39.25	13.76	81.42 -91.42	85.13

FIGURE 3: Details of Neural Network Classification using Neural Networks

2.4. Feature Extraction

A feature extraction method can now to be applied to the edge detected images. Three feature extraction methods are considered and their capabilities are compared in order to adopt one that is suitable for the proposed face emotion recognition problem. They are projection profile, contour profile and moments (Nagarajan et al., 2006).

Implementation Overview

For PCA to work properly, we have to subtract the mean from each of the data dimensions. The mean subtracted is the average across each dimension. So, all the x values have subtracted, and all the y values have subtracted from them. This produces a data set whose mean is zero as shown in the fig [2].

Data =

x	y
2.5	2.4
0.5	0.7
2.2	2.9
1.9	2.2
3.1	3.0
2.3	2.7
2	1.6
1	1.1
1.5	1.6
1.1	0.9

DataAdjust =

x	y
.69	.49
-1.31	-1.21
.39	.99
.09	.29
1.29	1.09
.49	.79
.19	-.31
-.81	-.81
-.31	-.31
-.71	-1.01

FIGURE 2:

The next step is to calculate the Co-Variance Matrix. Since the data is 2 dimensional, the covariance matrix will be 2 x 2. We will just give you the result as

$$cov = \begin{pmatrix} .616555556 & .615444444 \\ .615444444 & .716555556 \end{pmatrix}$$

Since the non-diagonal elements in this covariance matrix are positive, we should expect that both the x and y variable increase together.

2.5 Eigen Vectors and Eigen Faces

Since the covariance matrix is square, we can calculate the eigenvectors and eigenvalues for this matrix. These are rather important, as they tell us useful information about our data. Here are the eigenvectors and eigenvalues,

$$eigenvalues = \begin{pmatrix} .0490833989 \\ 1.28402771 \end{pmatrix}$$

$$eigenvectors = \begin{pmatrix} -.735178656 & -.677873399 \\ .677873399 & -.735178656 \end{pmatrix}$$

It is important to notice that these eigenvectors are both unit eigenvectors that is their lengths are both 1. This is very important for PCA. Most maths packages, when asked for eigenvectors, will give you unit eigenvectors. It turns out that the eigenvector with the highest eigen value is the principle component of the data set. In our example, the eigenvector with the largest eigen value is the one that pointed down the middle of the data. It is the most significant relationship between the data dimensions.

In general, once eigenvectors are found from the covariance matrix, the next step is to order them by eigen value, highest to lowest. This gives the components in order of significance. If we leave out some components, the final data set will have less dimensions than the original. To be precise, if we originally have n dimensions in our data, and so calculate n eigenvectors and eigen values, and then choose only the first p eigenvectors, then the final data set has only p dimensions. This is the final step in PCA is choosing the components (eigenvectors) that we wish to keep in our data and form a feature vector, we simply take the transpose of the vector and multiply it on the left of the original data set, transposed.

Final Data=RowFeatureVector x RowDataAdjust

where RowFeatureVector is the matrix with the eigenvectors in the columns transposed so that the eigenvectors are now in the rows, with the most significant eigenvector at the top, and RowDataAdjust is the mean-adjusted data transposed, ie. the data items are in each column, with each row holding a separate dimension. Final Data is the final data set, with data items in columns, and dimensions along rows.

The implementation chart is as in Fig [4],

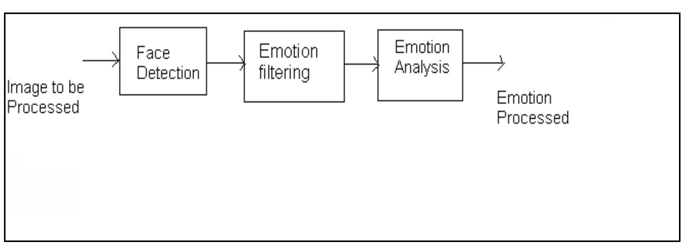


FIGURE 4: Implementation Chart

2.6 Face Detection

Face Detection follows the Eigen face approach. Once the eigenfaces have been computed, several types of decision can be made depending on the application. The steps involved in this approach are mentioned below: (1) Acquire initial set of face images (the training set). (2) Calculate Eigenvector from the training set keeping only M images (face space). (3) Calculate the corresponding distribution in the face space. The process is explained in the following flow diagram Fig [5],

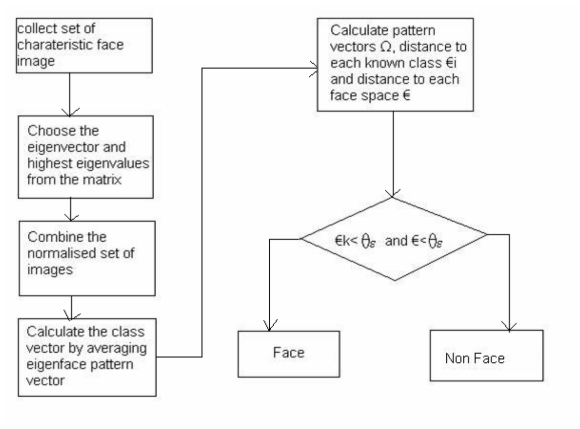


FIGURE 5: Flow Diagram

2.7 Emotion Filtering

Emotion Filtering is where in the detected image is projected over the various average faces such as Happy, Neutral, Sad, and Surprised. The filters provide the output values of the given input by comparing the input with the training samples present in the system. The input is compared to all the training sets of various emotions. The values obtained after the comparison in this module is provided to the analyzer.

2.8 Emotion Analyzer

This is the decision making module. The output of the emotion filter is passed into the emotion analyzer. There are threshold values set for each emotions. For example,

- If output value is equal to the threshold value, then it is considered as neutral
- If output value is greater than the threshold value, then it is considered as happy
- If output value is less than the threshold value, then it is considered as sad

Thus the emotion analyzer makes the decision based on the threshold values provided. By adding new threshold values one add a new emotion to the system.

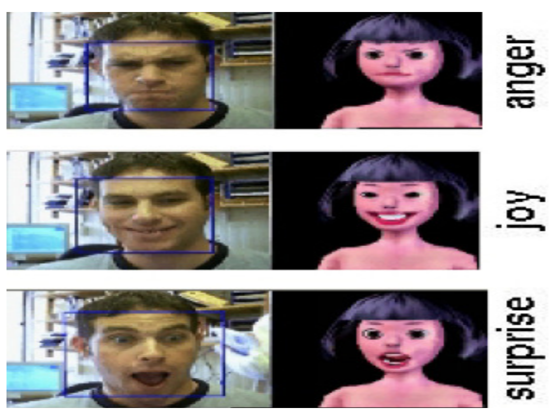


FIGURE 6: Sample Emotions

In our project, we can have nearly 300 images with which our software is completely trained. Some samples of the training set images are shown in the following figure (7). We have developed this system using Dotnet language. In our software we have developed a form with various input and functionality objects.

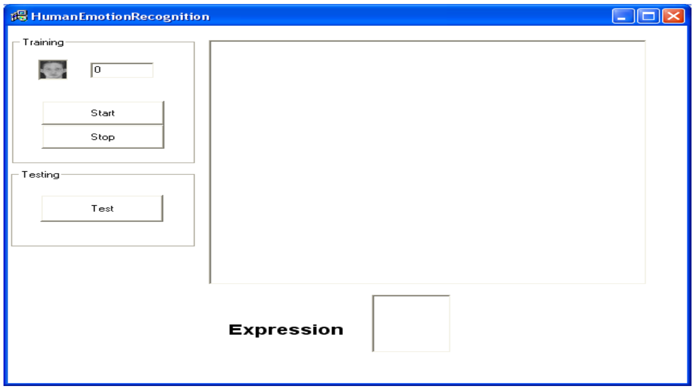


FIGURE 7

In the first phase our system runs the training sets by which system training is over. In the second phase we can stop the training and in the third phase we continue with the testing module. In the testing phase our system will start showing the emotions of the images continuously as in the following figure (8).

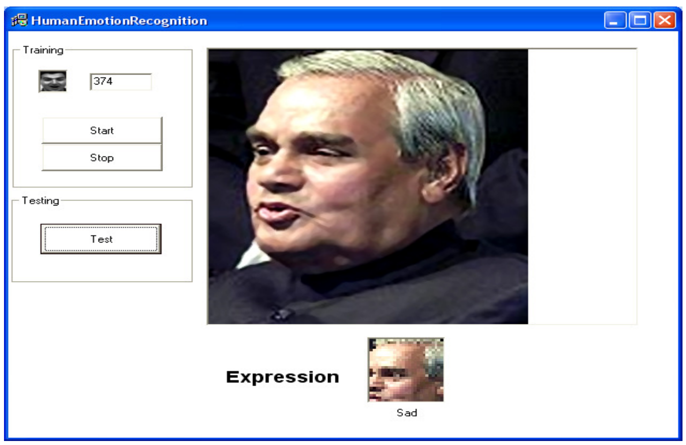


FIGURE 8

The expression in the above image seems to be sad, and the result of the system is also sad. Thus it is proven that our system is able to identify the emotions in the images accurately.

3. EXPERIMENTAL RESULTS

Face segmentation and extraction results were obtained for more than 300 face images, of non-uniform background, with excellent performance. The experimental results for the recognition stage presented here, were obtained using the face database. Comparisons were accomplished for all methods using face images. In our system, all the input images are getting converted to the fixed size pixels before going for the training session. However, this recognition stage has the great advantages of working in the compressed domain and being capable for multimedia and content-based retrieval applications.

Face detection and recognition can also be done using Markov random fields, Guided Particle Swarm Optimization algorithm and Regularized Discriminant Analysis based Boosting algorithm. In Markov Random Fields method the eye and mouth expressions are considered for finding the emotion, using the edge detection techniques. This kind of emotion analysis will not be perfect because the persons actual mood cannot be identified just by his eyes and mouth. In the Guided Particle Swarm optimization method video clips of the user showing various emotions are used, the video clip is segmented and then converted into digital form to identify the emotion. This is a long and tedious process. Our proposed system was implemented using C# programming language under .NET development framework. The implemented project has two modes, the training mode and the detection mode. The training mode is completed once if all the images are trained to the system. In the detection mode the system starts identifying the emotions of all the images which will be 97% accurate. The following figure shows the comparison of recognition rates.

Methods	Recognition Rates
Markov Random Fields	95.91
Guided Particle Swarm Optimization	93.21
Regularized Discriminant Analysis based Booting	96.51
Our Proposed method	97.76

FIGURE 9

Comparison of facial expression recognition using image database

The above results are shown as comparison graphs in the following Fig [11],

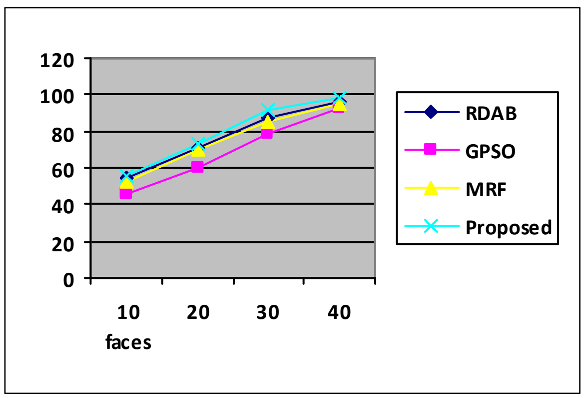


FIGURE 10

Here, we have taken number of training samples in the x-axis and recognition rate on the y-axis. The accuracy of emotion recognition is high in our proposed method compared to the other two algorithms.

4. CONCLUSION

In this paper we have presented an approach to expression recognition in the images. This emotion analysis system implemented using PCA algorithm is used in detection of human emotions in the images at low cost with good performance. This system is designed to recognize expressions in human faces using the average values calculated from the training samples. We evaluated our system in terms of accuracy for a variety of images and found that our system was able to identify the face images and evaluate the expressions accurately from the images.

REFERENCES

[1] M. Pantic, L.J.M. Rothkrantz, “Automatic Analysis of Facial Expressions”: The State of the Art, *IEEE Trans. Pattern Analysis and Machine Intelligence,* Vol. 22, No. 12, pp. 1424-1445, 2000

[2] G. Donato, M.S. Bartlett, J.C. Hager, P. Ekman, T.J. Sejnowski, “Classifying Facial Actions", IEEE *Trans. Pattern Analysis and Machine Intelligence*, Vol. 21, No. 10, pp. 974-989, 1999

[3] Face Recognition using Principle Component Analysis -Kyungnam Kim, Department of Computer Science, University of Maryland, and College Park, MD 20742, USA.

[4] M. Pantic, L.J.M. Rothkrantz, "Automatic Analysis of Facial Expressions": The State of the Art, IEEE *Trans. Pattern Analysis and Machine Intelligence*, Vol. 22, No. 12, pp. 1424-1445, 2000

[5] P. Viola and M. Jones. Robust real-time object detection. Technical Report 2001/01, Compaq Cambridge Research Lab, 2001.

[6] I. Essa and A. Pentland. Coding, analysis, Interpretation and recognition of facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):757–763, 1997.

[7] M. Turk and A. Pentland. "Eigenfaces for recognition". *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[8] G. Littlewort, I. Fasel, M. Stewart Bartlett, and J. R. Movellan. Fully automatic coding of basic expressions from video. Technical Report 2002.03, UCSD INC MPLab, 2002.

[9] M. Lades, J. C. Vorbr¨uggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. W¨urtz, and W. Konen. Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*, 42:300–311, 1993.

[10] J. F. Cohn, A. J. Zlochower, J. J. Lien, and T. Kanade. Feature-point tracking by optical flow discriminates subtle differences in facial expression. In *Proceedings International Conference on Automatic Face and Gesture Recognition*, pages 396–401, 1998.

[11] P. Ekman. Basic emotions. In T. Dalgleish and T. Power, editors, *The Handbook of Cognition and Emotion*. John Wiley & Sons, Ltd., 1999.

[12] P. Ekman and W. Friesen. "*Facial Action Coding System" (FACS): Manual*. Consulting Psychologists Press, Palo Alto, CA, USA, 1978.

[13] Y.-L. Tian, T. Kanade, and J. F. Cohn. Recognizing action units for Facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2):97–115, 2001.

[14] M. N. Dailey, G. W. Cottrell, and R. Adolphs. A six-unit network is all you need to discover happiness. In *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society*, Mahwah, NJ, USA, 2000. Erlbaum.

[15] Nagarajan R., S. Yaacob, P. Pandiyan, M. Karthigayan, M. Khalid & S. H. Amin. (2006). Real Time Marking Inspection Scheme for Semiconductor Industries. *International Journal of Advance Manufacturing Technology (IJAMT)*, ISSN No (online): 1433-3015.

[16] Karthigayan. M, Mohammed Rizon, Sazali Yaacob & R. Nagarajan,. (2006). An Edge Detection and Feature Extraction Method Suitable for Face Emotion Detection under Uneven Lighting, *The 2nd Regional Conference on Artificial Life and Robotics (AROB'06)*, July 14 - July 15, Hatyai, Thailand.

[17] Karthigayan M, M.Rizon, S.Yaacob and R. Nagarajan. (2007). On the Application of Lip features in Classifying the Human Emotions, *International Symposium on Artificial life and Robotics,* Japan.