

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/221786983>

Genetic Algorithm and Neural Network for Face Emotion Recognition

Chapter · May 2008

DOI: 10.5772/6186 · Source: InTech

CITATIONS

12

READS

1,908

4 authors, including:



Karthigayan Muthukaruppan

Manipal International University

45 PUBLICATIONS 140 CITATIONS

[SEE PROFILE](#)



Raghul Nagarajan

Madras Christian College

141 PUBLICATIONS 1,801 CITATIONS

[SEE PROFILE](#)



Sazali Yaacob

University of Kuala Lumpur

366 PUBLICATIONS 4,244 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Adaptive Fatigue and Alertness Model Using EEG Frequency Spindles Based on Neural Net-Fuzzy Systems [View project](#)



Bio-inspired mechanism for improving network systems perofromances [View project](#)

Genetic Algorithm and Neural Network for Face Emotion Recognition

M. Karthigayan, M. Rizon, R. Nagarajan and Sazali Yaacob,
*School of Mechatronics Engineering,
 Universiti Malaysia Perlis (UNIMAP),
 02600 Jejawi, Perlis,
 Malaysia*

1. Introduction

Human being possesses an ability of communication through facial emotions in day to day interactions with others. Some study in perceiving facial emotions has fascinated the human computer interaction environments. In recent years, there has been a growing interest in improving all aspects of interaction between humans and computers especially in the area of human emotion recognition by observing facial expressions. The universally accepted categories of emotion, as applied in human computer interaction are: Sad, Anger, Joy, Fear, Disgust (or Dislike) and Surprise. Ekman and Friesen developed the most comprehensive system for synthesizing facial expression based on what they called as action units (Li, 2001). In the early 1990's the engineering community started to use these results to construct automatic methods of recognizing emotion from facial expression in still and video images (Sebe, 2002). Double structured neural network has been applied in the methods of face detection and emotion extraction. In this, two methods are proposed and carried out; they are lip detection neural network and skin distinction neural network (Takimoto et al., 2003). Facial action coding (Panti & Patras, 2006) is given to every facial points. For example, code 23 is given for lip funnel, code 4 for eye brow lower, code 10 for chin raise etc. The cods are grouped for a specific facial emotion. In order to determine the category of emotion, a set of 15 facial points in a face-profile sequence has been recommended. The work performs both automatic segmentation of an input video image of facial expressions and recognition of 27 Action Units (AUs) fitted to facial points. A recognition rate of 87% has been reported. The motion signatures (Anderson & Peter, 2006) are derived and classified using support vector machines (SVM) as either non-expressive (neutral) or as one of the six basic emotions. The completed system demonstrates in two simple but effective computing applications that respond in real-time to the facial expressions of the user. The method uses edge counting and image correlation optical flow techniques to calculate the local motion vectors of facial feature (Liyanage & Suen, 2003). Cauchy Naïve Bayes classifier is introduced in classifying the face emotion. The person-dependent and person-independent experiments have demonstrated that the Cauchy distribution assumption typically provides better results than those of the Gaussian distribution assumption (Sebe, 2002).

The current literature on emotion detection through facial images indicates the requirement of two desired directions. One on the image processing techniques highly relevant for identifying facial features under uneven lighting and the other is on interpreting the face emotion through the processed facial features. Both of these problems are under taken in this paper. Lips and eyes are used as origins of extracting facial emotion features. The methods of image processing, filtering and edge detection that are suitable for feature extraction are proposed first. Then this processed image is utilized to identify certain optimum parameters through Genetic Algorithm (GA). These parameters are employed in the interpretation of emotion characteristics. A set of new fitness functions are also suggested in extracting the lip parameters through GA. Since the emotion detection algorithm can be made as an expert system through continuous processing of available data, the suggested method of emotion interpretation is considered as suitable for a personalized face. A South East Asian (SEA) subject is considered in this work to illustrate the process (Figure 1). Figure 2 shows the outline of the process flow of estimating the emotion through facial images.



Fig.1. The Angry Emotion of South East Asian

2. Face image processing

As the first step in image processing, the region of interest (ROI) of a lip and an eye have been selected in the acquired images. The ROI image is converted into grayscale image (0-255). Before obtaining the filtered grayscale image, a histogram equalization method has been applied. Histogram equalization (Rafael et al., 2003) improves the contrast in the grayscale and its goal is to obtain an uniform histogram. The histogram equalization method also helps the image to reorganize the intensity distributions. New intensities are not introduced into the image. Existing intensity values will be mapped to new values but the actual number of intensity pixels in the resulting image will be equal or less than the original number. In the image sequence, the histogram equalized image is filtered using average and median filters in order to make the image smoother. Finally, Sobel edge detection method is applied to the filtered image with a good level of success. However, due to the intensity variations of light exposed on the face, the segmentation process is not satisfactory. In the edge detected image of the whole face, the eyes are properly segmented where as the lips segmentations are poor. Hence, the histogram equalized image is split into

of lip ROI and eye ROI regions and then the regions are cropped from the full image. This has solved the problem of light intensity variations. Figure 3 and 4 show the edge detected eye and lips regions derived from their respective ROI areas.

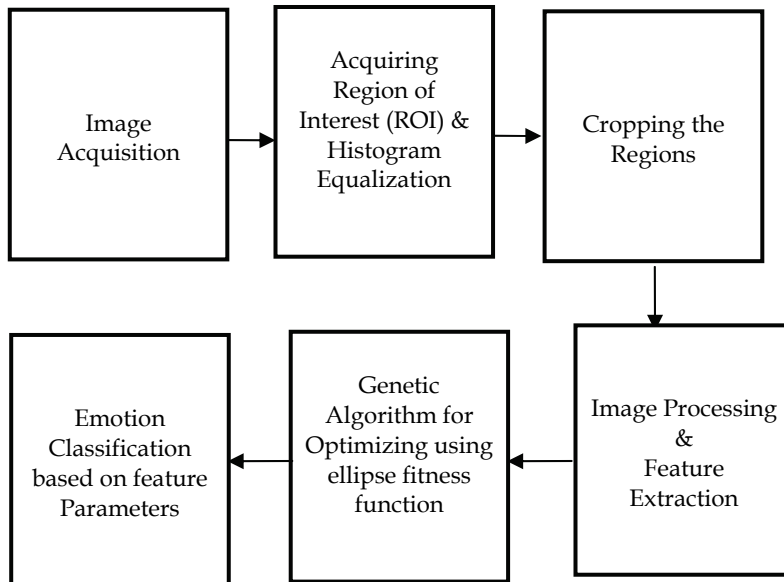


Fig. 2. Process Flow of Image Processing



Fig. 3. Sobel Edge Detected Eyes Region



Fig. 4. Sobel Edge Detected Lip Region

3. 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).

3.1 Projection profile

This feature extraction method is associated with the row-sum and column-sum of white pixels of edge identified image (Karthigayan et al., 2006). The pattern of row-sum (P_h) along the column and the pattern of column-sum (P_v) along the row of white pixels are defined as the feature of each region. These patterns are known as projection profiles. Let $S(m,n)$ represent a binary image of m rows and n columns. Then, the vertical profile is defined as the sum of white pixels of each column perpendicular to the x -axis; this is represented by the vector P_v of size n as defined by

$$p_{vj} = \sum_{i=1}^m s(i, j) \quad , \quad j = 1, 2, 3, \dots, n \quad (1)$$

The horizontal profile is the sum of white pixels of each row perpendicular to the y -axis; this is represented by the vector P_h of size m , where

$$p_{hi} = \sum_{j=1}^n s(i, j) \quad , \quad i = 1, 2, 3, \dots, m \quad (2)$$

3.2 Moments

The moments have been widely used in pattern recognition (Karthigayan et al., 2006). Several desirable properties that can be derived from moments are also applicable to face emotion analysis. Central moments processing time is faster than Zernike moments and moments invariant. Central moments of binary image for each column of the image orders can be obtained. The image orders can be of 2 or 3. In the order 1, moment values are zeros. On the other hand, orders more than 3 produce smaller and smaller moment values that will not increase the effectiveness of feature extraction.

Let $f(x,y)$ be an image. Then, the 2D continuous function of the moment of order $(p+q)$, M_{pq} , is defined as

$$M_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy \quad (3)$$

The central moment, μ_{pq} , of $f(x,y)$ is defined as.

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})^p (y - \bar{y})^q f(x, y) dx dy \quad (4)$$

where $\bar{x} = \frac{M_{10}}{M_{00}}$ and $\bar{y} = \frac{M_{01}}{M_{00}}$.

If $f(x,y)$ is a digital image then equation (4) becomes

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y) \quad (5)$$

where p and q are of nonnegative integer values. The moment values can be considered as extracted features.

3.3 Contour profile

This is one of the techniques used for object identification in the field of pattern recognition. The outer vertical and horizontal edge detected image black pixels in white background are counted. This count is named as the contour profile and can be used as the features of the lip and eye regions.

The performance of each of the above described feature extracting methods are compared with an objective of selecting one for our approach. The projection profile is found to perform well with regards to the processing time and is adopted here. The projection profile has also been found to have performed well in varied aspects in the earlier works (Karthigayan et al., 2006; Nagarajan et al., 2006; Karthigayan et al., 2007).

4. Face emotion recognition using genetic algorithm

In the early 1970s, John Holland, one of the founders of evolutionary computations, introduced the concept of genetic algorithm (GA) (Negnevitsky, 2002). The GA is a particular class of evolutionary algorithms. This is a heuristic approach used to find approximate solutions to solve problems through application of the principles from evolutionary biology. GA adopts biologically-derived techniques such as inheritance, mutation, natural selection, and recombination (or crossover). A population containing a number of trial solutions each of which is evaluated (to yield fitness) and a new generation is created from the better of them. The process is continued through a number of generations with the aim that the population evolves to contain an acceptable solution. GA is well known for optimization of nonlinear functions. It offers the best optimized value for any fitness function suitably selected for particular problems.

GA has been applied in varieties of applications which include image processing, control systems, aircraft design, robot trajectory generation, multiple fault diagnosis, traveling salesman, sequence scheduling and quality control wherein solutions to nonlinear optimization are required (Neo, 1997). Some aspects of vision system and image processing methodologies have been discussed towards approximating the face as a best ellipse using GA. In the feature extraction stage, the GA is applied to extract the facial features such as the eyes, nose and mouth, in a set of predefined sub regions. Some simulations have been carried out (Gary & Nithianandan, 2002). A method that extracts region of eyes out of facial image by genetic algorithm has also been suggested recently (Tani et al., 2001).

The human eye shape is more towards an ellipse (we call this as a regular ellipse). The edge detected eye image can be considered as an ellipse with variations. The minor axis is a feature of the eye that varies for each emotion. The major axis of the eye is more or less fixed for a particular person in varied emotions. The whitened area within the edge detected eye

image for one of the emotions of SEA is shown in Figure 5. The ellipse is parameterized by its minor and major axes, respectively, as "2a" (fixed) and "2b" (to be computed). This is shown in Figure 6 and is described by the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \quad (6)$$

The human lip shape is more towards a combination of two ellipses and we call this as an irregular ellipse. The word 'irregular' means that the ellipse has two different minor axes wherein the major axes remains the same. The edge detected lip image is considered as an irregular ellipse. Lengths of minor axes of the lip feature for each emotion are computed. Figure 7 shows the whitened area of edge detected lip image for a particular emotion of South East Asian. The major axis is "2a" (considered as fixed) and two minor axes are "2b1" and "2b2" (to be computed). This is shown in Figure 8 and is described by Equation (6) with b1 and b2 suitably substituted for top and bottom portions respectively.

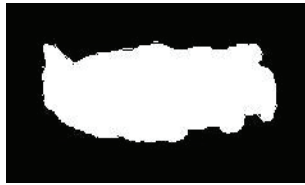


Fig. 5 Edge detected and whitened eye image

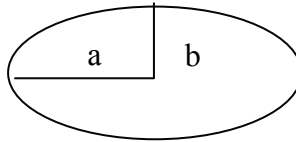


Fig. 6 The Regular Ellipse



Fig. 7. Edge detected and whitened lip image

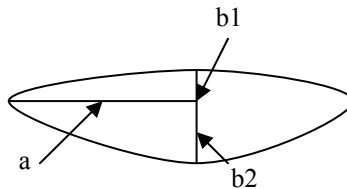


Fig 8. The Irregular Ellipse

4.1 Algorithm

GA is an iterative process (Negnevitsky, 2002). Each iteration is called generation. A chromosome of length of 6 bits and a population of 20 are chosen in our work. The selected chromosome is an approximate solution. Other selected parameters are listed in Table 1. The GA process is described in the following steps.

Step 1. Represent the problem variable domain as chromosome of a fixed length and population, with suitable cross over probability and mutation probability

Step 2. Define a fitness function to measure the performance, or fitness of an individual chromosome in the problem domain

Step 3. Randomly generate an initial population of chromosomes.

Step 4. Calculate the fitness of each individual chromosome.

Step 5. Select a pair of chromosomes for mating from the current population. Parent chromosomes are selected with a probability related to their fitness. Highly fit chromosomes have a higher probability of being selected for mating compared to less fit chromosomes.

Step 6. Create a pair of offspring chromosomes by applying the genetic operators - crossover and mutation

Step 7. Place the created offspring chromosomes in the new population

Step 8. Repeat from step 5 until the size of new chromosome population becomes equal to the size of the initial population

Step 9. Replace the initial chromosome population with the new population

Step 10. Go to step 4, and repeat the process until the termination criterion is satisfied.

4.2 Fitness function

A fitness function is a particular type of objective function that quantifies the optimality of a solution, a chromosome, so that the chromosome is ranked against all the other chromosomes. The fitness functions are derived from Equation(6). Equations(7) and (8) are fitness functions respectively for "b1" and "b2" to obtain optimum lip features. Equations(9) is the fitness function for "b" to obtain the optimum eye feature.

$$f(x) = \left(\sum_i^m \sum_j^n \text{col}(j) - 2\sqrt{X_1^2 \left(1 - \frac{\text{row}(i)^2}{a^2}\right)} \right)^2 \quad (7)$$

$\text{if } X_1 \geq 0$

$$\overline{f(x)} = \left(\sum_i^m \sum_j^n \text{col}(j) - 2\sqrt{X_2^2 \left(1 - \frac{\text{row}(i)^2}{a^2}\right)} \right)^2 \quad (8)$$

$\text{if } X_2 \leq 0$

$$f(x) = \left(\sum_i^m \sum_j^n \text{col}(j) - 2\sqrt{X^2 \left(1 - \frac{\text{row}(i)^2}{a^2}\right)} \right)^2 \quad (9)$$

In Equation (7) to (9), col(j) is the sum of white pixels occupied by jth column and row(i) is the sum of white pixels in ith row.

The lip and eye features have been given as inputs to the GA to find the optimized values of b_1 & b_2 and b . Considering the parameters indicated in Table 1, the variance of the distribution of the Gaussian Mutation can be controlled with two parameters, the scale and the shrink. The scale parameter determines the variance at the first generation. The shrink parameter controls how variance shrinks as generations go by. These are selected as 1.0 each. Scattered crossover creates a random binary vector. It then selects the genes where the vector is a 1 from the first parent and the genes where the vector is a 0 from the second parent and combines the genes to form the child. The optimization has been carried out for 5 times for each emotion. The process of GA is found to offer a set of optimized minor axis values X_1 , X_2 and X through fitness equations, Equations (7), (8) and (9). Table 2 indicates the manually measured values, b_1 , b_2 and b , and the corresponding optimized values, X_1 , X_2 and X . The emotion, based on X_1 , X_2 and X can now be estimated. The experiment results show that values of X_1 , X_2 and X are different for each emotion there by distinctions are possible.

Generation	250
Population size	20
Fitness scaling	Rank
Selection Function	Roulette
Mutation	Gaussian
Crossover	Scattered
Stall generation	50
Stall time	20

Table 1 Parameter Settings for GA Processing

Emotions	Manually Computed Mean Value (in pixels)			Optimized Mean Value by GA (in pixels)		
	b_1	b_2	b	X_1	X_2	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

Table 2 Optimized Value of the three Features

5. Emotion classification using neural network

Recently, there has been a high level of interest in applying artificial neural network for solving many problems (Negnevitsky, 2002; Pantic M. & Leon, 2000). The application of neural network gives easier solution to complex problems such as in determining the facial expression (Nagarajan, 2006). 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 2, X_1 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 X_1 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.

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 9. 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 $(3 \times 20 \times 3)$) and the other model with a structure of $(3 \times 20 \times 7)$. The output of $(3 \times 20 \times 3)$ is a 3-bit binary word indicating the seven emotional states. The output ($O_i, i=1,2,..., 7$) of $(3 \times 20 \times 7)$ 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 3. 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.

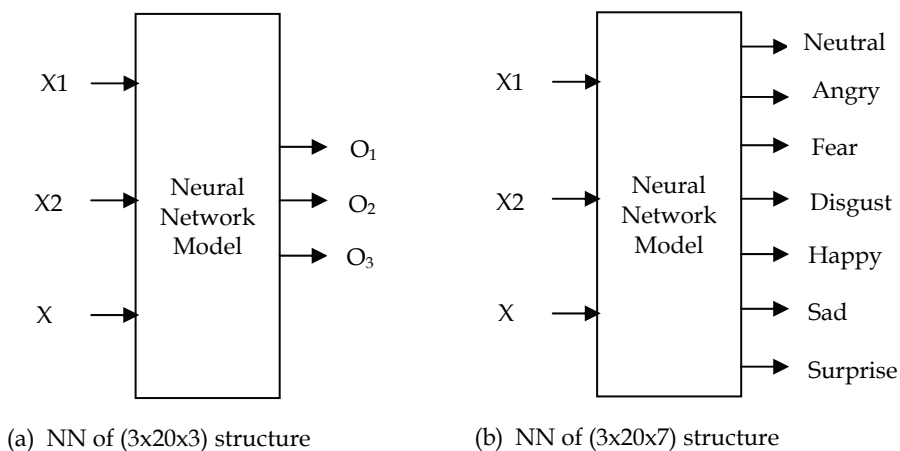


Fig. 9 Neural Network Structures

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

Table 3. Details of Neural Network Classification of Emotion

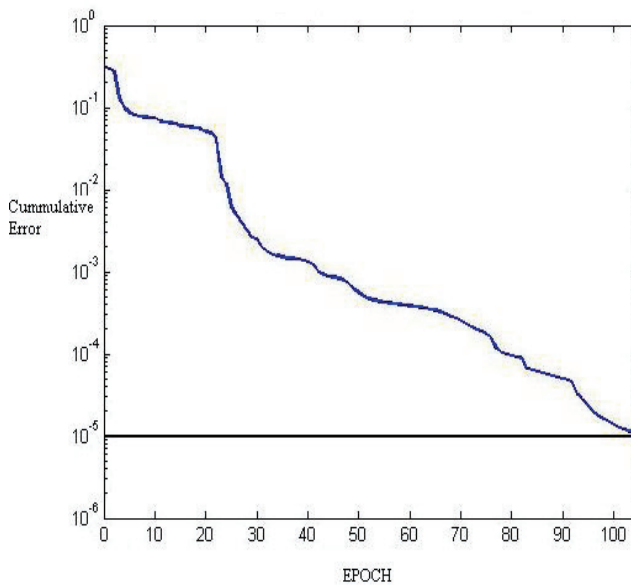


Fig. 10 Error vs Epoch Characteristic

6. Results and conclusion

In this study on a South East Asian face, the classification of six emotions and one neutral has been considered. The average and median filters are applied to smoothen the image. The Sobel edge detection is found to perform well since it offers a better segmentation even in non-structural light intensities. The eye and lip regions are used for the study on emotions. The GA is then applied to get the optimized values of the minor axes, X_1 and X_2 , of an irregular ellipse corresponding to the lips and the minor axis, X , of a regular ellipse related to eye, by using a set of new fitness functions. The ranges of optimized values of emotion

obtained from GA are found to overlap with each other resulting in an unacceptable classification. In order to overcome this problem, a NN is implemented to offer better classification. On an average of 10 trials of testing, the suggested NN processing has achieved about 85.13% and 83.57% of success rate for structure 3x20x7 and for structure 3x20x3 of NN models respectively. The successful classification even goes to the maximum of about 91.42% in the NN model of 3x20x7 structure. Even though the suggested methodology of face emotion detection and classification is general, the derived results are suitable only to a personalized South East Asian face. The software package can be developed as an expert emotion classification system for a personalized face. The applications of this emotion classification system are many such as from identifying criminals through a police enquiry to helping bedridden disabled dumb patients.

7. References

- Anderson K. & Peter W. McOwan (2006, February). A Real-Time Automated System for the Recognition of Human Facial Expressions. *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, 36(1), 96-105.
- Gary G. Yen & N. Nithianandan. (2002, May). Facial Feature Extraction Using Genetic Algorithm. *Proceedings of Congress on Evolutionary computation*, 2, 1895-1900.
- 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.
- 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.
- Liyanage C De Silva & Suen Chun Hui. (2003). Real-time Facial Feature Extraction and Emotion Recognition, *Proceedings of Fourth Pacific RIM Conference on Multimedia*, vol. 3, pp. 1310-1314, 15-18 Dec 2003.
- Li H. (2001, May 2-4). Computer Recognition of Human Emotion. *Proceedings of International Symposium on Intelligent Multimedia, Video and Speech Processing*, Hong Kong, 490-493.
- 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.
- Negnevitsky M.. (2002). *Artificial Intelligence*. Addison Wesley, Pearson Education Limited, England.
- Neo. (1997). http://neo.lcc.uma.es/TutorialEA/semEC/cap03/cap_3.html
- Panti M. & I. Patras. (2006, April). Dynamics of Facial Expression: Recognition of Facial Actions and Their Temporal Segments from Face Profile Image Sequences. *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, 36(2), 433-449.
- Pantic M. & Leon J. M Rothkrantz (2000, December). Automatic Analysis of Facial Expressions: The State of the Art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12), 1424-1445.
- Rafael C Gonzalez, Richard E. Woods & Steven L Eddins. (2003). *Digital Image Processing using MATLAB*, Prentice Hall.

- Sebe N., M. S. Iew, I. Cohen, Y. Sun, T. Gevers & T. S. Huang. (2004, May 17-19) .Authentic Facial Expression Analysis. *Proc. of the sixth International Conference on Automatic Face and Gesture Recognition (FGR'04)*, Seoul, Korea, 517-522.
- Tani H., K. Terada, S. Oe & J. Yamaguchi. (2001). Detecting of One's Eye from facial Image by Using Genetic Algorithm. *The 27th Annual Conference of the IEEE Industrial Electronics Society*, Colorado, USA 1937-1940.
- Takimoto H., Y. Mitsukura, M. Fukumi & N. Akamatsu.(2003, July 20-24). Face Detection and Emotional Extraction System Using Double Structure Neural Network. *Proc. of the International Joint Conference on Neural Networks*, Orlando, USA, 2, 1253-1257.