

## Simple Approach for Face Expression Recognition Using Dual Classifier for Enhanced Throughput

Desai Ghata Sanjeevkumar<sup>1</sup>, Pankaj Kumar Gautam<sup>2</sup>

<sup>1</sup>CSE Department, Parul Institute of Technology, Limda, Vadodara, India.

<sup>2</sup>CSE Department, Parul Institute of Technology, Limda, Vadodara, India.

**Abstract** — This paper presents a facial expression recognition system using dual classifier approach. The face is detected using viola jones algorithm. The Gabor filters are used for feature extraction. The dimensionality of the feature vector is reduced by the Principal Component Analysis (PCA) to remove redundant data that leads to unnecessary computation cost. The Support Vector Machine (SVM) and Bayesian Classifiers are used sequentially for expression classification. The performance of the proposed method is tested on public and largely used Cohn-Kanade database. The experiments show that proposed method gives promising results.

**KEYWORDS**- CLASSIFICATION, GABOR, SVM, ADABOOST, BAYESIAN CLASSIFIER

### I. INTRODUCTION

A recent challenge in designing computerized environments is to keep the human user at the core of the system. To have truly affective human-computer intelligent interaction systems (HCIs), the computer needs to be capable of interacting with the user in a natural way (the way in which two people interacts with each other). To recognize an emotional state, HCIs should interpret non-verbal behaviour like: voice, body gesture and facial expressions. Among three, facial expression is the most natural means of communicating human emotions, intentions and opinions to each other. [1] showed in their research work, that 55% of the emotional information is conveyed by facial expression alone. Remaining voice tone and spoken words conveys 38% and 7% of the information respectively. [2] did a psychological research on facial expression and they concluded that there are six basic facial expressions which are universal. The six basic expressions include happiness, sadness, disgust, anger, surprise and fear. Facial expression recognition system can broadly be categorized into three basic parts: Face detection, features extraction and classification.

The organization of the paper is as follows: section II presents the structure of the proposed method, section III demonstrates the detailed description about the method, section IV gives the experimental results of the proposed tested on the public and largely used databases CK and CK+ and VI conclusions are drawn.

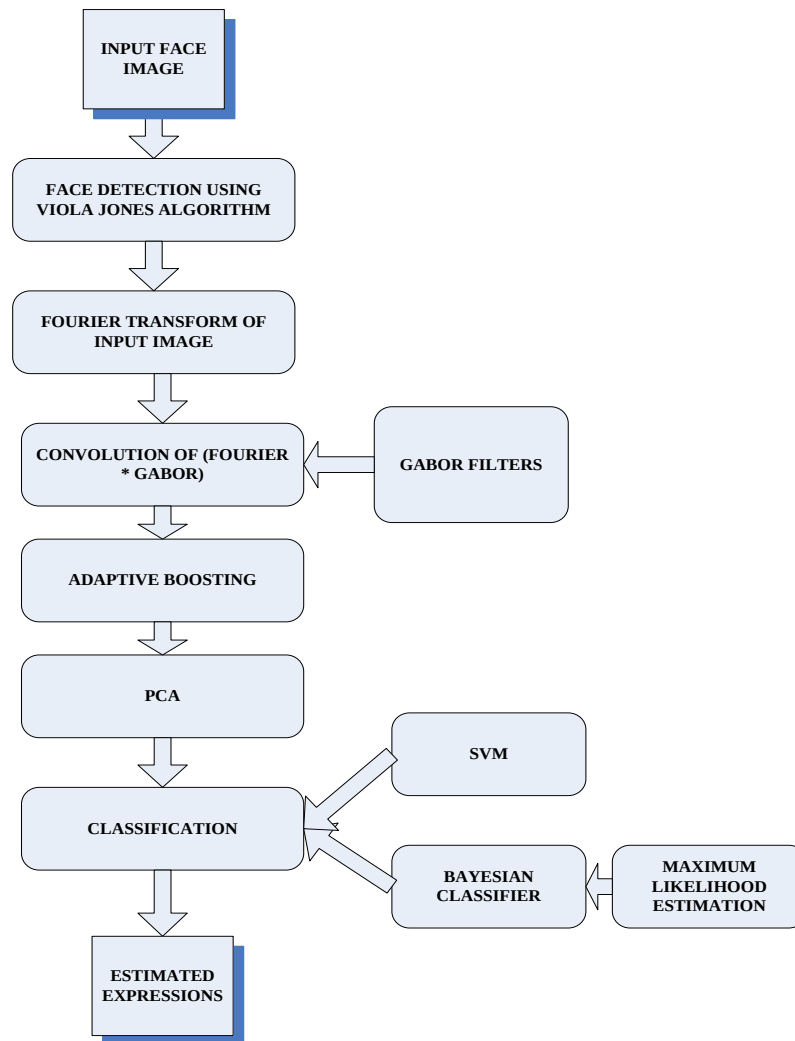
### II. SYSTEM STRUCTURE

Figure 1 shows the system structure of the proposed method. The viola jones algorithm is used for detecting face from the input image. The Fourier transform of the input image is done. The feature vector is extracted from the transformed face image using Gabor filters. From the large number of features the deserving features are selected using AdaBoost algorithm. The dimensionality of feature vector is reduced using PCA to remove redundant data that leads to unnecessary computation costs. Then the features are fed to the SVM classifier. The Bayesian classifier with the maximum likelihood estimation is used for the cross validation purpose. The dual classifier gives the estimated facial expressions.

### III. PROPOSED METHOD

#### Step 1: Face Detection Using Viola Jones:

First Component is face detection component. It detects the face from the given input images. The reason behind using Viola Jones algorithm have several advantages like extremely fast feature computation, efficient feature selection, scale and location invariant detector, instead of scaling the image itself, we scale the features etc. The face is detected by matching the Haar-like features in the database with these features in testing image. A Haar-like feature is the difference between the sums of the pixels in different rectangular regions which can be seen as the black and white regions in Figure 2. There are three different types of this feature, as shown: 2-region features, 3-region features, and 4-region features. In 2-region features, the difference is calculated between two regions. In 3-region features, the difference is calculated by the subtraction between the sum of left and right regions and the center region. In 4-region features, the sums of two region pairs in the diagonal are calculated and the Haar-like feature is the difference of these two sums. These Haar-like features represent the characteristics of the face by interpreting the face has different light and dark areas. For example, the eye region is darker than the nose region.



**Figure 1: Face Recognition Processing Flow**



**Figure 2: Face detection by Haar-like features [3]**

#### **Step 2: Fourier Transform of Input Image [17]**

The Detected face image goes through the Fourier transform. The Fourier transform is a representation of an image as a sum of complex exponentials of varying magnitudes, frequencies, and phases. The Fourier transform plays a critical role in a broad range of image processing applications, including enhancement, analysis, restoration, and compression.

If  $f(m,n)$  is a function of two discrete spatial variables  $m$  and  $n$ , then the two dimensional Fourier transformation of  $f(m,n)$  is defined by the relationship

$$f(\omega_1, \omega_2) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} f(m, n) e^{-j\omega_1 m} e^{-j\omega_2 n} \quad (3.2.1)$$

The variables  $\omega_1$  and  $\omega_2$  are frequency variables; their units are radians per sample.  $F(\omega_1, \omega_2)$  is often called the *frequency-domain* representation of  $f(m, n)$ .  $F(\omega_1, \omega_2)$  is a complex-valued function that is periodic both in  $\omega_1$  and  $\omega_2$ , with period  $2\pi$ . Because of the periodicity, usually only the range  $-\pi \leq \omega_1, \omega_2 \leq \pi$  is displayed. Note that  $F(0, 0)$  is the sum of all the values of  $f(m, n)$ . For this reason,  $F(0, 0)$  is often called the *constant component* or *DC component* of the Fourier transform.

### Step 3: Convolution

We know that images are generally complex and the magnitude is generally used for visualization. The convolution of Gabor filters is done by computing the Fourier transform of the image and multiplying with the Fourier transform of the Gabor filter in the frequency domain. The Gabor wavelet representation of images allows description of spatial frequency structure in the image while preserving information about spatial relations. Let  $r(x, y)$  be the original image  $f(x, y)$  convolved with a 2D Gabor function  $g(x, y)$

$$r(x, y) = f(x, y) * g(x, y)$$

Where  $*$  denotes the convolution operator and with  $g(x, y)$  being a 2D Gabor filter defined as

$$g_{\lambda, \phi, \sigma, \gamma}(x, y) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos(2\pi \frac{x'}{\lambda} + \phi) \quad (3.2.2)$$

Where  $x' = x \cos \theta + y \sin \theta$  and  $y' = -x \sin \theta + y \cos \theta$

The parameters used in Eq. (2.2.2.1) are defined as follows. Wavelength ( $\lambda$ ); this is the wavelength of the cosine factor of the Gabor filter kernel, Orientation(s) ( $\theta$ ); it specifies the orientation of the normal to the parallel stripes of a Gabor function. Its value is specified in degrees. Valid values are real numbers between 0 and 360. The Phase offset(s) ( $\phi$ ) in the argument of the cosine factor of the Gabor function is specified in degrees. Valid values are real numbers between 180 and 180. The Aspect ratio ( $\gamma$ ) called more precisely the spatial aspect ratio, specifies the ellipticity of the support of the Gabor function.

### Step 4: Adaptive Boosting

Due to the large size of the Gabor wavelets, it is not practically possible to use all the wavelets as input to our classifier for fear of misclassification and possible system crash [4]. The AdaBoost feature reduction algorithms have special speed advantage in increasing classification process. Thus we formulated an AdaBoost-based algorithm to select a few deserving portions of the wavelets.

Assuming the extracted Gabor features are represented by a total of  $i \in (1, 2, \dots, N)$  appearance features. Then the image  $I$  is represented as  $\Phi_i = \{(x_n, y_n)\}_{n=1}^N$  configured by the parameters  $z, \mu, v$ . The positive sets  $\Phi^{(+)}$  and the negative sets  $\Phi^{(-)}$  are denoted by  $\Phi^{(+)} = \{(x_n, y_n)\}_{n=1}^N \subset R^1 \times (\pm 1)$  and  $\Phi^{(-)} = \{(x_n, y_n)\}_{n=1}^N \subset R^1 \times (\pm 1)$  respectively, where  $x_n$  is the  $n$ th data sample containing  $J$  features, and  $y_n$  is its corresponding class label. To train the vectors  $\|G\|$ , which is denoted by  $\Phi_{(u, v, z)}$  over a distribution  $D$ , we simply determined the weights of all the feature vectors  $\{(x_n, y_n)\}_{n=1}^N = \Phi^{(+)} + \Phi^{(-)}$ . This gives us a threshold  $\lambda$  which indicated the decision hyperplane.  $\lambda$  is computed as:

$$\lambda = \frac{\sum_{\Phi_i \in \Phi^{(+)}} D(i, \Phi_{(\mu, v, z)})}{\|\sum_{\Phi_i \in \Phi^{(+)}} D(i, \Phi_{(\mu, v, z)})\|} + \frac{\sum_{\Phi_i \in \Phi^{(-)}} D(i, \Phi_{(\mu, v, z)})}{\|\sum_{\Phi_i \in \Phi^{(-)}} D(i, \Phi_{(\mu, v, z)})\|} \quad (3.2.3)$$

A sample is positive or client if it is located at the positive half of  $\lambda$  (which is the majority decision), otherwise it is a negative or an imposter. The status is reversed if the minority of the positive instances is rather located in the positive half space. Let  $c$  be denoted by clients and  $p$  be the imposters. For a given training dataset containing both positive and negative samples, where each sample is  $(S_i, y_i); y_i \in \{\pm 1\}$  represents the corresponding class label, the feature selection algorithm is formulated as follows [4]:

- Initialize sample distribution  $D_0$  by weighting every training sample equally such that the initial weight  $w_{1,i} = 1/2c, 1/2p$  for  $y = 1$  and  $-1$ , respectively.
- For the iteration  $t = 1, 2, \dots, T$ , where  $T$  is the final iteration, do:
  - Normalize the weights,  $w_{t,i} \leftarrow \frac{w_{t-1,i}}{\sum_{i=1}^N w_{t-1,i}}$ , where  $w_t$  a probability distribution and  $N$  is the total number of features.
  - Train a weak classifier  $h_t$  for feature  $j$ , which uses a single feature. The training error  $\epsilon_t$  is estimated with respect to  $w_t$  such that:
 
$$\epsilon_t = \sum_i w_{t,i} |h_t(x_i) - y_i|^2$$
  - Select the hypothesis  $h_t^1$  with the most discriminating information, that is to say, the hypothesis with the least classification error  $\epsilon_t^1$  on the weighted samples.

- Compute the weight  $\omega_t$  that weights  $h_t^1$  by its classification performance as:
  - $\omega_t = \frac{1}{2} \ln[\frac{1}{\epsilon^1_t} - 1]$
- The weight distribution is then updated and normalized.
- The final feature selection hypothesis  $H(S)$  which is a function of the selected features is denoted by:

$$H(S) = \text{sgn}[\sum_{t=1}^T \omega_t h_t^1 S^1_t]$$

- The selected features represent samples of the facial deformation patterns of the expressive face.

#### Step 5: PCA

Principle Component Analysis is then applied to extracted features for dimensionality reduction. Let's consider an application where we have  $N$  images each with  $n$  pixels. We can write our entire data set as an  $N \times n$  data matrix  $D$ . Each row of  $D$  represents one image of our data set. For example we may have  $D$ . First step is to move the origin to mean of the data. Then subtract the mean image from each image of the data set to create the mean centred data matrix  $U$ . Compute the covariance matrix from the mean centred data matrix:

$$\Sigma = U^T U / (N - 1) \quad (3.2.4)$$

Solve for  $\phi$  and  $\Lambda$  that satisfy:

$$\Sigma = \phi \Lambda \phi^T \quad (3.2.5)$$

Where  $\phi$  is matrix of Eigen vectors and  $\Lambda$  is matrix of eigenvalues. If we normalise the eigenvectors, then the system of vectors  $\phi$  forms an orthonormal basis, that is to say:

$$\forall \phi_i, \phi_j \in \phi, \phi_i \cdot \phi_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad (3.2.6)$$

It is in effect an axis system in which we can represent our data in a compact form. Achieve size reduction by selecting  $m$  largest Eigen values of  $\Sigma$ . An  $n \times m$  matrix  $\Phi_{pca} = [\phi_1, \phi_2, \dots, \phi_m]$  performs the PCA projection. For any given image  $\mathbf{p}_x = [i_1, i_2, \dots, i_n]$  we can find a corresponding point in the PCA space by computing

$$\mathbf{p}_\phi = (\mathbf{p}_x - \mu_x) \cdot \Phi_{pca} \quad (3.2.7)$$

The  $m$ -dimension vector  $\mathbf{p}_\phi$  is all we need to represent the image.

#### Step 6: Classification:

The main factor of the proposed algorithm is the dual classifier used. It is the combination of Support Vector Machine (SVM) and Bayesian Classifier.

Bayesian Classifier:

The features outputs from PCA are fed to the Bayesian classifier. We use baseline classifier based on data reduction and classification using a Bayesian classifier. We fit multivariate normal distributions to all the classes using Maximum Likelihood Estimation and classify based on posteriors.

Bayesian classifier a simple probabilistic classifier based on applying Bayes' theorem.

$$p(h|d) = \frac{p(d|h)p(h)}{p(d)} \quad (3.2.8)$$

Where,  $p(h)$  = prior belief (probability of hypothesis before seeing any data),  $P(d|h)$  = likelihood (probability of data if the hypothesis  $h$  is true),  $P(d)$  = data evidence (marginal probability of the data),  $P(h|d)$  = posterior (probability of hypothesis after having seen the data)

SVM Classifier:

An SVM is a method of supervised learning for classification by using a hyper plane to divide the space into two regions, each region classifying one type of element [3]. For training, from the negative training samples and positive training samples, a margin separator is defined by mathematical functions. Using this margin separator, the testing sample is considered as positive case or negative case. From the feature points the SVM can classify the expressions according to given facial database. Finally the resulting expressions are displayed. Once the training is completed the classifier can now classify the face expression for the given input image of face.

#### IV. Experiment Results

We have used the Cohn-Kanade Database for the training of classifier and testing purpose. The CK and CK+ database contains 593 sequences across 123 subjects which are FACS coded at the peak frame. All sequences are from the neutral face to the peak expression. This DB contains only posed expressions. The subjects were instructed to perform a series of 23 facial displays that included single action units and action unit combinations. Each begins from a neutral or nearly neutral face. For each, an experimenter described and modelled the target display. Six were based on descriptions of prototypic emotions (i.e., joy, surprise, anger, fear, disgust, and sadness). These six tasks and mouth opening in the absence of other action units were annotated by certified FACS coders. The figure 3 contains the results of classified expressions using the dual classifier, which is the combination of the Bayesian classifier and the Support Vector Machine classifier. In the following images the “Actual Emotion” shows the actual emotion of the image and the “Predicted emotion” shows the emotion recognized by our proposed method.

Figure 3(a) shows the result for input image that is indicated by “Figure 1” whose actual emotion is “Surprise” and our method predicted it correctly that is indicated by “Predicted emotion”. In Figure 3(b) the same is done for the input image of the expressions “sad”.

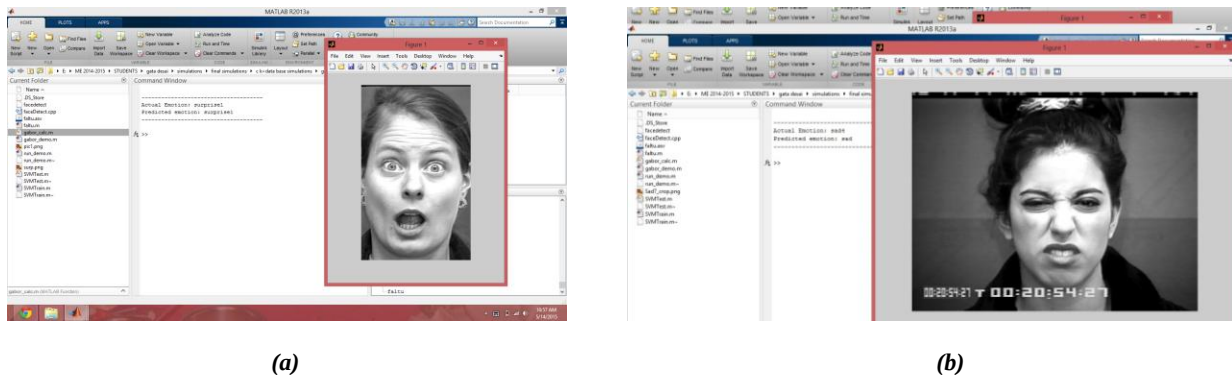


Figure 3: The Face Expression Detection using proposed method for input expressions:  
(a) Surprise (b) Sad

**Confusion Matrix:** A confusion matrix is a specific table layout that allows visualization of the performance of an algorithm. Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class. The name stems from the fact that it makes it easy to see if the system is confusing two classes.

Here we are considering the six different expressions: happy, sad, disgust, fear, angry and surprised. The Confusion matrix is shown in the Table 1.

Table 1: Confusion Matrix

|          |          | Prediction |          |
|----------|----------|------------|----------|
|          |          | Negative   | Positive |
| Classify | Negative | $T_N$      | $F_N$    |
|          | Positive | $F_P$      | $T_P$    |

In Above matrix the  $T_N$  indicates True Negative that is the test result indicates that result is negative and it is actually negative.  $F_N$  indicates false negative that is an error in which a test results indicates that result is negative, when in reality it is positive. The  $F_P$  indicates false positive that is an error in which a test indicates that result is positive but in reality it is negative.  $T_P$  indicates True Positive that is the test result indicates that result is positive and it is actually positive. We have applied our method on 50 different images of the standard CK and CK+ database. From the experiment we can derive the table 2.

Table 2: The Classification result on 50 images

| Classification Value | For 50 images |
|----------------------|---------------|
| $T_N$                | 24            |
| $F_P$                | 2             |
| $F_N$                | 4             |
| $T_P$                | 20            |

**Classification Accuracy:** Classification accuracy gives the rate of face expression recognition by the system. It gives the percentage of correctly detected images. The equation for calculating the classification accuracy is:

$$\text{Classification Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \quad (5.2.1)$$

From table 2 and equation 5.2.1 we can calculate the classification accuracy. The classification accuracy of our proposed method is 88%.

## V. CONCLUSION

Humans recognize facial expressions virtually without effort or delay, but reliable expression recognition by machine is still a challenge. For achieving optimal pre-processing, feature extraction or selection, and classification methods are required because of input data variability. The work that was shown above gives an accuracy of around 88.00 % on standard CK and CK+ Database. The accuracy is very much better than those conventional existing algorithms and obviously comes from the fact of multiple classifiers that have been used in the proposed system. The successful implementation of this system may be beneficial to many research scholars who are working in the field of Audio visual recognition. The work can be more focused as far as parameter like accuracy is concerned. Also some other feature based algorithms can be tried and results can be compared with our method and can be looked upon if it gives better results. The system can be made application specific and can be used into various applications which can be further used into fields: security, biometrics, etc. Other applications like Gestures based on emotions like feeling hot, cold and according to that gestures can be given to control the room temperature etc.

## REFERENCES

- [1] A. Mehrabian, Nonverbal communication, Aldine, 2007.
- [2] P. Ekman, W. V. Friesen, and J. C. Hager, "Facial action coding system," A Human Face, Salt Lake City, 2002.
- [3] Linh Tuan Dang, Eric W. Cooper, Katsuari Kamei, "Development of Facial Expression Recognition for Training Video Customer Service Representatives", 2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) July 6-11, 2014, Beijing, China.
- [4] Ebenezer Owusu, Yongzhao Zhan, Qi Rong Mao, "A neural-AdaBoost based facial expression recognition system", Expert Systems with Applications 41 (2014) 3383–3390.
- [5] Yuan Ren, "Facial Expression Recognition System," Waterloo, Ontario, Canada, 2008.
- [6] Ms. Aswathy.R, "A Literature review on Facial Expression Recognition Techniques", IOSR Journal of Computer Engineering (IOSR-JCE), Volume 11, Issue 1, May. - Jun. 2013, PP 61-6.
- [7] Anima Majumder, Laxmidhar Behera and Venkatesh K. Subramanian, "Local Binary Pattern Based Facial Expression Recognition Using Self-Organizing Map," International Joint Conference on Neural Networks (IJCNN), 2014.
- [8] Brian C. Lovell, Shaokang Chen, Ting Shan, "Real-time Face Detection and Classification for ICCTV," 2007.
- [9] G. Hemalatha, C.P. Sumathi, "A Study of Techniques for Facial Detection and Expression Classification", International Journal of Computer Science & Engineering Survey (IJCSES) Vol.5, No.2, April 2014.
- [10] Philipp Michel, Rana El Kaliouby, "Real Time Facial Expression Recognition in Video using Support Vector Machines", ICM'03, November 5–7, 2003, Vancouver, British Columbia, Canada.
- [11] Faten Bellakhdhari, Kais Loukil, Mohamed ABID, "Face recognition approach using Gabor Wavelets, PCA and SVM", IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 2, No 3, March 2013.
- [12] Vinay Bettadapura, "Face Expression Recognition and Analysis: The State of the Art", Georgia Institute of Technology.
- [13] Rania Salah El-Sayed, Prof. Dr. Ahmed El Kholy, Prof. Dr. Mohamed Youssri El-Nahas "Robust Facial Expression Recognition via Sparse Representation and Multiple Gabor filters", International Journal of Advanced Computer Science and Applications, Vol. 4, No.3, 2013.
- [14] Piyanuch Silapachote, Deepak R. Karupiah, and Allen R. Hanson, "Feature Selection Using Adaboost for Face Expression Recognition", University of Massachusetts Amherst Amherst, MA 01003, USA.
- [15] Adin Ramirez Rivera, Jorge A. Rojas Castillo, and Oksam Chae, "Recognition Of Face Expressions Using Local Principal Texture Pattern", IEEE 2012

### Websites:

- [16] <http://www.face-rec.org/>
- [17] [www.mathworks.com](http://www.mathworks.com)
- [18] <http://www.ece.northwestern.edu/>

### Books:

- [19] Rafael C. Gonzalez and Richard E. Woods, —Digital Image Processing, 3<sup>rd</sup> Edn, Pearson Publishing Company, 2009.
- [20] Rafael C. Gonzalez and Richard E. Woods and Steven L. Eddins, —Digital Image Processing using MATLAB, Pearson Publishing Company, 2002.

### Thesis:

- [21] Yang Yong "Facial Expression Recognition And Tracking Based On Distributed Locally Linear Embedding And Expression Motion Energy", National University Of Singapore, 2006.