

Face Localization and Recognition in Varied Expressions and Illumination

Hui-Yu Huang, Shih-Hang Hsu

Abstract—In this paper, we propose a robust scheme to work face alignment and recognition under various influences. For face representation, illumination influence and variable expressions are the important factors, especially the accuracy of facial localization and face recognition. In order to solve those of factors, we propose a robust approach to overcome these problems. This approach consists of two phases. One phase is preprocessed for face images by means of the proposed illumination normalization method. The location of facial features can fit more efficient and fast based on the proposed image blending. On the other hand, based on template matching, we further improve the active shape models (called as IASM) to locate the face shape more precise which can gain the recognized rate in the next phase. The other phase is to process feature extraction by using principal component analysis and face recognition by using support vector machine classifiers. The results show that this proposed method can obtain good facial localization and face recognition with varied illumination and local distortion.

Keywords—Gabor filter, improved active shape model (IASM), principal component analysis (PCA), face alignment, face recognition, support vector machine (SVM)

I. INTRODUCTION

OR the advance of technology development, human beings have inherent unique biological characteristics, such as fingerprint, face, or iris, which can gradually replace the card or key to control any items. This biologic information can easy to capture by means of lens device. However, the lighter factor extremely affected the information precision. In face recognition, illumination variation is usually existent in the captured face image. It is well known that the variations of illumination may change face appearance dramatically so that the variations between the images of the same face. Hence, there are many studies have been worked this effect on face recognition recently [1, 2, 3]. If these factors are considers, the face recognition rate van be improved and be more robust.

Owing the illumination factor, in this paper, we propose an approach to against light influence in face recognition based on template matching method of facial features for location precision. In face recognition, we combine the support vector machine to classify and simultaneously keep the good recognition rate.

The rest of the paper is organized as follows. Section 2 presents the proposed method. Experimental results and performance evaluation are presented in Sections 3. Finally, Section 4 concludes this paper.

Hui-Yu Huang is with Department of Computer Science and Information Engineering, National Formosa University, Yunlin 632, Taiwan (phone: +886-5-631-5590; fax: +886-5-633-0456; e-mail: anne.huang@ieee.org).

Shih-Hang Hsu was with Department of Computer Science and Information Engineering, National Formosa University, Yunlin 632, Taiwan.

II. PROPOSED SCHEME

Facial localization of landmark feature points often suffers from a variety of illumination and occlusion influences, in order to reduce these factors, a technology combining Gabor filter and hierarchical image blending is proposed in this paper. After performing this technology, we can obtain an image of illumination normalization and then be advantage of the localization and recognition of face. Figure 1 shows the diagram of system flowchart. In face recognition procedure, it consists of principal component analysis (PCA) for feature extraction and support vector machine (SVM) for the classification. Thus this approach can effectively achieve the location of feature points and recognition on the frontal face, so that illumination influence or local distortion will be reduced.

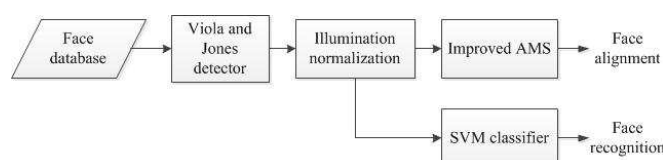


Fig. 1 System flowchart

A. Processing of Illumination Normalization

The problem of illumination variation is usually existent and an important factor in the study of face recognition. Recently, the diversification of light conditions under the theme of face recognition [2] indicated, feature extraction is imperfect in the case of light exposure or low light whether the Gaussian filter or histogram equalization method were used. Hence, in this paper, we will propose an effective method to solve this factor to improve the recognition rate and localization accuracy of face image.

The advantage of wavelet transform is to overcome the limitations of traditional Fourier transform, so that it can work in time domain and frequency domain to analyze the data. Gabor filters, which are generated form a wavelet expansion of the Gabor kernels [4], exhibit desirable attributes of spatial locality and frequency domains optimally. In this paper, we use the Gabor wavelet filter to modulate the orientations and frequencies in order to reduce the light influence. The Gabor wavelet filter is defined as

$$\varphi_{\mu,\nu}(z) = \frac{\|k_{\mu,\nu}\|^2}{\sigma^2} \left(\frac{\|k_{\mu,\nu}\|^2 \|z\|^2}{2\sigma^2} \right) (ik_{\mu,\nu}z) \exp\left(-\frac{\sigma^2}{2}\right) \quad (1)$$

where μ and ν denote the orientation and scale of the Gabor kernels, including (frequency) and σ (bandwidth) parameters. The wave vector is defined as

$k_{\mu, \nu} = k_{\nu} \cdot e^{i\phi_{\mu}}$, where $k_{\nu} = \frac{k_{\max}}{f^{\nu}}$ and $\phi_{\mu} = \frac{\pi\mu}{8}$. k_{\max} is the maximum frequency, and f is the spacing factor between kernels in frequency domain. The term $k = 2^{-\frac{s}{2}} (2^{-\pi})$ is represented each scale value is Gabor wavelet transform. The term $-e^{\left(\frac{\sigma^2}{2}\right)}$ is represented the deduction illumination noise.

Assuming that the image is filtering by Gabor filter within different scales and phase angles, the convolution operation will be performed and its formula is expressed as

$$G_{\mu, \nu}(x, y) = f(x, y) * \varphi_{\mu, \nu}(x, y) \quad (2)$$

where $f(x, y)$ represents the input image as shown in Fig. 2(a), $\varphi_{\mu, \nu}(x, y)$ represents 2-D Gabor filter.

Generally, considering the frontal face image processing using Gabor filter, the orientation (μ) divided into eight different phase angles, and the scale (ν) classified five different scales to obtain forty kernels in Gabor filter. Using these forty kernels to filter the frontal face image, we can get forty feature images and then compute the average feature image [5] by Eq.(3). Figure 2 presents an example image and the corresponding average feature image.

$$O(x, y) = \frac{1}{N} \sum_{s=0}^4 \sum_{\theta=0}^7 G_{s, \theta}(x, y), \quad (3)$$

where N denotes the total number of kernels.

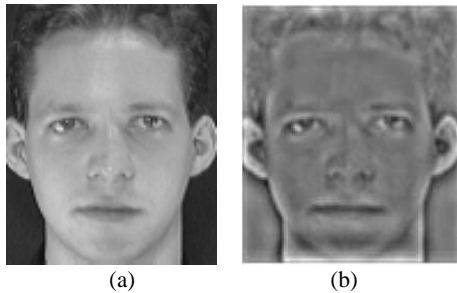


Fig. 2 (a) An example image, (b) the average feature image

For frequency field, the characteristics of high-frequency information represent the contour and texture features of an image.

However, these properties are important information in face recognition. Hence, in this paper, we utilize local Gabor filter to extract face features, it not only reduces the computational time in storage for the amount of data but also decreases the extracting time of features.

Owing to Gabor filter has the characteristic of angle symmetry, in order to avoid the redundancy operation and retain the characteristic of high-frequency information, here, we chose the orientation in the range $[90, 180]$ and three smaller scales to cover more high frequency signal. Assuming an original image is of size 46×56 pixels, Table I shows the comparison results

with 40 filters, 24 filters, and 12 filters used by us. In this paper, we select the last four orientations $\mu = (4, 5, 6, 7)$ and three scales $\nu = (0, 1, 2)$ to retain the important information and perform our experiments.

TABLE I
 FEATURE DIMENSION AND EXTRACTION TIME WITH THE DIFFERENT GABOR KERNELS SET. DIM. DENOTES FEATURE DIMENSION WHICH IS EQUAL TO IMAGE SIZE \times THE NUMBER OF FILTERS

| kernels | 5 \times 8 filters | 3 \times 8 filters | 3 \times 4 filters |
|----------|----------------------|----------------------|----------------------|
| ν | 0,1,2,3,4 | 0,1,2 | 0,1,2 |
| μ | 0,1,2,3,4,5,6,7 | 0,1,2,3,4,5,6,7 | 4,5,6,7 |
| Dim. | 12,288,000 | 7,372,800 | 3,686,400 |
| Time (s) | 3.880 | 1.816 | 0.909 |

In addition, the average feature image obtained which usually needs to complex computation, hence, we propose image blending technology to overcome this problem and simultaneously this technology can retain much edge information to against illumination factor. About image blending technology, we take the real-part response of the Gabor filter and hierarchical image blending to recover the lost edge features. Generally, image blending [6] is used to the source image and destination image to synthesize an image. According to Eq. (4), it makes the pixels to interact blending to get the virtual image (I_{ν}).

$$I_{(\nu)} = (1 - M) I_s + M \times I_t, \quad (4)$$

$$0 \leq M \leq 1,$$

where I_s and I_t represent pixel value on source image and destination image, respectively. Parameter (M) denotes the rate of interaction process.

After adequately adjusting the ratio of interpolation parameter M for the source image and destination image, we can obtain the blended image. Thus an illumination normalization image can be achieved. Figure 3 shows a hierarchical diagram of image blending; the determination of parameter M is depended on the user. In this study, it is set 0.5.

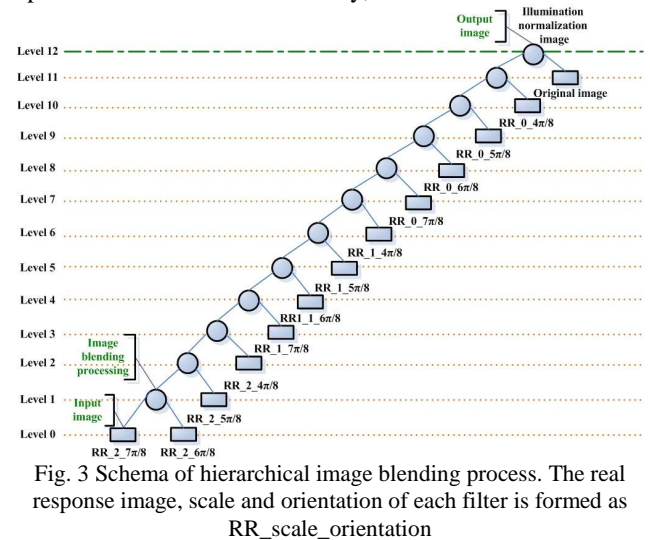


Fig. 3 Schema of hierarchical image blending process. The real response image, scale and orientation of each filter is formed as RR_scale_orientation

B. Face Alignment

Active shape models (ASM) [3] is a model-based feature matching method to constrain the shape of an object in an image. The ASM is primary based on the shape of objects as training samples, and then uses this information to find the best match of the mean shape model to the data in a new image and to obtain the transformation matrix about deformation process.

In this paper, we use the template with 68 landmarks as the align feature points, assuming a given training sample set $\Omega = \{X_1, X_2, \dots, X_N\}$ where N is the number of training samples, X_i is the shape vector with (x_i, y_i) coordinate, the coordinates of all feature points are concatenated into 2×68 -dimensional vector represented as X

$$X = (x_1, x_2, \dots, x_{68}, y_1, y_2, \dots, y_{68})^T, \quad (5)$$

where (x_i, y_i) are the coordinates of the i th landmark.

Then, we gather together feature points of all training images indicated as the vector, next the coordinates about those of points are adjusted to the centroid of feature points. And the training sets are aligned to the specified reference image selected from the training sample, thus it can make the scale regularization. After alignment processing, the principal component analysis can be used to deal with the aligned shape vectors. The mean shape is expressed

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad (6)$$

and the covariance matrix S is defined as

$$S = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T \quad (7)$$

The eigenvalues and the corresponding eigenvectors of the covariance matrix S denote as $(\lambda_1, \dots, \lambda_s)$ and (p_1, \dots, p_s) , respectively. The first t eigenvalues satisfying the $\sum_{i=1}^t \lambda_i \geq \alpha \sum_{i=1}^n \lambda_i$ are selected, where α is a selected feature ratio within the total number of features. Here, it is set 0.95 to 0.98. The recording first t eigenvectors can be formed as a matrix $\phi = (p_1, p_2, \dots, p_t)$, Finally, the shape vector can be obtained as the following formula.

$$X = \bar{X} + \phi b, \quad (8)$$

where b is the eigenvector corresponding to the formation of the shape parameter set, it is expressed as

$$b = \phi^T (X - \bar{X}). \quad (9)$$

The allowance of the similar shape is in the range of $-3\sqrt{\lambda_i} < b_i < 3\sqrt{\lambda_i}$, $i \leq t$.

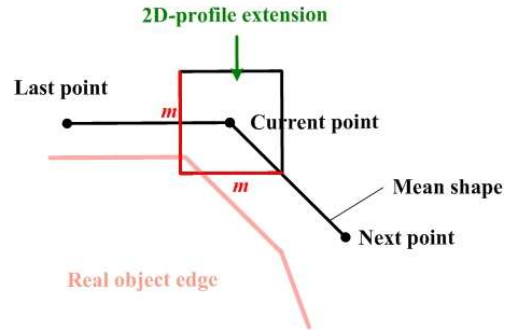


Fig. 4 Face alignment using two-dimensional profile

After processing training model, the mean shape and the transformation matrix can be obtained and further applied to search the facial features. The mean shape will be projected on the target area by using a two-dimensional structure profile to accurately locate to feature points; Figure 4 shows a two-dimensional profile diagram for the location of feature points. Figure 5 shows an example of two-dimensional profile for convergence processing of face aligning.

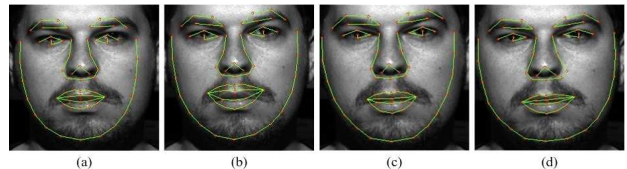


Fig. 5 Example of 2D profile for convergence process of face aligning. The number of iterations is 0, 3, 8, 24, respectively

After adjusting all feature points, in order to refine the location of feature points, we use an adaptive affine transform (AT) of global shape model to update position accuracy of feature points. The adaptive affine transform is defined as Eq. (10). By this way, let the initial position of next search be close to the positioned place. At the same time, it can modify the shape parameter b which could be the shape model more fitting the updated feature points. By iterations of the above procedure until model convergence, then we can get the consistent with the shape which fits to the current target.

For the adaptive affine transform procedure, the scale value (s_c) and rotation angle (θ_c) are firstly computing corresponding to the reference image x_{image} and test image y_{image} by using affine transform, respectively. Then the scale factor and rotation angle by using the Eq. (10) can be obtained. The x_c and y_c denote the gravity difference between x_{image} and y_{image} , respectively.

$$AT \begin{pmatrix} x \\ y \end{pmatrix} = \begin{bmatrix} x_i + x_c \\ y_i + y_c \end{bmatrix} + \begin{bmatrix} s \cdot s_c \cos(\theta + \theta_c) & s \cdot s_c \sin(\theta + \theta_c) \\ -s \cdot s_c \sin(\theta + \theta_c) & s \cdot s_c \cos(\theta + \theta_c) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (10)$$

For the current feature point (x, y) , the $\theta + \theta_c$ is the angle of rotation and $s \cdot s_c$ is the scale factor, the displacement unit is $(x_i + x_i)$ and $(y_i + y_i)$. Next, in order to calculate feature points of image, we utilize the 4-level multi-resolution pyramid strategy to search the number of rectangular side of the sampling points, and simultaneously to update the shape parameter b until the shape fits the model as a new point. Until over 95% feature points in $1/2$ image are found or equal to the number of iterations of the maximum. In our experiments, the number of iterations of the termination condition is set 24.

C. Face Recognition

Based on radial basis function (RBF) kernel function for SVM, it is only necessary two parameters (cost function (c) and test kernel function (Γ)) to adjust the model calibration. Because the input vector value is stayed in the region $[0, 1]$, the system can greatly reduce the complexity computation and has a high predictive ability. However, in order to avoid improper selection of parameters which may be easy to cause the over-fitting occurrence, here, we adopt K -fold cross-validation method [7] to evaluate the classification performance. All samples are divided into training set and test set.

III. RESULTS

In this section, we will demonstrate our proposed method using two public JAFFE and Yale_B database to evaluate the performance for localization and recognition rate.

A. Data Setting

In our experiments, we adopt the Japanese Female Facial Expression (JAFFE) Database [8] including rich expressions and Yale_B [1] including a variety of light conditions to estimate the system performance. Table II shows the total number of samples on the database, the number in a class, and the number of samples in a single-class. The experiment procedures are firstly splitting the samples into the training and test classes by 10-fold cross-validation method and then to evaluate the average recognition rate. The rule of 10-fold cross-validation strategy is that all samples are divided into 10 parts in which the nine-tenths of samples as training set and the rest part as an identification of test set, thus after operating ten times, the average recognized results can be more credible.

TABLE II
DATASET WITH TWO DATABASES

| Dataset | JAFFE | Yale_B |
|--------------------------|-------|--------|
| Total samples | 200 | 2280 |
| Class | 10 | 38 |
| Sample No. of each class | 20 | 60 |



Fig. 6 (a) Yale_B database in case of harsh light. (b) Results of normalized illumination (a)

B. Results of Illumination Normalization

Figure 6 shows some of results after normalizing illumination. From Fig. 6, it is clear that illumination factor is eliminated and simultaneously the edge features are enhanced. However, these characteristics are to provide the benefit information in the following processes of location and recognition and can yield the good achievement.

C. Evaluating Performance for Localization

In order to evaluate the difference of facial feature points location between ASM and IASM, we adopt the average localization error (E) to measure, and it is defined as

$$E = \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m |P_{i,j} - P'_{i,j}|, \quad (11)$$

where $P_{i,j}$ denotes the j th manually-labeled feature point in the i th test image from m samples. $P'_{i,j}$ denotes the correspond fitting position by ASM searching. In addition, based on Eq. (11), the improved ratio (I) is used to present the improved percentage of the proposed IASM algorithm compared with that of ASM method. It is expressed as

In Eq. (12), when I is positive, the proposed IASM method is better than the ASM method.

$$I = \frac{E_{ASM} - E_{IASM}}{E_{ASM}} \times 100\% \quad (12)$$

D. Localization Results

In this paper, we adopt two labeled modes for face image to estimate the performance of feature points location, mode 1 (Database_1) is selected manually feature points positioned on the two eyes like Fig. 7(a), mode 2 (Database_2) is used the Viola-Jones (V-J) detector [9] to detect and locate the face region like Fig. 7(b).

Table III shows the JAFFE and Yale_B database for location results. The results show that our proposed IASM method has the significant improvement effect in JAFFE or Yale_B database. Figure 8 presents the located results for some of cases of JAFFE database by using IASM.

An example with the low light image shows in Fig. 9 (a), the result after performing illumination normalization is shown in Fig. 9(b). Comparing the normal light image shown in Fig. 10, it is clear that our proposed method can specifically present the face contour.

Generally, when one image affected by light source such as Fig. 9(a), the ASM method cannot efficiently detect the face position and thus doesn't work in face alignment. However, based on our proposed method for illumination normalization pre-processing and IASM, it can successfully detect the face position and clearly locate the facial feature points, such as Fig. 11.

TABLE III
IMPROVED PERFORMANCE WITH DIFFERENT DATABASE

| Mode | ASM | IASM | I |
|---------|-------|-------------|--------|
| JAFFE_1 | 3.72 | 2.13 | 42.74% |
| JAFFE_2 | 11.63 | 5.33 | 54.14% |

| | | | |
|----------|-------|--------------|--------|
| Yale_B_1 | 5.56 | 4.07 | 26.79% |
| Yale_B_2 | 30.44 | 10.98 | 63.92% |

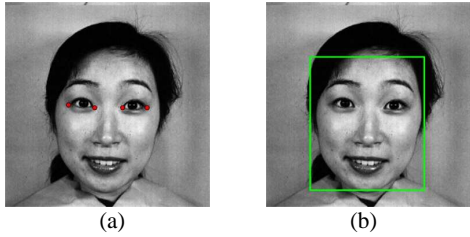


Fig. 7 Modes of location. (a) Mode 1: Locating on the two eyes, (b) mode 2: V-J face location

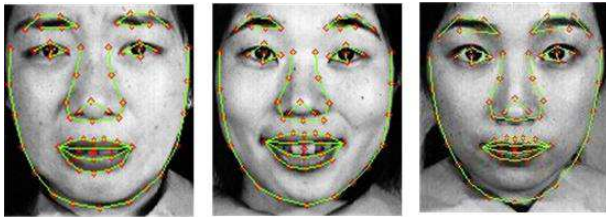


Fig. 8 JAFFE face localization using IASM



Fig. 9 Example of Yale_B illumination normalization image. (a) Example of the low light image, (b) illumination normalization result

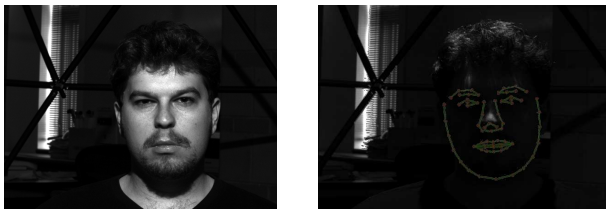


Fig. 10 Normal light image Fig. 11 Localization result for Fig. 9(a)

E. Recognition results

For recognition process, we use the receiver operating characteristic (ROC) (or called ROC curve) [10] to analyze the recognizing results. The ROC is a graphical plot of sensitivity, or true positive rate vs false positive rate for a binary classifier system. Here, we denote ROC_I. In addition, the ROC can also be represented by plotting false positive rate (FPR) vs false rejection rate (FRR). Here we denote ROC_II. Figure 12 shows the ROC_I results. From Fig. 12, the recognition result of Yale_B after processing illumination normalization can increase the sensitivity. ROC_II curve shows that the equal error rate (EER) also decreased significantly. The EER is defined as $EER=FPR=FRR$. Table 4 presents the recognition accuracy rate, ERR, and mean execution time (MT) compared with illumination normalization. After processing illumination normalization, recognition rate can be increased to 83.33%, obviously, EER is reduced to 0.21.

TABLE IV
 RECOGNITION EVALUATION. SYMBOL * INDICATES THAT THE DATABASE BE PROCESSED BY ILLUMINATION NORMALIZATION

| Database | Samples | Accuracy | EER | MT (s) |
|----------|-------------|---------------|--------------|--------------|
| JAFFE | 200 | 100.0% | 0.071 | 0.371 |
| Yale_B | 2280 | 66.71% | 0.398 | 6.254 |
| *JAFFE | 200 | 100.0% | 0.071 | 0.385 |
| *Yale_B | 2280 | 83.33% | 0.210 | 4.660 |

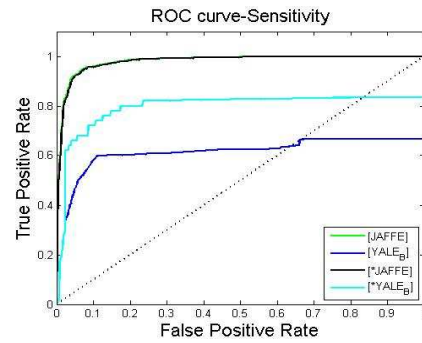


Fig. 12 Recognition results using the ROC_I curve

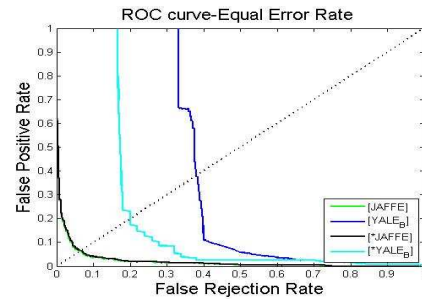


Fig. 13 Recognition results using the ROC_II curve

IV. CONCLUSIONS

In this paper presents a robust method for face alignment and recognition for illumination biases and multi-expressions. The system aims to align the face features more exactly and to gain the recognition rate. This approach can fast capture the important high-frequency information by using local Gabor filter and retain each of real-part response profile messages by using hierarchical image blending, thus face localization and recognition can be done after normalizing illumination. In the future, we will further improve the current method to deal with a part of occluded face images as well as localizing and recognizing those of images.

ACKNOWLEDGMENT

This work was supported in part by the National Science Council of Republic of China under Grant No. NSC99-2221-E-150-064.

REFERENCES

- [1] A. S. Georghiades, D. J. Kriegman, and P. N. Belhumeur, "From few to many: illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 23, no. 6, pp. 643-660, 2001.
- [2] RM Makwana, "Illumination invariant face recognition: A survey of passive methods," *Procedia Computer Science*, vol. 2, pp. 101-110, 2010.

- [3] S. Marcel and J. Keomany, "Robust-to-illumination face localization using active shape models and local binary patterns," *IDIAP Research Report*, pp. 6-47, 2006.
- [4] X. Ling, Y. Wang, Z. Zhang, and Y. Wang, "On-line signature verification based on Gabor features," in *Proc. of the 19th Annual Wireless and Optical Communications*, 2010, pp. 1-4.
- [5] H. R. Hsu, *Apply an adaptive center selection algorithm to radial basis function committee machine for face recognition*, Master Thesis, National Yunlin University of Science and Technology, 2008.
- [6] V. Rankov, R. J. Locke, R. J. Edens, P. R. Barber, and B. Vojnovic, "An algorithm for image stitching and blending," in *Proc. of SPIE*, no. 5701, 2005, pp. 190-199.
- [7] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. of Fourteenth Int. Joint Conf. on Artificial Intelligence*, pp. 1137-1143, 1995.
- [8] M. J. Lyons, S. Akamasku, M. Kamachi, and J. Gyoba, "Coding facial expressions with Gabor wavelets," in *Proc. of Int. Conf. on Automatic Face and Gesture Recognition*, pp. 200-205, 1998.
- [9] P. Viola and M. Jones, "Robust real-time object detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137-154, 2004.
- [10] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol.27, no. 8, pp. 861-874, 2006.