Face Alignment and Recognition Under Varying Lighting and Expressions Based on Illumination Normalization

Hui-Yu Huang and Shih-Hang Hsu

Abstract In this paper, we propose an efficient approach to perform face alignment and recognition under lighting and expression conditions based on illumination normalization. For face representation, lighting influence and variable expressions, especially the accuracy of facial localization and face recognition, are the important factors. Hence, the proposed approach aims to overcome these problems. This approach consists of two parts. One is to normalize illumination for face image. The other is to extract feature by means of principal component analysis and recognize face by means of support vector machine classifiers. Experimental results demonstrate that our approach can obtain a good facial alignment and face recognition with varying lighting, local distortion, and expressions.

Keywords Gabor-based filter · Improved active shape model (IASM) · Principal component analysis (PCA) · Face alignment · Face recognition · Support vector machine (SVM)

1 Introduction

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 will cause the mistake recognition. Hence, there are many studies have been worked this effect on face recognition recently [1–5]. If these factors are considers, the face recognition rate can be improved and be more robust.

Makwana [2] proposed a survey of passive methods for illumination invariant face recognition. Authors discussed some methods, such as subspace-based statistical methods, illumination invariant representation methods, model based methods, etc.

H.-Y. Huang (✉) · S.-H. Hsu
National Formosa University, 64, Wun-Hua Road, Huwei, Yunlin 632, Taiwan
e-mail: anne.huang@ieee.org; hyhuang@nfu.edu.tw

Marcel et al. [3] proposed a combined ASM with the different local binary patterns (LBP) method to address the problem of locating facial features in images of frontal faces. Lin et al. [4] proposed a face recognition scheme using Gaborface-based 2D-PCA classification based on 2D Gaborface matrices instead of transformed 1D feature vectors. In addition, in order to detect illumination effect, Zhang et al. [5] proposed a wavelet-based face recognition method to reduce lighting factor.

Owing the effect of lighting factor, it is very challenging problem for face alignment and recognition, in this paper, we propose an approach to against lighting conditions based on template matching method for précising location of facial features. In face recognition, we combine the support vector machine to classify and still keep a good recognition rate.

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

2 Proposed Method

Facial localization of landmark feature points often suffers from a variety of illumination and occlusion influences, in order to reduce these factors, we propose our approach to solve these problems.

2.1 Normalizing Illumination

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-based wavelet filter to modulate the orientations and frequencies in order to reduce the lighting influence. The Gabor-based wavelet filter is defined as

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

where μ and v denote the orientation and scale of the Gabor kernels, $z = (x, y)$ including φ (frequency) and σ (bandwidth) parameters. The wave vector $k_{\mu,v}$ is defined as $k\mu, v = kv \cdot e^{i\phi\mu}$, where $k_v = \frac{k_{max}}{f^v}$ 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 $k_v = 2^{-\frac{v}{2}}(\frac{\pi}{2})$ represents each scale value is Gabor-based wavelet transform.

The term $-e^{\left(-\frac{\sigma^2}{2}\right)}$ represents the deduction illumination noise.

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

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

where $f(x, y)$ represents the input image, $\varphi_{\mu,v}(x, y)$ 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 (v) 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).

$$O(x, y) = \frac{1}{N} \sum_{i=0}^{4} \sum_{\theta=0}^{7\pi/8} G_{i,\theta}(x, y), \tag{3}$$

where N denotes the total number of kernels.

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, in our experiments, we select the last four orientations $\mu = (4, 5, 6, 7)$ and three scales $v = (0, 1, 2)$ to retain the important information and perform our experiments.

2.2 Hierarchical Image Blending

In addition, the average feature image obtained which usually needs to complex computation, hence, we propose image blending technology to overcome this problem, at the same time, 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.
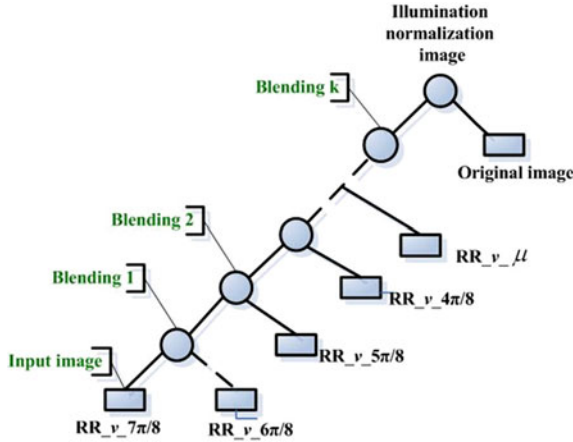
Fig. 1 Schema of hierarchical image blending process. The real response image (RR), scale (v) and orientation (μ) in Gabor filter are formed as RR_v_μ

Generally, image blending [6] is used to the source and destination images to synthesize a new image. According to Eq. (4), it makes the pixels to interact blending to get the virtual image (I_v).

$$I_v = (1 - M)I_s + M \times I_t, \quad 0 \le M \le 1, \tag{4}$$

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

After adequately adjusting 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 1 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.

2.3 Aligning Processing

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 $\Omega = \{X_1, X_2, \ldots, X_N\}$ 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

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

where (x_i, y_i) are as coordinates of the ith landmark in a face image.

Then, we sum up feature points of all training images denoted as a 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 presented the aligned average shape vectors. The mean of the n aligned shapes (X_i) is expressed as $\overline{X}$

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i, \tag{6}$$

and the covariance matrix S is defined as

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

The eigenvalues and the corresponding eigenvectors of the covariance matrix S denote as $(\lambda_i, \ldots, \lambda_s)$ and $(p_i, \ldots, p_s)$, 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 $\phi = (p_1, p_2, \ldots, p_t)$. Finally, the shape vector can be obtained as the following formula.

$$X = \overline{X} + \phi b, \tag{8}$$

where b is the eigenvector corresponding to the formation of the shape parameter set. The allowance of the similar shape is in the range of $-3\sqrt{\lambda_i} < b_i < 3\sqrt{\lambda_i}, \ i \leq t$.

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; Fig. 2 shows a two-dimensional profile diagram for the location of feature points.

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. (9). 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}

Fig. 2 Face alignment using
two-dimensional profile
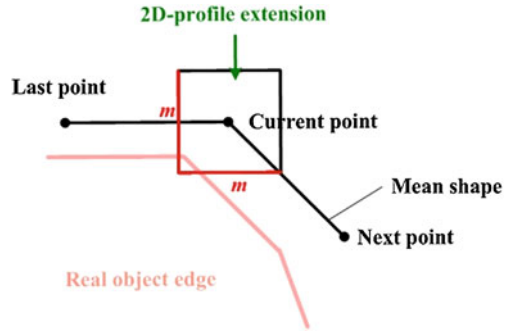
and test image y_{image} by using affine transform, respectively. Then the scale factor
and rotation angle by using the Eq. (9) can be obtained. The x_c and y_c denote the
gravity difference between x_{image} and y_{image}, respectively. The s is a scalar factor
corresponding to affine transform.

$$AT\begin{pmatrix} x \\ y \end{pmatrix} = \begin{bmatrix} x_t + x_c \\ y_t + y_c \end{bmatrix} + \begin{bmatrix} s \cdot s_c \cos(\theta + \theta_c) & s \cdot s_c \sin(\theta + \theta_c) \\ -s \cdot s_c \sin(\theta + \theta_c) & s \cdot s_c \cos(\theta + \theta_c) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (9)$$

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 $(x_t + x_i)$ and $(y_t + y_i)$. 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.

2.4 Recognition Processing

Based on radial basis function (RBF) kernel function for SVM, it is only necessary
two parameters (cost function (c) and test kernel function (*Gamma*)) to adjust the
model calibration. Because the input vector value is stayed in the range [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.

Table 1 Dataset with two database

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

3 Experimental Results

In our experiments, we adopt the JAFFE [8] database including rich expressions and Yale_B [1] including a variety of lighting conditions to estimate the system performance. Table 1 shows the total number of samples in the database, number of 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 means of 10-fold cross-validation method and then to compute 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, after operating ten times, the average recognized results can be more credible. 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

$$E = \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} |P_{i,j} - P'_{i,j}|, \tag{10}$$

where $P_{i,j}$ denotes the jth manually-labeled feature point in the ith test image from m samples. $P'_{i,j}$ denotes the correspond fitting position by ASM searching. In addition, based on Eq. (10), 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

$$I = \frac{E_{ASM} - E_{IASM}}{E_{ASM}} \times 100\%. \tag{11}$$

In Eq. (10), when I is positive, the proposed IASM method is better than the ASM method.

In this paper, we adopt two labeled modes for face image to estimate the performance of the location of feature points, mode 1 (Database_1) is selected manually feature points positioned on the two eyes shown in Fig. 3a, mode 2 (Database_2) is used the Viola-Jones (V-J) detector [9] to detect and locate the face region shown in Fig. 3b. Table 2 shows the localization errors (E) and improved result for JAFFE and Yale_B database compared ASM to IASM. From Table 2, it is evident that our proposed IASM method has the significant improvement effect in JAFFE or Yale_B database. Figure 4 presents the located results for some of cases of JAFFE database compared ASM to IASM.
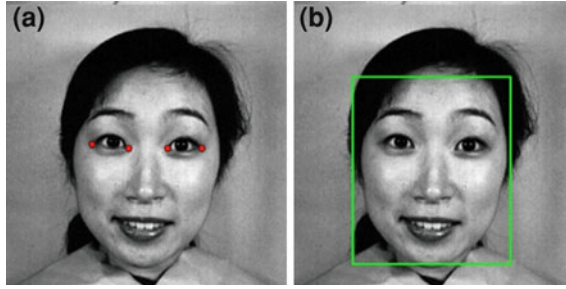
Fig. 3 Mode of locations. **a** Mode 1 locating on the two eyes, **b** mode 2 V-J face location

Table 2 Improved performance with different database

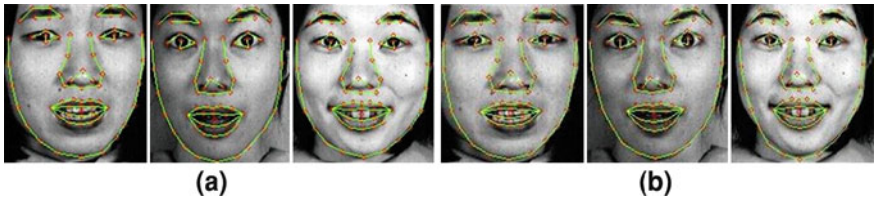| Mode	E		I (%)
	ASM	IASM	
JAFFE_1	3.72	**2.13**	42.74
JAFFE _2	11.63	**5.33**	54.14
Yale_B_1	5.56	**4.07**	26.79
Yale_B _2	30.44	**10.98**	63.92

Fig. 4 JAFFE face localization comparison of ASM (**a**) and IASM (**b**). **a** ASM result **b** IASM result

For recognition process, we use the receiver operating characteristic (ROC curve) [10], as shown in Fig. 5, to analyze the recognizing results. The ROC is a graphical plot of sensitivity, or true positive rate versus false positive rate for a binary classifier system. For Fig. 5, the recognition result of Yale_B after processing illumination normalization can increase the sensitivity. Table 3 presents the recognition accuracy rate, and mean execution time (MT) compared whether illumination normalization or not. After processing illumination normalization, recognition rate can be increased to 83.33 %.
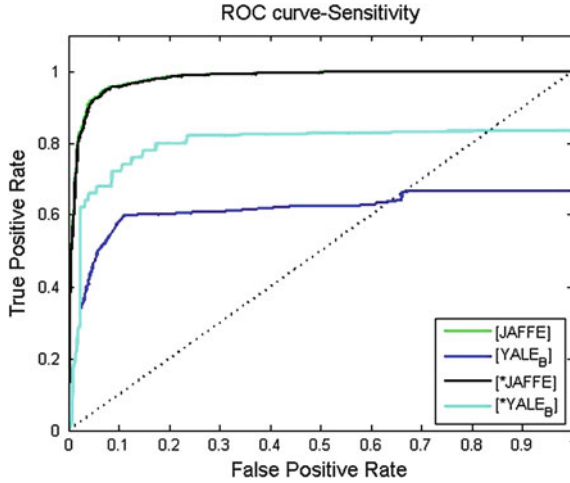
Fig. 5 Recognition results using the ROC curve

Table 3 Recognition evaluation

Database	Samples	Accuracy (%)	MT (s)
JAFFE	200	100.0	0.371
Yale_B	2280	66.71	6.254
*JAFFE	**200**	**100.0**	**0.385**
*Yale_B	**2280**	**83.33**	**4.660**

Symbol * indicates that the database has been processed by illumination normalization

4 Conclusions

In this paper, we have presented a face alignment and recognition under varying lighting conditions and multi-expressions based on illumination normalization. The purpose of this system is to align the facial features more exactly and to improve the recognition rate. Experimental results show that our proposed method is feasible and efficient to achieve face alignment and recognition.

References

1. Georghiades, A.S., Kriegman, D.J., Belhumeur, P.N.: From few to many: illumination cone models for face recognition under variable lighting and pose. IEEE Trans. Pattern Anal. Mach. Intel. **23**, 643–660 (2001)
2. Makwana, R.M.: Illumination invariant face recognition: a survey of passive methods. Procedia Comput. Sci. **2**, 101–110 (2010)
3. Marcel, S., Keomany, J.: Robust-to-illumination face localization using active shape models and local binary patterns. IDIAP Research Report, pp. 6–47 (2006)
4. Lin, W., Li, Y., Wang, C., Zhang, H.: Face recognition using Gabor face-based 2DPCA and (2D) 2PCA classification with ensemble and multichannel model. In: Proceedings of IEEE Conference of Computational Intelligence in Security and Defense Applications, pp. 1–6 (2007)
5. Zhang, T., Fang, B., Yuan, Y., Tang, Y.Y., Shang, Z., Li, D., Lang, F.: Multiscale facial structure representation for face recognition under varying illumination. Pattern Recogn. **42**, 251–258 (2009)
6. Ling, X., Wang, Y., Zhang, Z., Wang, Y.: On-line signature verification based on Gabor features. In: Proceedings of the 19th Annual Wireless and Optical Communications, pp. 1–4 (2010)
7. Kohavi, R.: A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of Fourteenth International Joint Conference on Artificial Intelligence, pp. 1137–1143 (1995)
8. Lyons, M.J., Akamasku, S. Kamachi, M., Gyoba, J.: Coding facial expressions with Gabor wavelets. In: Proceedings of International Conference on Automatic Face and Gesture Recognition, pp. 200–205 (1998)
9. Viola, P., Jones, M.: Robust real-time object detection. Int. J. Comput. Vis. **57**, 137–154 (2004)
10. Fawett, T.: An introduction to ROC analysis. Pattern Recogn. Lett. **27**, 861–874 (2006)