Expression-Invariant Face Recognition with Accurate Optical Flow

Chao-Kuei Hsieh[1], Shang-Hong Lai[2], and Yung-Chang Chen[1]

[1] Dept. of Electrical Engineering, National Tsing Hua University, Taiwan
[2] Dept. of Computer Science, National Tsing Hua University, Taiwan
d903915@oz.nthu.edu.tw, lai@cs.nthu.edu.tw,
ycchen@ee.nthu.edu.tw

Abstract. Face recognition is one of the most intensively studied topics in computer vision and pattern recognition, but few are focused on how to robustly recognize expressional faces with one single training sample per class. In this paper, we modify the regularization-based optical flow algorithm by imposing constraints on some given point correspondences to obtain precise pixel displacements and intensity variations. By using the optical flow computed for the input expressional face with respect to a referenced neutral face, we remove the expression from the face image by elastic image warping to recognize the subject with facial expression. Numerical validations of the proposed method are given, and experimental results show that the proposed method improves the recognition rate significantly.

Keywords: Face recognition, expression invariant, accurate optical flow.

1 Introduction

Face recognition has been studied for the past few decades. A comprehensive review of the related works can be found in [1]. Pose, illumination, and expression variations are three essential issues to be dealt with in the research of face recognition. In this paper, we focus on the issue of facial expression on face recognition performance.

For expression-invariant face recognition, the algorithms can be roughly divided into three categories: the subspace model based [2], morphable model based, and optical flow based approaches. In the first category, Tsai and Jan [2] applied subspace model analysis to develop a face recognition system that is tolerant to facial deformations. Note that multiple training images in each class were needed for this method.

Some other approaches were proposed to warp images to be the same shape as the ones used for training. Ramachandran et al. [3] presented pre-processing steps for converting a smiling face to a neutral face. Li et al. [4] applied a face mask for face geometry normalization and further calculated the eigenspaces for geometry and texture separately, but not all images can be well warped to a neutral image because of the lack of texture in certain regions, like closed eyes. Moreover, linear warping

H.H.S. Ip et al. (Eds.): PCM 2007, LNCS 4810, pp. 78–87, 2007.

was usually applied, which was not consistent to the non-linear characteristic of human expression movements.

Another approach is to use optical flow to compute the face warping transformation. However, it is difficult to learn the local motion within feature space to determine the expression changes of each face, since different persons have expressions with different ways. Martinez [5] proposed a weighting method that weights independently those local areas which are less sensitive to expressional changes. The intensity variations due to expression may mislead the calculation of optical flow. A more considerable drawback is that the optical flow field to every reference image is calculated, which may require huge computational cost.

Our aim in this paper is to solve the expression-invariant face recognition under the circumstance that there is only one neutral training sample per class. Under such condition, subspace approach is not appropriate for insufficient training data. In order to avoid inconsistent linear warping in model-based algorithm, our previous accurate optical flow algorithm [6] is applied for expression normalization. Not only pixel deformations but also intensity variations are computed in this algorithm. In our proposed system, the recognition accuracy was significantly improved and the optical flow computation for each input face is required only once with reference to a given universal neutral face. We show our experimental results on the Binghamton University 3D Face Expression (BU-3DFE) Database [8].

The remaining parts of this paper are organized as follows. We will introduce our proposed system in section 2, including the constrained optical flow estimation and the expression-invariant face recognition system. Section 3 presents the experimental results on applying the proposed optical flow estimation and the expression-invariant face recognition algorithm. Section 4 concludes this paper.

2 Proposed Face Recognition System

In this section, an expression-invariant face recognition system based on the constrained optical flow estimation is proposed.

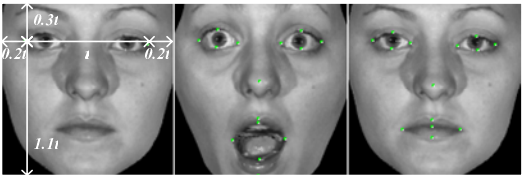
2.1 Feature Selection and Face Alignment

The main purpose of this work is to apply expression normalization on frontal faces for face recognition. We labeled 15 feature points manually (as shown in Fig. 1(b) and Fig. 1(c)) to achieve approximate face alignment, though there exist some automatic feature extraction methods [7]. With the labeled points, the distance between the outer corners of both eyes was used as the reference to scale the face images, as shown in Fig. 1(a).

2.2 Accurate Optical Flow Computation

The original formulation of the gradient-based regularization method proposed by Horn and Schunck involved minimizing an energy function of the following form:

$$E(u,v)=\int_{\Omega}\left[\left(I_x u+I_y v+I_t\right)^2+\lambda\left(u_x^2+u_y^2+v_x^2+v_y^2\right)\right]d\mathbf{x}$$



(a) (b) (C)

Fig. 1. (a) Face region selection; (b) 14 feature points on surprising face; (c) 14 feature points on neutral face

where I is the image intensity function, $\left[u(\mathbf{x},t), v(\mathbf{x},t)\right]^T$ is the motion vector to be estimated, subscripts x, y, and t denote the direction in the partial derivatives, Ω represents the image domain and λ is a parameter controlling the degree of smoothness. With the generalized dynamic image model, the optical flow constraint can be extended to [6]:

$$I_x(\mathbf{r})u(\mathbf{r}) + I_y(\mathbf{r})v(\mathbf{r}) + I_t(\mathbf{r}) + m(\mathbf{r})I(\mathbf{r}) + c(\mathbf{r}) = 0 \tag{1}$$

where $m(\mathbf{r})$ and $c(\mathbf{r})$ are the multiplier and offset fields of the scene brightness variation field and $\mathbf{r} = [x, y, t]^T$ is a point in a spatiotemporal domain. Thus, the energy function to be minimized in our algorithm can be written in a discrete form as follows:

$$\begin{aligned} f(\mathbf{u}) = &\sum_{i\in D}\left(\frac{I_{x,i}u_i + I_{y,i}v_i + I_{t,i} + m_i I_i + c_i}{\sqrt{I_{x,i}^2 + I_{y,i}^2 + I_i^2 + 1}}\right)^2 \\ &+ \lambda\sum_{i\in D}\left(u_{x,i}^2 + u_{y,i}^2 + v_{x,i}^2 + v_{y,i}^2\right) + \mu\sum_{i\in D}\left(m_{x,i}^2 + m_{y,i}^2 + c_{x,i}^2 + c_{y,i}^2\right) \end{aligned} \tag{2}$$

where the subscript i denotes the ith location, vector $\boldsymbol{u}$ is the concatenation of all the flow components u_i and v_i and all the brightness variation multiplier and offset fields m_i and c_i, λ and μ are the parameters controlling the degree of smoothness in the motion and brightness fields, and D is the set of all the discretized locations in the image domain.

We further developed a dynamic smoothness adjustment scheme to effectively suppress the smoothness of constraint at motion boundaries and brightness variation boundaries in the multiplier and offset fields, thus yielding the following energy function:

$$\begin{aligned} f(\mathbf{u}) = &\sum_{i\in D} w_i\left(\frac{I_{x,i}u_i + I_{y,i}v_i + I_{t,i} + m_i I_i + c_i}{\sqrt{I_{x,i}^2 + I_{y,i}^2 + I_i^2 + 1}}\right)^2 \\ &+ \lambda\sum_{i\in D}\left(\alpha_{x,i}u_{x,i}^2 + \alpha_{y,i}u_{y,i}^2 + \beta_{x,i}v_{x,i}^2 + \beta_{y,i}v_{y,i}^2\right) + \mu\sum_{i\in D}\left(\gamma_{x,i}m_{x,i}^2 + \gamma_{y,i}m_{y,i}^2 + \delta_{x,i}c_{x,i}^2 + \delta_{y,i}c_{y,i}^2\right) \end{aligned} \tag{3}$$

where $\alpha_{x,i}$, $\alpha_{y,i}$, $\beta_{x,i}$, $\beta_{y,i}$, $\gamma_{x,i}$, $\gamma_{y,i}$, $\delta_{x,i}$, $\delta_{y,i}$, are the weights for the corresponding components of the *i*th smooth constant along *x*- and *y*- directions.

Equation (3) can be rewritten in a matrix-vector form as $f(\mathbf{u}) = \mathbf{u}^T K\mathbf{u} - 2\mathbf{u}^T\mathbf{b} + \mathbf{c}$, thus minimizing this quadratic and convex function is equivalent to solving a large linear system $\boldsymbol{K}\mathbf{u} - \mathbf{b} = 0$, where $\mathbf{b} = \begin{bmatrix} -\mathbf{e}_{xt}^T & -\mathbf{e}_{yt}^T & -\mathbf{e}_{It}^T & -\mathbf{e}_t^T \end{bmatrix}^T \in \hbar^{4wh}$ (*w* and *h* denote image width and height, and $\mathbf{e}_{xt}$, $\mathbf{e}_{yt}$, $\mathbf{e}_{It}$, $\mathbf{e}_t$ are all *wh*-dimensional vectors with entries $w_i I_{x,i} I_{t,i} / N_i$, $w_i I_{y,i} I_{t,i} / N_i$, $w_i I_i I_{t,i} / N_i$, and $w_i I_{t,i} / N_i$, respectively. $N_i = I_{x,i}^2 + I_{y,i}^2 + I_i^2 + 1$ is the normalization term and w_i is the weight.) and

$$\boldsymbol{K} = \begin{bmatrix} \lambda \boldsymbol{K}_{s,\alpha} + \boldsymbol{E}_{xx} & \boldsymbol{E}_{xy} & \boldsymbol{E}_{xI} & \boldsymbol{E}_x \\ \boldsymbol{E}_{xy} & \lambda \boldsymbol{K}_{s,\beta} + \boldsymbol{E}_{yy} & \boldsymbol{E}_{yI} & \boldsymbol{E}_y \\ \boldsymbol{E}_{xI} & \boldsymbol{E}_{yI} & \mu \boldsymbol{K}_{s,\gamma} + \boldsymbol{E}_{II} & \boldsymbol{E}_I \\ \boldsymbol{E}_x & \boldsymbol{E}_y & \boldsymbol{E}_I & \mu \boldsymbol{K}_{s,\delta} + \boldsymbol{E} \end{bmatrix} \in R^{4wh \times 4wh}$$

The matrix $\boldsymbol{K} \in R^{4wh \times 4wh}$ is a symmetric positive-definite matrix, where $\boldsymbol{E}_{xx}$, $\boldsymbol{E}_{xy}$, $\boldsymbol{E}_{xI}$, $\boldsymbol{E}_x$, $\boldsymbol{E}_{yy}$, $\boldsymbol{E}_{yI}$, $\boldsymbol{E}_y$, $\boldsymbol{E}_{II}$, $\boldsymbol{E}_I$, and $\boldsymbol{E}$ arise from the optical flow constraints and they are all $wh \times wh$ diagonal matrices with diagonal entries $w_i I_{x,i}^2 / N_i$, $w_i I_{x,i} I_{y,i} / N_i$, $w_i I_{x,i} I_i / N_i$, $w_i I_{x,i} / N_i$, $w_i I_{y,i}^2 / N_i$, $w_i I_{y,i} I_i / N_i$, $w_i I_{y,i} / N_i$, $w_i I_i^2 / N_i$, $w_i I_i / N_i$, and w_i / N_i, respectively. We approximate the partial derivative of the flow field using forward difference, thus the matrix $\boldsymbol{K}_{s,\alpha}$, which comes from the smoothness constraint, takes the following form:

$$\boldsymbol{K}_{s,\alpha} = \begin{bmatrix} \boldsymbol{G}_1 & \boldsymbol{H}_1 & & \\ \boldsymbol{H}_1 & \boldsymbol{G}_2 & \hbar & \\ & \hbar & \hbar & \boldsymbol{H}_{w-1} \\ & & \boldsymbol{H}_{w-1} & \boldsymbol{G}_w \end{bmatrix} \in R^{wh \times wh}$$

where the submatrices $\boldsymbol{G}_i$ and $\boldsymbol{H}_i$ are all of size $h \times h$. The other matrices $\boldsymbol{K}_{s,\beta}$, $\boldsymbol{K}_{s,\gamma}$, and $\boldsymbol{K}_{s,\delta}$ are obtained with α replaced by β, γ, and δ, respectively. The detailed definitions of these submatrices can be found in [6].

We apply the preconditioned conjugate gradient algorithm to solve this linear system efficiently. To accelerate the convergence speed of the conjugate gradient algorithm, the preconditioner $\boldsymbol{P}$ must be carefully designed. It must be a good approximation to $\boldsymbol{K}$ and a fast numerical method is required to solve the linear system $\boldsymbol{Pz} = \boldsymbol{r}$ in each iteration of the preconditioned conjugate gradient algorithm. The incomplete Cholesky factorization is a good choice with the preconditioner P is given in the following form $\boldsymbol{P} = \boldsymbol{L}\boldsymbol{L}^T \approx \boldsymbol{K}$. The detailed steps of this algorithm (ICPCG) are summarized as follows:

(a) Initialize $\mathbf{u}_0$; compute $\mathbf{r}_0 = b - \boldsymbol{K}\mathbf{u}_0$; $k = 0$.

(b) Solve $\boldsymbol{P}\mathbf{z}_k = \mathbf{r}_k$; $k = k + 1$.

(c) If $k=1, \mathbf{p}_1=\mathbf{z}_0$; else compute $\beta_k=\mathbf{r}_{k-1}^T\mathbf{z}_{k-1}/\mathbf{r}_{k-2}^T\mathbf{z}_{k-2}$, and update $\mathbf{p}_k=\mathbf{z}_{k-1}+\beta_k\mathbf{p}_{k-1}$

(d) Compute $\alpha_k=\mathbf{r}_{k=1}^T\mathbf{z}_{k-1}/\mathbf{p}_k^T\boldsymbol{K}\mathbf{p}_k$.

(e) Update $\mathbf{r}_k=\mathbf{r}_{k-1}-\alpha_k\boldsymbol{K}\mathbf{p}_k$, $\mathbf{u}_k=\mathbf{u}_{k-1}+\alpha_k\mathbf{p}_k$.

(f) If $\mathbf{r}_k\approx\mathbf{0}$, stop; else go to step (b).

In our application, the initial value of $\mathbf{u}_0$ is calculated using RBF interpolation from all the feature points. This optical flow algorithm has been proven superior to others. However, the proposed ICPCG algorithm cannot guarantee the computed optical flow is consistent to the motion at the corresponding feature points. Therefore, we modify the optimization problem to be

$$\begin{aligned} &\textit{minimize } f(\mathbf{u})=\mathbf{u}^T K\mathbf{u}-2\mathbf{u}^T\mathbf{b}+\mathbf{c}, \\ &\textit{subject to } u_{(x_i,y_i)}=\bar{u}_i, \textit{ and } v_{(x_i,y_i)}=\bar{v}_i, \forall(x_i,y_i)\in S \end{aligned} \tag{4}$$

where S is the set of feature points and $(\bar{u}_i,\bar{v}_i)$ is the specified optical flow vector at the i-th feature point. Instead of using the Lagrange multiplier, we modify the original procedure of ICPCG to satisfy these specified data constraints. After initializing the first guess of $\mathbf{u}_0$ in step (a), we reset the motion vectors at the feature points to be the specified flow vectors. After solving z_k in step (b), the z values for the facial feature pixels in S is then set to 0. By doing so, the motion vector value of the feature points will not be changed in the decision step (c) of the next gradient direction.

2.3 Face Recognition Flowchart

Fig. 2 shows the flowchart of our proposed face recognition system (lower part) and the benchmark system using PCA dimension reduction (upper part).

In our proposed system, only the optical flow field for the input face image with respect to the universal neutral face image NE_1 is required to compute for each

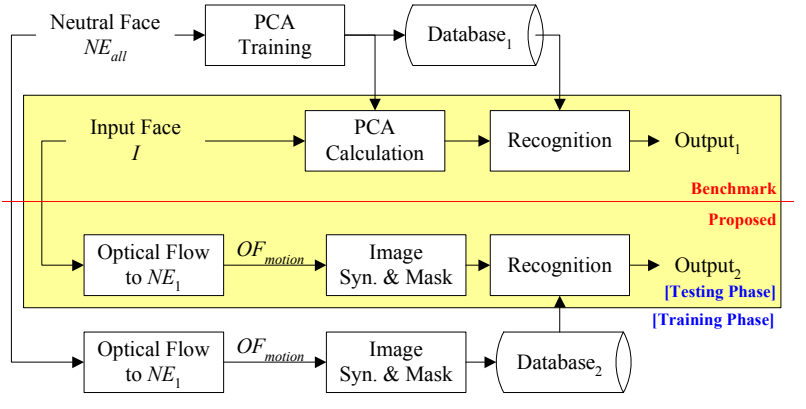


Fig. 2. Flowchart of the proposed face recognition system

training and testing image. In the image synthesis block, we discard the intensity part of optical flow, m_i and c_i, and take the motion part only, u_i and v_i, for neutral face warping. A mask is then put on the synthesized image to remove some border regions. Such procedure is applied on both training phase and testing phase as the pre-processing step. It can unify the geometry coordinate of training neutral images in the training phase and neutralize the input expressional image in the testing phase. A simple direct subtraction and the nearest-neighbor classification was employed in the face recognition.

3 Experimental Results

Our experiments were performed on the Binghamton University 3D Face Expression (BU-3DFE) Database [8].

Fig. 3. Sample images in BU-3DFEDB. The left-top most is the neutral face. The others are sad, fear, disgust, happy, sad, and surprise expressions in columns; level1 to level 4 in rows.

The BU-3DFE database contains the face images and 3D face models of 100 subjects (56 females and 44 males) with neutral face and 6 different expressions (happy, disgust, fear, happy, sad, and surprised) at different levels (from level 1-weak to 4-strong), but only 2D face images were used in our experiments. Fig. 3 shows the 25 normalized face images of one subject with the proposed method.

3.1 De-expression with Accurate Optical Flow

We try to verify the effectiveness of expression normalization with accurate optical flow by checking the face recognition rates. In the training phase for original data, we first use PCA to calculate the reduced vector for all the 100 neutral images of each subject. In the testing phase, the input image is projected to the PCA subspace and

classified by the nearest neighbor as shown in Fig. 2. The average recognition rate is 56.88%, and the recognition results are shown in Fig. 4 for all expressions and levels separately. The surprising and disgust expressions had the worst results, and the recognition rate is decreased while the degree of facial expression gets stronger.

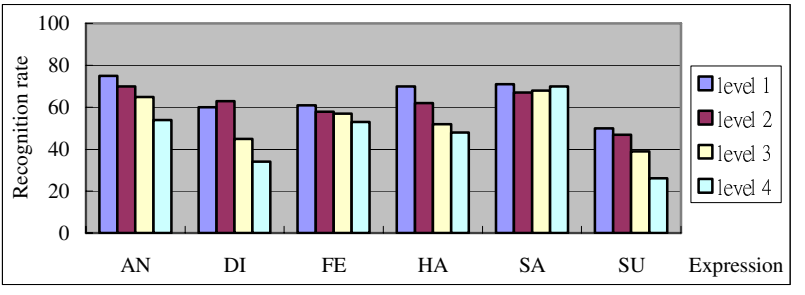


Fig. 4. Recognition rates of original images under different expressions and levels (with average recognition rate 56.88%)

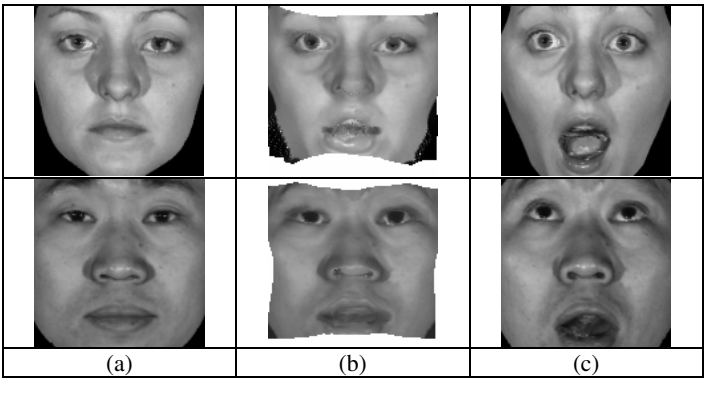


Fig. 5. Illustrations of expression normalization using accurate optical flow, (a) original neutral faces; (b) expression normalized faces from (c) to (a); (c) surprising face at level 4

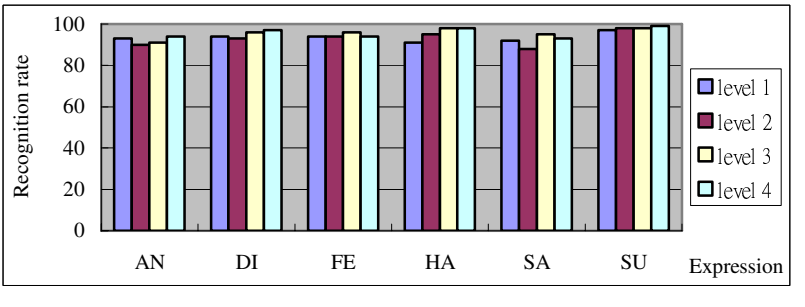


Fig. 6. Recognition rates of reconstructed face with original optical flow to known subject. (average recognition rate: 94.5%)

With the accurate optical flow between the expression and neutral faces of the same subject, we can generate a synthesized neutral face image as shown in Fig. 5. The recognition rates by using the synthesized neutral face images with the optical flow warping are shown in Fig. 6, which can achieve an average value of 94.5%.

The recognizing process is basically same as that in the benchmark test, except that regions in the training neutral faces corresponding to the black and white parts are unified to white. As shown in the result, the constrained optical flow estimation is excellent for removing expression deformation even with exaggerative facial motions. For example, the recognition rate for surprising level 4 is improved from 26% to 99%.

3.2 Face Recognition with Synthesized Neutral Face

In the face recognition experiment, the constrained optical flow estimation algorithm is employed to find the warping transformation for the facial expression image and synthesize the corresponding neutral face image for classification. Since intensity variation is also involved in the optical flow computation, the synthesized image looks just like the person with neutral face even if the reference neutral face belongs to someone else, like Fig. 7. In the experiments, the universal neutral face image NE_1 was arbitrarily selected as the neutral face of #1 subject, and the comparison was based on the average recognition rates.

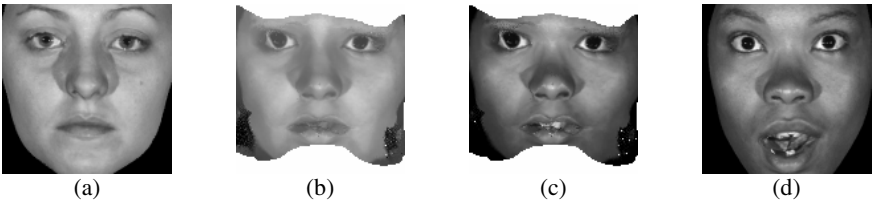


Fig. 7. Illustrations of accurate optical flow from one person to another, (a) referenced neutral face; (b) synthesized face from (d) to (a) with full optical flow; (c) synthesized face from (d) to (a) with motion optical flow only; (d) input image of surprising face at level 4

Fig. 8. Illustrations of accurate optical flow from one person to another, (a) referenced neutral face; (b) mask image; (c) bilinear warped face to from (a) NE_1; (d) synthesized face with only motion vector from (a) to NE_1

Therefore, as mentioned in section 2.3, only motion information in the optical flow estimation was used for neutral face image synthesis. As shown in Fig. 7(c), there will be some missing area in the synthesized image due to large motion, and a mask is employed to reduce the area of image comparison.

Fig. 9 and Fig. 10 depict the recognition results with partially masked synthesized image using different methods, PCA and direct subtraction. Obviously, the direct subtract algorithm outperforms the PCA method, which means that the synthesized images are very similar to each other due to the same reference neutral image.

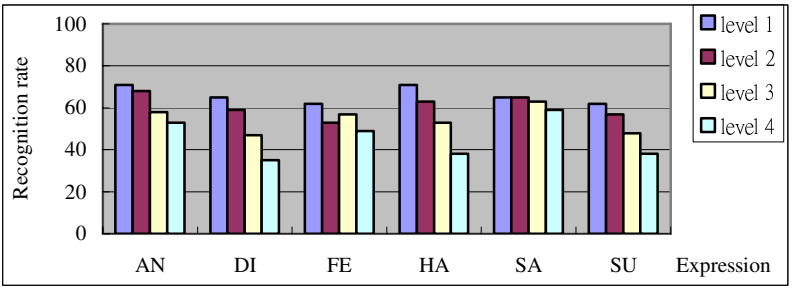


Fig. 9. Recognition rate from the synthesized neural face image by using the estimated optical flow only, compared to synthesized neutral face by using the same method, by using PCA dimensional reduction with 95% energy preservation. The average recognition rate is 57.19%.

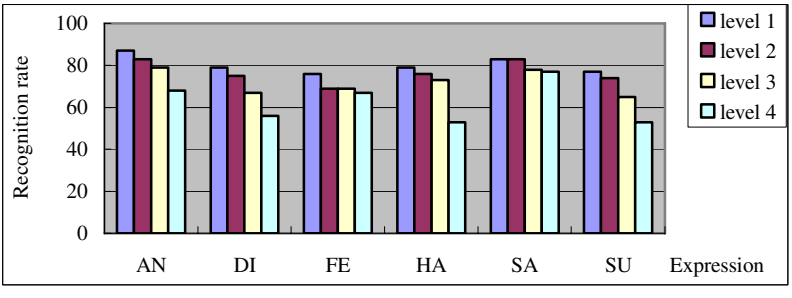


Fig. 10. Recognition rate from the synthesized neural face image by using the estimated optical flow only, compared to synthesized neutral face by using the same method, by using direct difference and nearest neighbor method. The average recognition rate is 73.48%.

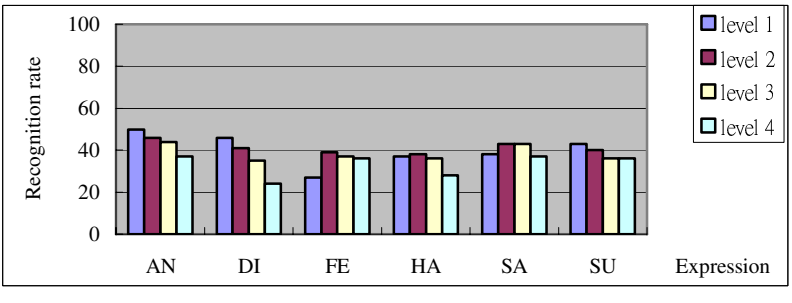


Fig. 11. Recognition rate from the synthesized face image by using the estimated optical flow only, compared to bilinearly warped neutral face, by using direct difference and the nearest neighbor classification. The average recognition rate is 39.02%.

Therefore, direct subtraction is more suitable in this application. Our proposed method (Fig. 10) is better than the benchmark result (Fig. 4) by 17% in average.

Different pre-processing procedures are performed in our experiments for comparison: one is bilinear warping from neutral faces (Fig 8(a)) to the universal neutral image NE_1, like Fig. 8(c); the other is by using the same procedure for the testing image, e.g. Fig. 8(d). The experimental results are shown in Fig. 11 and Fig. 10, respectively. We can see that the second way is better than the first one, since the face geometry is more uniform in the processed testing image.

4 Conclusion and Future Works

In this paper, we apply the constrained optical flow estimation for removing the facial expression for expression-invariant face recognition. With some modification in the optical flow computation, we can guarantee that the motion vector of the feature points to be consistent during the ICPCG iterations. The proposed method can improve the recognition rate about 17% under the circumstance that only one training neutral image of each subject is available, and the optical flow computation is performed only once for each testing image.

The lighting condition is restricted in our experiments. It is apparent that accurate optical flow varies dramatically when the illumination variation is large, which may influence the expression-normalized face image. In the future works, we plan to compensate the global lighting effect by using the spherical harmonics technique. After removing the illumination variations, the optical flow can be subsequently used to estimate the warping transformation due to facial expression.

References

1. Li, S.Z., Jain, A.K.: Handbook of Face Recognition. Springer, NewYork (2005)
2. Tsai, P.-H., Jan, T.: Expression-Invariant Face Recognition System Using Subspace Model Analysis. In: 2005 IEEE Conference on Systems, Man and Cybernetics (October 10-12, 2005) vol. 2, pp. 1712–1717 (2005)
3. Ramachandran, M., Zhou, S.K., Jhalani, D., Chellappa, R.: A Method for Converting a Smiling Face to a Neutral Face with Applications to Face Recognition. In: Proceedings of ICASSP 2005, vol. 2 (March 18-23, 2005)
4. Li, X., Mori, G., Zhang, H.: Expression-Invariant Face Recognition with Expression Classification. In: CRV 2006. The 3rd Canadian Conference on Computer and Robot Vision 2006 (June 07-09, 2006)
5. Martinez, A.M.: Recognizing expression Variant Faces from a Single Sample per Class. In: CVPR 2003. Proceedings of 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (June 18-20, 2003)
6. Teng, C.-H., Lai, S.-H., Chen, Y.-S., Hsu, W.-H.: Accurate optical flow computation under non-uniform brightness variations. Computer Vision and Image Understanding 97, 315–346 (2005)
7. Lin, S.C., Li, M.J., Zhang, H.J., Cheng, Q.S.: Ranking prior Likelihood Distributions for Bayesian Shape Localization Framework. In: Proceedings of ICCV 2003 (2003)
8. Yin, L., Wei, X., Sun, Y., Wang, J., Rosato, M.J.: A 3D Facial Expression Database For Facial Behavior Research. In: FGR 2006. Proceedings of 7th International Conference on Automatic Face and Gesture Recognition, pp. 211–216 (April 10-12, 2006)