Expression-Invariant Face Recognition With Constrained Optical Flow Warping

Chao-Kuei Hsieh, Shang-Hong Lai, *Member, IEEE*, and Yung-Chang Chen, *Fellow, IEEE*

***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 compute precise pixel displacements and intensity variations. By using the optical flow computed for the input expression variant face with respect to a reference neutral face image, we remove the expression from the face image by elastic image warping to recognize the subject with facial expression. Experimental validation is given to show that the proposed expression normalization algorithm significantly improves the accuracy of face recognition on expression variant faces.**

***Index Terms*—Constrained optical flow, expression invariance, expression normalization, face recognition.**

I. Introduction

FACE recognition has been studied for the past few decades, from initially the 2-D methods to some recent 3-D methods. Comprehensive reviews of the related works can be found in [25] and [36]. In 2-D appearance-based face recognition methods, different dimension reduction techniques have been applied. Principal component analysis (PCA) [31] is used to find a t-dimensional subspace whose basis vectors correspond to the maximal variance direction in the original image space. Linear discriminant analysis (LDA) [18] finds the vectors to form the subspace that best discriminate among classes. The between-class scatter matrix and the within-class scatter matrix are considered in this algorithm. Independent component analysis (ICA) [1] minimizes both second-order and higher-order dependencies in the input data and attempts to find the basis along which the data are statistically independent. Some other discussions are made on handling the singularity problem when the LDA was implemented. Beyond these, kernel methods were proposed. Some nonlinear projection preprocesses were used to increase the dimension for discrimination and took higher order correlation into account, before adopting the traditional PCA or LDA method. Kernel-PCA and Kernel-LDA techniques were developed based on this idea [33]. Recently, discriminative common vector (DCV) [2] was proposed to project the data onto the null space, rather than the range space, of the covariance matrix. Furthermore, nonlinear kernel feature was applied on the DCV once again [10], which can provide better performance. In all of the dimension reduction algorithms, multiple training images are required for each subject.

Even though the 2-D face recognition methods have been actively studied in the past two decades, there are still inherent disadvantages and drawbacks. Pose, illumination, and expression variations are three essential issues to be dealt with in the research of face recognition. It was shown that recognition rate will drop dramatically when the head pose and illuminating variations are too large, or there is expression on the face. Facial expression can be considered as a gesture executed by the facial muscles or the feeling expressed on a person's face (e.g., Fig. 6). Common facial expressions include anger, disguise, fear, happiness, sadness, and surprise.

To overcome the problem of face recognition under varying illumination, several methods have been proposed. The illumination cone methods [8] exploit the fact that the set of images of an object at a fixed pose under all possible illumination conditions forms a convex cone in the high-dimensional image space. The spherical harmonic-based technique [27] employs a simple closed-form basis representation for the irradiance in terms of spherical harmonic coefficients of the incident illumination. In addition, the quotient image-based approach [28] uses the definition of an illumination invariant signature image to enable an analytic generation of the image space with varying illumination.

Many efficient algorithms have been proposed to overcome the pose variation problem. Kim and Kittler proposed the locally linear discriminant analysis (LLDA) method [13]. The underlying idea is that global nonlinear data structures are locally linear and local structures can be linearly aligned. Rama and Tarres used the P2CA [24] and PLDA [25] to find the correspondence between the texture map and input image. Chai *et al.* proposed a local linear regression (LLR) method [3], which extends the basic idea of linear object class to generate the virtual frontal view from a given non-frontal image. Hsieh and Chen also proposed a kernel-based pose invariant face recognition algorithm [12] by integrating the nonlinearity of kernel function and linear assumption of LLR to further improve the performance.

To date, there was not much research effort on overcoming the expression variation problem in face recognition, although facial expression recognition has been studied for many years. The first paper to fully address the expression invariant problem is proposed by Martinez [19] based on a probabilistic approach.

Manuscript received September 12, 2007; revised November 18, 2008. Current version published May 15, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Horace Ho-Shing IP.

C.-K. Hsieh and Y.-C. Chen are with the Department of Electrical Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan (e-mail: d903915@oz.nthu.edu.tw; ycchen@ee.nthu.edu.tw).

S.-H. Lai is with the Department of Computer Science, National Tsing Hua University, Hisnchu 30013, Taiwan (e-mail: lai@cs.nthu.edu.tw).

Digital Object Identifier 10.1109/TMM.2009.2017606

It weights all predefined local face regions based on how much each region is affected by facial expression. A training procedure was employed for each specific expression, but different subjects with different culture may have different optimal weighting values. Some face recognition systems based on video sequences [9], [23], [25] can handle a wide range of facial expressions, since a video sequence contains more information than a single face image. A conventional face recognition system based on a single face per person, however, is valuable in many applications. For expression-invariant face recognition on images, the algorithms can be roughly divided into three categories: the subspace model based, morphable model based, and optical flow based approaches. In the first category, Tsai and Jan [30] applied subspace model analysis to develop a face recognition system that is insensitive 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. The idea of separately modeling texture and geometry information has been applied in active shape model and active appearance model (ASM/AAM) [4], [5]. Face geometry is defined via a set of feature points in ASM, while face texture can be warped to the mean shape in AAM. Ramachandran *et al.* [26] presented preprocessing steps to convert a smiling face to a neutral face. Li *et al.* [16] 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 was usually applied, which was not consistent to the nonlinear characteristics of facial expression movements.

Another approach is to use optical flow to compute the face warping transformation. Optical flow has been used in the task of expression recognition [6], [32]. 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 in different motion styles. Martinez [21] proposed a weighting method that independently weighs the local areas which are less sensitive to expressional changes. The intensity variations due to expression may mislead the calculation of optical flow. A precise motion estimation method was proposed in [20], which can be further applied for expression recognition. However, the proposed motion estimation did not consider intensity changes due to different expressions. In addition, it requires the optical flow estimation of an expression variant face to all reference images, which is very computationally intensive.

In this paper, we focus on the problem of face recognition from a single 2-D image of a face with an arbitrary facial expression. Note that this work is not about facial expression recognition. For many practical face recognition problem settings, like using a passport photo for face identification at custom security or identifying a person from a photo on the ID card, it is difficult to gather multiple training images for each subject, especially under different expressions. Therefore, our goal is to solve the expressive face recognition problem under the condition that the training database contains only neutral face images with one neutral face image per subject. Under such a condition, subspace learning approach is not appropriate because of insufficient training data. In order to avoid inconsistent linear warping in the model-based approach, we propose a facial expression normalization procedure based on constrained optical flow estimation. Not only pixel deformations, but also intensity variations are considered in the optical flow estimation algorithm. Based on the proposed facial expression normalization technique, the recognition accuracy can be significantly improved and the optical flow computation for each input face is required only once with reference to a given standard neutral face. We will show superior face recognition results by using the proposed algorithm on the Binghamton University 3D Face Expression (BU-3DFE) Database [34] with manually labeled facial feature points, since we do not attempt to solve the automatic facial landmark localization problem in this work.

The remaining of this paper is organized as follows. We describe the core optical flow computational technique in Section II, including the motivation and constrained optical flow estimation. The proposed expression-invariant face recognition system is presented in Section III. Section IV gives the experimental results by applying the proposed expression-invariant face recognition algorithm. Finally, we conclude this paper in the last section.

II. Optical Flow Computational Technique

A. Motivation and Basic Idea

In this paper, we focus on to the problem of recognizing a person from an expressive face image when the training database contains only neutral face images with only one neutral face image per person. Traditional learning techniques cannot be applied when only one sample image per class is available. Instead, image matching techniques are employed in this case. If all images are normalized to zero mean and unit variance, the correlation between an image $\boldsymbol{I}$ and another image $\boldsymbol{T}$ is directly related to the Euclidean distance in the original space, i.e., $\|\boldsymbol{I}-\boldsymbol{T}\|$ [7]. For the case with one sample per class, a test image is classified based on nearest-neighbor search in the training set. However, the distance metrics usually used in the nearest-neighbor search, including the correlation and Euclidean distance, cannot account for the deviations due to facial expression, which significantly degrade the face recognition accuracy.

Some authors have proposed different approaches [16], [21] to deal with variations due to facial expression. Martinez [21] proposed to compute the optical flow between the testing and training face images, use the optical flow at each pixel to determine the associated weight, and compute a weighted Euclidean distance, denoted by $wc(c) = \|\mathbf{W}_c(\mathbf{I}_c - \mathbf{T})\|$, by using the optical flow between the testing face image $\mathbf{T}$ and a training neutral image $\mathbf{I}_c$ of subject c as the weighting coefficients in the weight matrix $\mathbf{W}_c$. However, it cannot guarantee that the computed optical flow corresponds to the exact pixels in different images, since the intensity variations due to facial expression may yield errors in the computed optical flow. Another problem with this approach is that the optical flow computation between

the input image and all training images is computationally intensive, especially when the total number of subjects in the database is large.

Another approach is using feature point correspondence to remove the facial expression with appropriate image warping to achieve expression-invariant face recognition. Li *et al.* [16] designed a simplified and triangulated mask in an expression-invariant face recognition system. In this system, some important face features, such as eye corners, nose tip, and mouse corners are extracted. Then, all the other vertices on the mask can be fitted using the symmetry and common knowledge of face structures, and the one-to-one correspondences among the vertices for describing the facial motion are established. Finally, an affine warping for the texture in each triangle is determined from an expressive face to a reference neutral face. The main problem with this approach is that only the geometric variations, but no intensity variations, of the corresponding feature points are considered in the expression removal. In addition, affine warping is insufficient to describe nonlinear facial expression motion.

In this paper, we propose to combine the advantages of the above two approaches: the unambiguous correspondence of feature point labeling and the flexible representation of optical flow computation. With doing so, the correspondences of facial feature points in expression variant images can provide reliable constraints for the optical flow computation of the entire face image, and the optical flow can well describe the nonlinear facial expression motion. Meanwhile, the pixel intensity change caused by expression is another issue to be addressed. When handling the feature point correspondence, not only the position movements but also the intensity changes should be considered at the same time, that is $I_2(\boldsymbol{a}_i+\delta\boldsymbol{a}_i) = I_1(\boldsymbol{a}_i)-I_1(\boldsymbol{a}_i)\times m(\boldsymbol{a}_i)- c(\boldsymbol{a}_i)$, where I_1 and I_2 are image intensity functions, $\boldsymbol{a}_i$ is the i^{th} feature point, $\delta\boldsymbol{a}_i$ is the displacement vector, m and c denote the multiplier and offset fields. Among these parameters, I_1 and I_2 can be obtained from images, feature point movement $\delta\boldsymbol{a}_i$ can be obtained from feature point extraction, but $m(\boldsymbol{a}_i)$ and $c(\boldsymbol{a}_i)$ need to be further determined. In turn, other triangular vertices and pixels inside the mask triangles can be treated in the same way, i.e., $I_2(\boldsymbol{b}_i + \delta\boldsymbol{b}_i) = I_1(\boldsymbol{b}_i) - I_1(\boldsymbol{b}_i) \times m(\boldsymbol{b}_i) - c(\boldsymbol{b}_i)$, but with $\delta\boldsymbol{b}_i$ undetermined. Note that $\delta\boldsymbol{b}_i$, $m(\boldsymbol{b}_i)$, and $c(\boldsymbol{b}_i)$ should be related to $\delta\boldsymbol{a}_i$, $m(\boldsymbol{a}_i)$, and $c(\boldsymbol{a}_i)$. In summary, the problem is to solve the variables $m(\boldsymbol{a}_i)$, $c(\boldsymbol{a}_i)$, $\delta\boldsymbol{b}_i$, $m(\boldsymbol{b}_i)$, and $c(\boldsymbol{b}_i)$ under the constraint of motion vector $\delta\boldsymbol{a}_i$ at some sparse feature points as well as some smoothness constraints for neighboring points in the image plane.

B. Regularization-Based Optical Flow Computation

1) Energy Function and Adaptive Smoothness Adjustment: In order to formulate a more general equation and solve the problem systematically, we leave discussion of the constraints on the motion vectors at feature points behind, and consider the optical flow estimation on all the pixels. Under the intensity conservation constraint of conventional optical flow, the difference between the corresponding points in different images should be zero, that is $I(\mathbf{r} + \delta\mathbf{r}) - I(\mathbf{r}) = \delta I(\mathbf{r}) = 0$, The difference $\delta I(\mathbf{r})$ of intensity values can be defined as $I_x(\mathbf{r})u(\mathbf{r}) + I_y(\mathbf{r})v(\mathbf{r}) + I_t(\mathbf{r})$, where I is the image intensity function, $[u(\mathbf{x},t), v(\mathbf{x},t)]^T$ is the motion vector to be estimated, subscripts x, y, and t denote the spatiotemporal direction in the partial derivatives, with $\mathbf{x} = [x,y]^T$ being a point in the spatial domain, and $\mathbf{r} = [x,y,t]^T$ is a point in the spatiotemporal domain. The original formulation of the gradient-based regularization method proposed by Horn and Schunck [11] involves minimizing an energy function of the following form:

$$E(u,v)=\int_{\Omega}\left[(I_x u+I_y v+I_t)^2+\lambda\left(u_x^2+u_y^2+v_x^2+v_y^2\right)\right]d\mathbf{x} \tag{1}$$

where Ω represents the 2-D image domain and λ is a parameter controlling the degree of smoothness in the flow field.

The intensity conservation constraint, however, is generally not feasible in many circumstances. Therefore, with the generalized dynamic image model (GDIM) proposed by Negahdaripour and Yu [22], i.e., $I(\mathbf{r}+\delta\mathbf{r})-I(\mathbf{r}) = \delta I(\mathbf{r}) = -m(\mathbf{r})I(\mathbf{r})-c(\mathbf{r})$, the optical flow constraint can be generalized to [29]

$$I_x(\mathbf{r})u(\mathbf{r}) + I_y(\mathbf{r})v(\mathbf{r}) + I_t(\mathbf{r}) + m(\mathbf{r})I(\mathbf{r}) + c(\mathbf{r}) = 0 \tag{2}$$

where $m(\mathbf{r})$ and $c(\mathbf{r})$ denote the multiplier and offset factors of the scene brightness variation field. Teng *et al.* [29] proposed to minimize the following discrete energy function to compute the optical flow:

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

where the subscript i denotes the ith location, vector $\mathbf{u}$ is the concatenation of all the flow components u_i and v_i and all the brightness variation multiplier and offset factors 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.

Furthermore, an adaptive smoothness adjustment scheme was developed [29] to effectively suppress the smoothness of constraint at motion boundaries and brightness variation boundaries in the multiplier and offset factors. The strength of smoothness constraint is controlled by assigning an appropriate weight for each component of the constraint, thus yielding the following energy function:

$$\begin{aligned} f(\mathbf{u}) = &\sum_{i\in D} w_i\left(\frac{I_{x,i}u_i + I_{y,i}v_i + I_{t,i} + m_i I_i + c_i}{\sqrt{I_{x,i}^2 + I_{y,i}^2 + I_i^2 + 1}}\right)^2 \\ &+ \lambda\sum_{i\in D}\left(\alpha_{x,i}u_{x,i}^2 + \alpha_{y,i}u_{y,i}^2 + \beta_{x,i}v_{x,i}^2 + \beta_{y,i}v_{y,i}^2\right) \\ &+ \mu\sum_{i\in D}\left(\gamma_{x,i}m_{x,i}^2 + \gamma_{y,i}m_{y,i}^2 + \delta_{x,i}c_{x,i}^2 + \delta_{y,i}c_{y,i}^2\right) \end{aligned} \tag{4}$$

where w_i is the weight for the ith data constraint, $\alpha_{x,i}, \alpha_{y,i}, \beta_{x,i}, \beta_{y,i}, \gamma_{x,i}, \gamma_{y,i}, \delta_{x,i}, \delta_{y,i}$, are the weights for the corresponding components of the ith smooth constant along x- and y-directions. The weighted least squares energy function can improve robustness and accuracy in the optical flow estimation. This energy function is quadratic and convex for fixed w_i, thus allowing for applying the precondictioned conjugate gradient algorithm [29]. The weight w_i is determined based on the iterative re-weighted least square estimation algorithm in robust estimation, given by

$$w_i = \frac{2\sigma_r^2}{2\sigma_r^2 + r_i^2}, \quad r_i \triangleq \frac{I_{x,i}\widehat{u}_i + I_{y,i}\widehat{v}_i + I_{t,i} + \widehat{m}_i I_i + \widehat{c}_i}{\sqrt{I_{x,i}^2 + I_{y,i}^2 + I_i^2 + 1}} \tag{5}$$

where $[\widehat{u}_i, \widehat{v}_i, \widehat{m}_i, \widehat{c}_i]^T$ is the current estimated flow vector, and σ_r is selected as the standard deviation of the normalized constraint residue r_i. The other weights, including $\alpha_{x,i}$, $\alpha_{y,i}$, $\beta_{x,i}$, $\beta_{y,i}$, $\gamma_{x,i}$, $\gamma_{y,i}$, $\delta_{x,i}$, and $\delta_{y,i}$, are determined in a similar way with the corresponding residues replaced appropriately. For the details, please refer to [29].

2) Numerical Minimization Algorithm: Equation (4) can be rewritten in a matrix-vector form as $f(\mathbf{u}) = \mathbf{u}^T \boldsymbol{K}\mathbf{u} - 2\mathbf{u}^T\mathbf{b} + \mathbf{c}$. By setting the first order deviation to zero, minimizing this quadratic and convex function is equivalent to solving a large linear system $\boldsymbol{K}\mathbf{u} = \mathbf{b}$. The incomplete Cholesky preconditioned conjugate gradient (ICPCG) has been employed to solve this type of large and sparse linear system very efficiently [29]. The overall optical flow computation algorithm is summarized as follows:

1) Set w_i, $\alpha_{x,i}$, $\alpha_{y,i}$, $\beta_{x,i}$, $\beta_{y,i}$, $\gamma_{x,i}$, $\gamma_{y,i}$, $\delta_{x,i}$, and $\delta_{y,i}$ to 1; set $\mathbf{u}_0 = \mathbf{0}$; $k = 0$; $l = 0$.
2) Form the matrix $\boldsymbol{K}$ and vector $\mathbf{b}$.
3) Apply ICPCG by using $\mathbf{u}_0$ as the initial vector to obtain the solution $\widehat{\mathbf{u}}$.
4) Update w_i, $\alpha_{x,i}$, $\alpha_{y,i}$, $\beta_{x,i}$, $\beta_{y,i}$, $\gamma_{x,i}$, $\gamma_{y,i}$, $\delta_{x,i}$, and $\delta_{y,i}$; $k = k + 1$; $\mathbf{u}_0 = \widehat{\mathbf{u}}$.
5) If $k < N$, go to step 2).
6) If $l = M$, stop; else $k = 0$, $l = l + 1$, set $w_i = 1$, $\mathbf{u}_0 = \widehat{\mathbf{u}}$, go to step 2).

In this algorithm, $\mathbf{u}_0$ is the initial solution of the ICPCG algorithm. For fixed weights with $w_i, \alpha_{x,i}, \alpha_{y,i}, \beta_{x,i}, \beta_{y,i}, \gamma_{x,i}, \gamma_{y,i}, \delta_{x,i}$, and $\delta_{y,i}$, the corresponding energy function is quadratic and convex, thus the ICPCG algorithm converges to a unique solution no matter what the initial solution $\mathbf{u}_0$ is. However, a good initial solution can significantly reduce the total number of iterations required to converge to the final solution. For the problem on facial expression images, the initial value of $\mathbf{u}_0$ is computed by using radial basis function (RBF) interpolation from all the feature points. Kim *et al.* [14] also proposed a robust algorithm for motion estimation under illumination variations. Their method was formulated in a robust estimation framework and they employed the nonlinear gradient descent method to minimize the associated nonlinear energy function. The adopted robust optical flow estimation method [29] is formulated into a matrix-vector form and the solution can be solved efficiently by using the ICPCG algorithm. However, this method yields good results only when the displacements between consecutive images are small, though this problem can be relieved by adopting the coarse-to-fine strategy for solving the function minimization problems.

C. Constrained Optical Flow Computation

As mentioned in the previous section, we describe an ICPCG-based algorithm for accurate optical flow computation for all pixels in the image domain. However, the motion vectors of the facial feature points were used only as references for interpolating the initial optical flow $\mathbf{u}_0$ by using RBF interpolation. The ICPCG algorithm cannot guarantee the final converged optical flow to be consistent with the motion vectors at these feature points. Therefore, we include the hard constraints at these sparse locations in the same regularization framework and formulate it as a constrained optimization problem. Thus, the unconstrained optimization problem in the original formulation of the optical flow estimation is modified to a constrained optimization problem given as follows:

$$\begin{aligned} &minimize\ f(\mathbf{u}) = \mathbf{u}^T K\mathbf{u} - 2\mathbf{u}^T\mathbf{b} + \mathbf{c} \\ &subject\ to\ u(x_i, y_i) = \overline{u}_i,\ and\ v(x_i, y_i) = \overline{v}_i \\ &\qquad \forall (x_i, y_i) \in S \end{aligned} \tag{6}$$

where S is the set of feature points and $(\overline{u}_i, \overline{v}_i)$ 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 with an incomplete Cholesky preconditioner $\mathbf{P}$ to solve the above constrained optimization problem as follows:

1) Initialize $\mathbf{u}_0$ by RBF interpolation on the motion vectors $(\overline{u}_i, \overline{v}_i)$ at the facial feature points in S; set $k = 0$, $u(x_i, y_i) = \overline{u}_i$, and $v(x_i, y_i) = \overline{v}_i$, $\forall (x_i, y_i) \in S$; compute $\mathbf{r}_0 = b - \boldsymbol{K}\mathbf{u}_0$.
2) Solve $\boldsymbol{P}\mathbf{z}_k = \mathbf{r}_k$; $k = k + 1$. Set $\mathbf{z}_k(x_i, y_i) = 0$, $\forall (x_i, y_i) \in S$.
3) 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 $\mathbf{p}_k = \mathbf{z}_{k-1} + \beta_k \mathbf{p}_{k-1}$.
4) Compute $\alpha_k = \mathbf{r}_{k=1}^T \mathbf{z}_{k-1} / \mathbf{p}_k^T \boldsymbol{K}\mathbf{p}_k$.
5) Update $\mathbf{r}_k = \mathbf{r}_{k-1} - \alpha_k \boldsymbol{K}\mathbf{p}_k$, $\mathbf{u}_k = \mathbf{u}_{k-1} + \alpha_k \mathbf{p}_k$.
6) If $\mathbf{r}_k \approx \mathbf{0}$, stop; else go to step 2).

As summarized above, we reset the motion vectors at the feature points to be the specified flow vectors after initializing the initial guess for $\mathbf{u}_0$ in step 1). In turn, after solving $\boldsymbol{z}_k$ in step 2), the entries in the vector $\mathbf{z}$ corresponding to the facial feature pixels in S are then set to 0. By doing so, the motion vector values at the facial feature points will not be changed in the solution update step 3). Note that the combination of the feature point correspondence and the optical flow computation is accomplished in the proposed constrained optical flow estimation via solving the above constrained optimization problem.

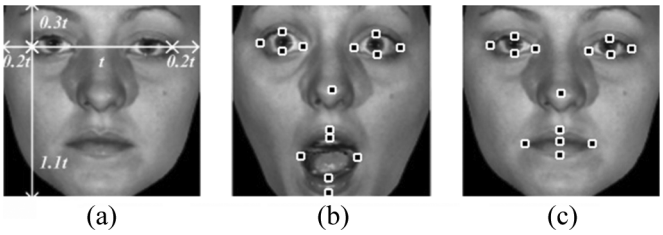


Fig. 1. (a) Face region selection. The 15 feature points on (b) a surprised face image and (c) a neutral face image.

III. Proposed Face Recognition System

As mentioned in Section II, the normalized correlation between an image $\boldsymbol{I}$ and another image $\boldsymbol{T}$ is equivalent to the Euclidean distance in the original image space $\|\boldsymbol{I}-\boldsymbol{T}\|$, if both images are normalized to zero mean and unit variance. An image in the test set can be classified by assigning to the label of the closest sample in the training set. However, the correlation or Euclidean distance is not well suited for the cases when there is facial expression variation between the training and testing images. We proposed a constrained optical flow estimation algorithm to alleviate the variation between expressive and neutral faces. Thus, the main purpose of this work is to achieve expression normalization on an expressive face image before identifying the face image.

A. Feature Selection and Face Alignment

The first procedure is to align the coordinates of images and define the facial feature points used in our constrained optical flow algorithm. We labeled 15 feature points, including four points for each eye, one at the nose tip, and the other six around the mouth region, as shown in Fig. 1(b) and (c). Even though automatic facial feature extraction methods have been well studied [17], all the points are manually labeled in our implementation to reduce the uncertainty of face alignment and feature selection. With the labeled points, the distance between the outer corners of both eyes is used as the reference to normalize face images (0.2, 0.2, 0.3 and 1.1 times to left, right, top, and bottom, respectively), as shown in Fig. 1(a).

B. Face Recognition Flowchart

After the image alignment and normalization of input testing image $\boldsymbol{T}$ and reference image $\boldsymbol{R}_c$, where $c = 1, 2, \ldots, C$, and C is the total number of subjects in the face database, we can further formulate the problem as follows:

$$\arg\min_{c} \|\boldsymbol{R}_c - Syn(\boldsymbol{T}; OF(\boldsymbol{T}; \boldsymbol{R}_c))\| \tag{7}$$

which is to determine the class c that minimizes the difference between the reference image $\boldsymbol{R}_c$ and the synthesized neutral image from the testing facial expression image $\boldsymbol{T}$ by using $\boldsymbol{R}_c$ as the reference. $OF(\boldsymbol{T}; \boldsymbol{R})$ operator is to apply the proposed constrained optical flow estimation algorithm to estimate the pixel motion vectors from one image $\boldsymbol{T}$ to a reference image $\boldsymbol{R}$, which is depicted in Fig. 2(a). Although the motion vectors and intensity variation coefficients are obtained in our optical flow estimation, only the motion information is used for the face synthesis. The function $Syn(\boldsymbol{T}; \mathbf{V})$ denotes the synthesis operator which warps the input image $\boldsymbol{T}$ to a new one through the motion vector $\mathbf{V}$, as depicted in Fig. 2(b). Note that this operator is a geometric warping determined by the optical flow and the brightness compensation is not used in the face synthesis. We further define the operation, $Syn(\boldsymbol{T}; OF(\boldsymbol{T}; \boldsymbol{R}_c))$, as $OF - Syn$ operator and symbolize it as depicted in Fig. 2(c). Under such procedure, the input image $\boldsymbol{T}$ can be normalized to the reference image of each class, and the variation between images due to expression can be reduced. We also define the subtraction operator as depicted in Fig. 3(a). The face regions are extracted from the images with a mask, as shown in Fig. 3(b), to eliminate the border regions. The mask is obtained from the intersection region of all the warped face images in the database to the global neutral face. The dashed dimension reduction block PCA is optional before subtraction and norm calculation ($\|\boldsymbol{x}\| \triangleq \sqrt{\boldsymbol{x}^T \cdot \boldsymbol{x}}$).

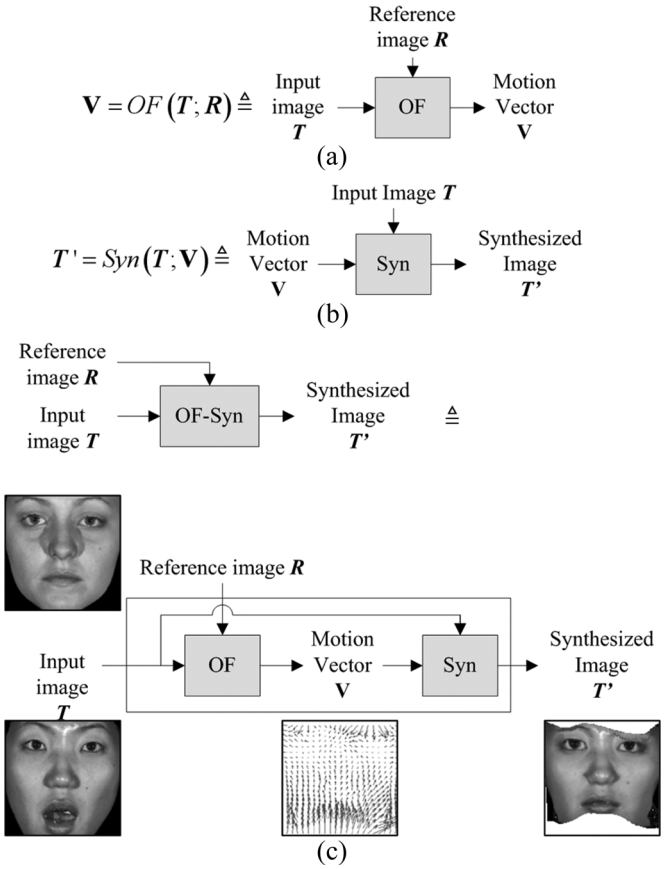


Fig. 2. Symbolizations of (a) OF, (b) Syn, and (c) $OF - Syn$ operators.

According to (7), we can design an expression-insensitive face recognition system as shown in Fig. 4, which comprises C OF-Syn operators and the nearest-neighbor classifier. In other words, C optical flow fields need to be computed with

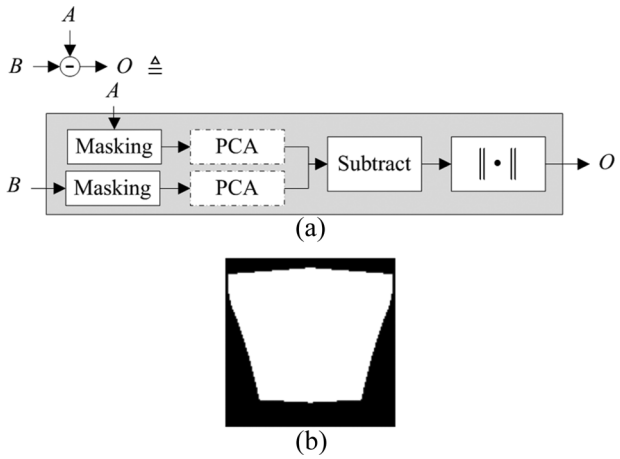


Fig. 3. (a) Flow chart of the subtraction operator and (b) the mask image.

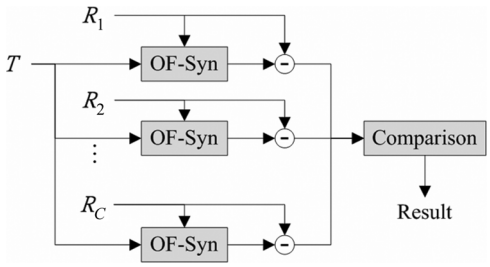


Fig. 4. Expression-invariant face recognition flow chart according to (7).

respect to the neutral images of all classes. It is very time-consuming and computationally intensive, especially when the total number of classes is large.

To further improve the computational efficiency, we modify the face recognition approach given in (7) as follows:

$$\arg\min_c \|Syn(\boldsymbol{R}_c; OF(\boldsymbol{R}_c; \boldsymbol{NE}_0)) - Syn(\boldsymbol{T}; OF(\boldsymbol{T}; \boldsymbol{NE}_0))\| \tag{8}$$

where $\boldsymbol{NE}_0$ is a standard neutral face image. This is based on the idea that if the synthesized image $\boldsymbol{T}'_c = Syn(\boldsymbol{T}; OF(\boldsymbol{T}; \boldsymbol{R}_c))$ is similar to $\boldsymbol{R}_c$, the synthesized images of $\boldsymbol{T}'_c$ and $\boldsymbol{R}_c$ to the same reference image, $\boldsymbol{NE}_0$, must be close to each other. Note that the operation of $Syn(\boldsymbol{T}'_c; OF(\boldsymbol{T}'_c; \boldsymbol{NE}_0))$ can be skipped, since we can find the synthesized result directly by applying $Syn(\boldsymbol{T}; OF(\boldsymbol{T}; \boldsymbol{NE}_0))$, as depicted in Fig. 5(a). Specifically speaking, instead of transforming the input image to the neutral image of each class, we now transform all images to a standard coordinate as $\boldsymbol{NE}_0$. The modified system flow chart is shown in Fig. 5(b). Although there are $C+1$ $OF-Syn$ operators in total, the C $OF-Syn$ operations among them can be performed in advance, thus only one optical flow computation between two face images is needed in the testing or recognition phase.

IV. Experimental Results

Our experiments were performed on the Binghamton University 3D Face Expression (BU-3DFE) Database [34]. The

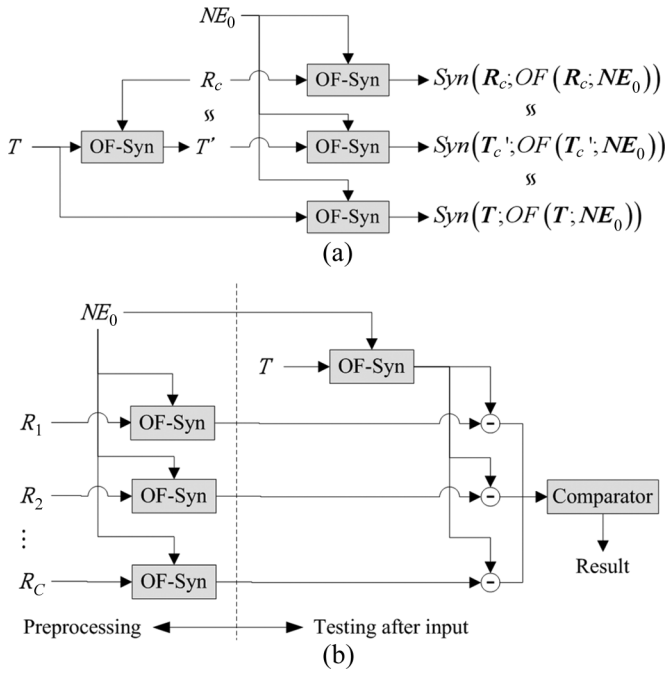


Fig. 5. (a) Expression normalization for all face images via optical flow warping with respect to a standard reference face image and (b) the proposed expression-invariant face recognition flow chart according to (8).

Fig. 6. Sample images in BU-3DFEDB. The left-top most is the neutral face. The others are the face images with angry, disgust, fear, happy, sad, and surprise expressions in columns from left to right with increasing levels in rows from top to bottom.

BU-3DFE database contains the face images and 3-D face models of 100 subjects (56 females and 44 males), each with a neutral face and six different expressions (happy, disgust, fear, happy, sad, and surprised) at different levels, from level-1 (weakest) to 4 (strongest), but only the 2-D face images were used in our experiments. All the images are normalized according to the procedures described in Section III-A and resized to 128 × 128 pixels. Fig. 6 shows the 25 normalized face images of one subject after the normalization procedure. The training database contains only neutral face images with only one neutral image for each class, i.e., 100 face images for 100 classes in total. There are 24 expression variant images for each subject, and we used totally 2400 images for all 100 subjects for testing.

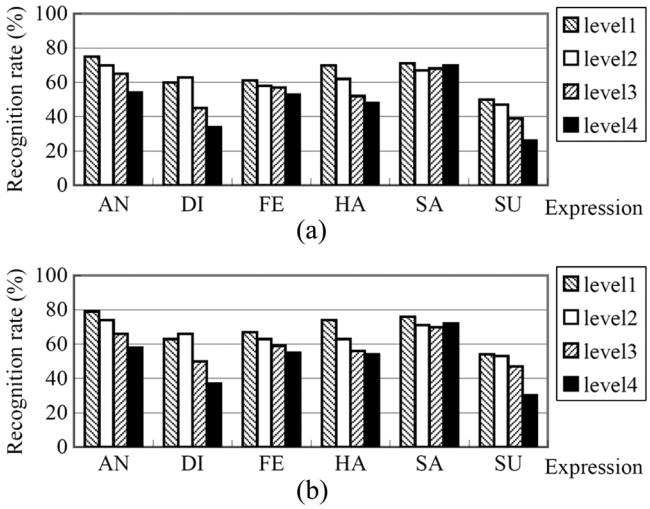


Fig. 7. (a) Recognition rates from the original images after PCA reduction under different expressions and levels (with average recognition rate 56.88%). (b) Recognition rates from the original images by direct subtraction under different expressions and levels (with average recognition rate 60.71%).

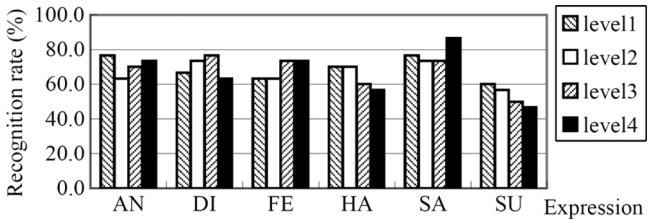


Fig. 8. Recognition rates from weighted optical flow result proposed by Martinez [21] (with average recognition rate 67.36%).

A. Benchmark Test

In the training phase for the original data, we first use PCA to compute the low-dimensional vector for all the 100 neutral images for all subjects. In the training phase, we use all the neutral images of all subjects in the training dataset for computing the PCA subspace by preserving 95% energy, which yields 51 eigenvectors. In the testing phase, the input image is projected to the PCA subspace and classified by the nearest neighbor classifier. The average recognition rate is 56.88%, used as a benchmark, and the recognition results are shown in Fig. 7(a) for all different expressions and levels separately. We can see that the surprise and disgust expressions give the worst results, and the result indicates that higher facial expression levels lead to worse recognition rates, as we discussed about the feasibility of using $||\boldsymbol{I}-\boldsymbol{T}||$ for face recognition before. Fig. 7(b) shows the result of direct subtraction without PCA preprocessing, which is slightly better than Fig. 7(a). We also implemented Martinez's method [21] with the optical flow obtained by our proposed method. In the experiment, we adopt the weighting of $w_i = MAX_F - ||F_i||$, where F_i is the magnitude of optical flow of each pixel and $MAX_F = \max_{\forall i} ||F_i||$. The average recognition rate is 67.36%, and the recognition results are shown in Fig. 8. Although the recognition rate is better, the time consumption is much longer since the optical flow computation of the test face image to all the face images in the training dataset is needed.

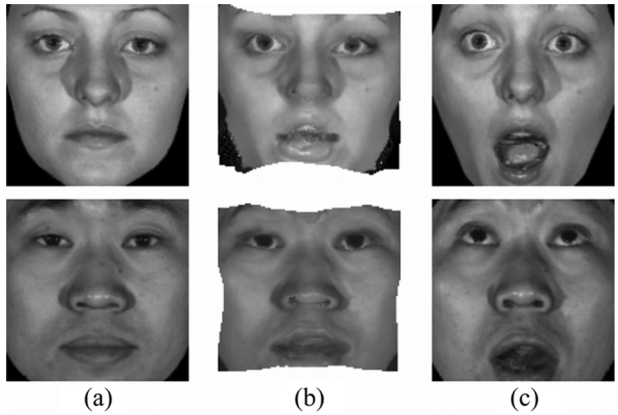


Fig. 9. Illustrations of expression normalization using constrained optical flow: (a) original neutral faces, (b) expression normalized faces with motion vector from (c) to (a), and (c) the face images with surprise expression at level 4.

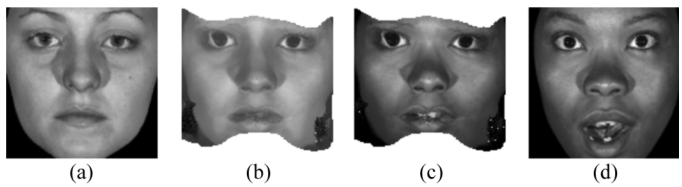


Fig. 10. Illustrations of constrained optical flow from one person to another: (a) referenced neutral face, (b) synthesized face from (d) to (a) with optical flow and brightness variation coefficients, (c) synthesized face from (d) to (a) with motion optical flow only, and (d) input image of surprising face at level 4.

B. Expression Normalization With Constrained Optical Flow Computation

In this experiment, we demonstrate the effectiveness of expression normalization by applying the proposed constrained optical flow computation through the face recognition accuracy. With the computed optical flow between the expressive and neutral faces of the same subject, we can generate a synthesized neutral face image as shown in Fig. 9. By using the proposed constrained optical flow estimation algorithm, we can obtained both motion vector and illumination coefficients simultaneously during the computation. Fig. 9(b) illustrates an example of applying the constrained optical flow for expression normalization with motion vector only. By using the above procedure, the synthesized neutral face looks similar to its real neutral face, except some blank regions. On the other hand, since intensity variation is also involved in the optical flow computation, the synthesized face image using both motion vector and illumination coefficients will look just like the person with neutral face, even the reference neutral face belongs to someone else, as shown in Fig. 10. In this case, the discrimination capability would be diminished. Therefore, as mentioned in Section III-B, only motion information in the optical flow estimation is used for neutral face image synthesis. Fig. 11 illustrates some expression-normalized images by using the original optical flow estimation and the proposed constrained optical flow algorithm, respectively. It is evident from Fig. 11 that the proposed constrained optical flow algorithm provides much better results in expression removal.

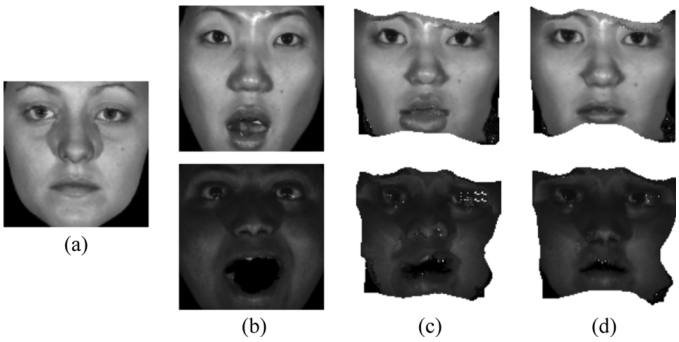


Fig. 11. Illustrations of expression normalization using optical flow warping with/without feature point constraints. (a) Reference face image, (b) input expression variant faces, (c) the de-expressed face images by using the warping determined by using the regularization-based optical flow estimation [29], and (d) the de-expressed face images by using the constrained optical flow warping.

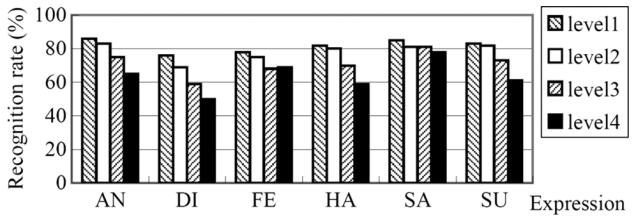


Fig. 12. Recognition rate from the synthesized neutral face image by using the estimated optical flow only, compared to synthesized neutral face by using the same method, under the flow as depicted in (7) and Fig. 5 (without PCA procedure). The average recognition rate is 73.67%.

C. Face Recognition With Expression Normalization

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. As shown in Fig. 10(c), there is some missing area in the synthesized image due to large motion, thus a mask shown in Fig. 3(b) is employed to restrict the area of image comparison. The recognition process is basically the same as that in the benchmark test, except that the blank regions in the synthesized face images within the mask are ignored in the image comparison.

We first examined the face recognition formulation of (7), which is to determine the class c that minimizes the difference between the reference image $\boldsymbol{R}_c$ and the synthesized neutral image from the testing facial expression image $\boldsymbol{T}$ by using $\boldsymbol{R}_c$ as the reference. The recognition results are shown in Fig. 12, with 73.67% average recognition rate. Although the result is better than the benchmark, the optical flow fields need to be computed with respect to the neutral images of all classes in the training dataset. It is time-consuming and computationally intensive.

We then modified the face recognition approach as depicted in Fig. 5(b) by using a neutral face image $\boldsymbol{NE}_0$ as the standard reference image in the optical flow computation. We first examined the face recognition system composed of the optical flow based expression normalization as shown in Fig. 2, the 500 face image comparison in Fig. 3(a) with the PCA block, and a standard reference image in the expression normalization process as shown in Fig. 5. Note that the standard neutral face image $\boldsymbol{NE}_0$ is conveniently selected as the neutral face of the first subject. Fig. 13

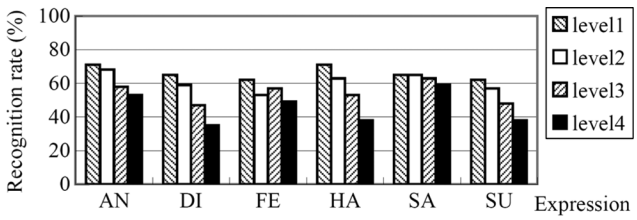


Fig. 13. Recognition rate from the synthesized neutral face image by using the estimated optical flow only, compared to synthesized neutral face by using the same method, by using PCA dimensional reduction with 95% energy preservation. The average recognition rate is 57.19%.

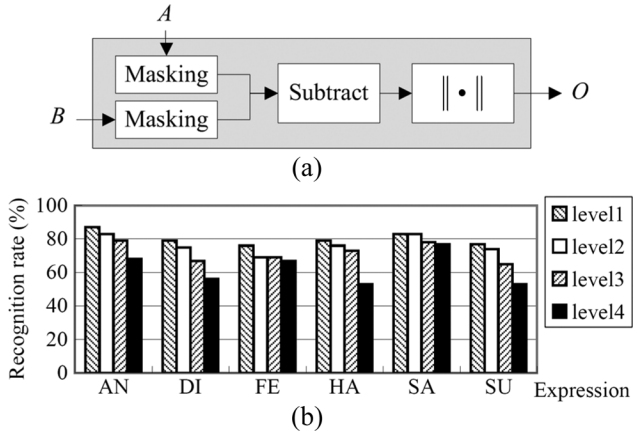


Fig. 14. (a) Modified function block from Fig. 4(a). (b) Recognition results from the synthesized neutral face images by using the constrained optical flow warping and direct differencing with a nearest neighbor classifier. The average recognition rate is 73.48%.

shows the recognition results for this face recognition system, which can provide only slightly better accuracy than the benchmark system. The PCA block in Fig. 3(a) is removed from the image comparison, as shown in Fig. 14(a), in the following experiment. For the face recognition system composed of the components in Figs. 2, 14(a), and 5, the recognition results by using direct differencing between partially masked synthesized images is depicted in Fig. 14(b). Obviously, the direct differencing scheme (with 73.48% average recognition rate) outperforms the PCA differencing scheme in our experiment. This means that the synthesized face images after the expression normalization process have very similar structure due to the same reference neutral image in the optical flow computation. A dimension reduction procedure may obliterate some delicate but significant information in the face images, thus degrading the recognition performance. Therefore, direct subtraction is more suitable here. The accuracy of the proposed face recognition system, shown in Fig. 14(b), is significantly superior to that of the benchmark system, with the recognition results given in Fig. 7, by 17% in average.

Under the same circumstance (using the first subject as $\boldsymbol{NE}_0$, the system composed of components in Figs. 2, 14(a), and 5), the recognition result using the regularization-based optical flow estimation [29] without hard constraints on facial feature points is shown in Fig. 15 with 67.14% average recognition rate. The result is consistent with the expression normalization result shown

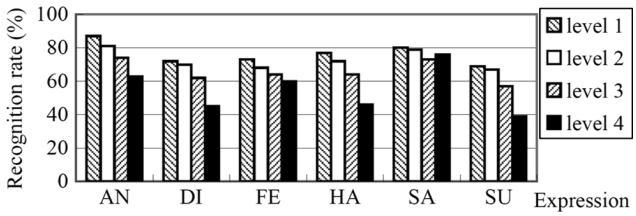


Fig. 15. Recognition results by using the regularization-based optical flow computation [29] without hard constraints on the facial feature points. The average recognition rate is 67.14%.

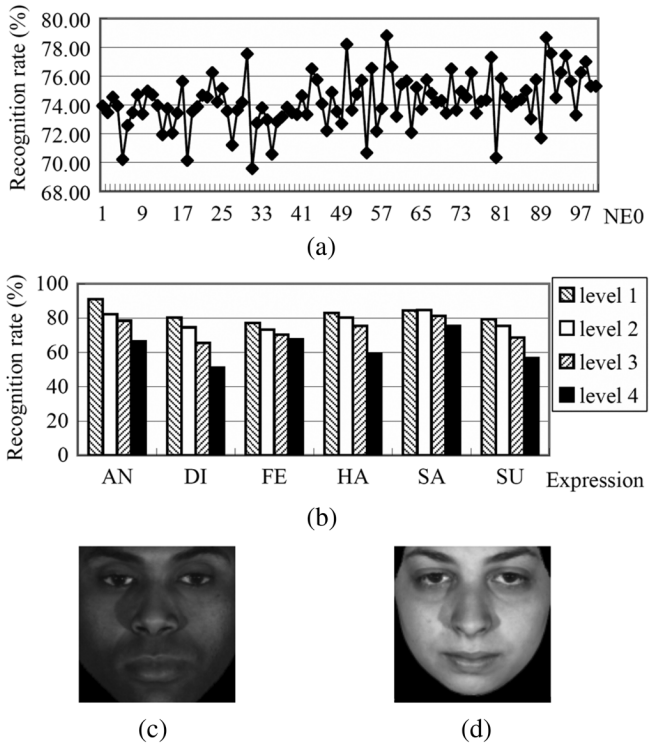


Fig. 16. (a) Recognition rates of using each the 100 subjects in the database as the standard reference image $\boldsymbol{NE}_0$. (The highest value is 78.79%, the lowest one is 69.58%, with 74.26% in average and 1.83% standard deviation.) (b) Average recognition rates of (a) in different expressions and different levels. (c) Reference neutral face with the best result, and (d) the reference neutral face of the worst result.

in Fig. 11. This justifies that the proposed constrained optical flow estimation provides better face recognition results.

Another issue is the selection of the standard reference neutral face image. Will different reference face image $\boldsymbol{NE}_0$ lead to significantly different result? We used each of all 100 subjects in the database as the reference neutral face to perform the face recognition experiment. In other words, the recognition experiment is repeated 100 times for all possible choices of reference neutral face, and the recognition results are shown in Fig. 16. We can see that the results are slightly vibrating with 74.26% accuracy in average and 1.83% standard deviation.

We also applied the traditional linear discriminant analysis (LDA) method with multiple training images and nearest neighbor classifier to the BU-3DFUDB database, and the results are summarized in Table I. It is evident that the face recognition rate with multiple training images for each subject is higher than that with only one training image. In addition, the recognition rates are further improved with the expression-normalized images obtained by the proposed algorithm.

TABLE I
AVERAGE RECOGNITION RATE USING LDA METHOD WITH DIFFERENT NUMBERS OF TRAINING SAMPLE. THE EXPERIMENT IS PERFORMED ON THE ORIGINAL BU-3DFUDB IMAGES AND THE EXPRESSION-NORMALIZED IMAGES BY USING THE PROPOSED METHOD, RESPECTIVELY

Training Number	LDA (Original)	LDA (Expression-normalized)
2	83.39	87.69
3	88.41	92.18
4	90.33	93.76
5	92.20	94.65

D. Illumination and Feature Point Disturbance

In this section, the performance of the proposed face recognition system under different illumination and feature point selection disturbances is examined.

For illumination variations, we first impose some synthetic lights in images by using Acrobat Photoshop, as shown in the 1st row of columns (b) and (c) of Fig. 17. The recognition results of the benchmark test (base 1) and the proposed method (case 1) are shown in Fig. 18. As the result shows, the recognition rates for both methods are dramatically reduced under different lighting conditions. To alleviate the problem due to illumination variations, we used a first-order polynomial $ax + by + c$ to approximate the current lighting condition from the face image, and then performed a lighting compensating procedure, as shown in the 3rd and 4th rows of Fig. 17. We can see that the recognition rates were improved after the lighting compensation (base 2 and case 2 in Fig. 18).

To simulate the inaccuracies in feature point extraction, we introduce a uniformly distributed random error, ranging from ± 1 pixel to ± 6 pixels in each direction for all feature points after the face alignment procedure we mentioned in Section III-A, and perform the face recognition experiments. The recognition rate is reduced with the increase of the facial feature error range, as shown in Fig. 19. With a small value of feature point error, e.g., ± 3 pixels, the result can still remain above 70%. However, by applying the extended active shape model feature detector [37] to automatically detect the facial feature points, we can only obtain an average recognition rate of 45%, for the proposed algorithm, as shown in Fig. 20. The degrade in the face recognition accuracy is because the off-the-shelf facial feature detector does not work well on face images with expressions, and large errors in feature point detection significantly influence the face alignment and the constrained optical flow computation. Thus, a robust facial feature detector would be essential to apply the proposed algorithm in practice.

V. CONCLUSION AND FUTURE WORKS

In this paper, we proposed a new algorithm for expression-invariant face recognition with one neutral face image per class in the training dataset. The basic idea is to combine the advantages of the feature point labeling in the model-based algorithms and the flexibility of the optical flow computation to estimate the geometric deformation for expressive face images. After the

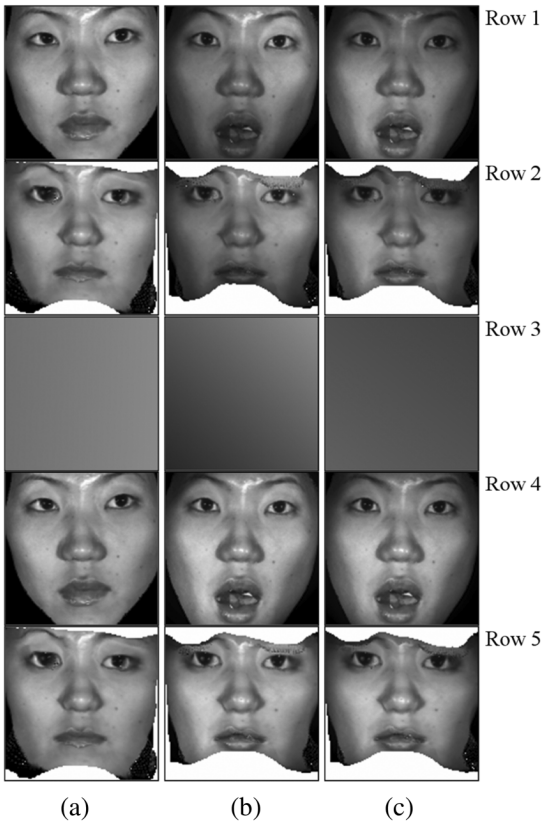


Fig. 17. Illustrations of synthesized illuminated images using Acrobat Photoshop and processing results. (a) Original neutral face, and (b) and (c) are two different synthesized illuminated images. 1st row: neutral image and input expressive image; 2nd row: expression normalized image from the 1st row; 3rd row: estimated lighting condition; 4th row: lighting compensated image of the 1st row; 5th row: expression normalized image from the 4th row.

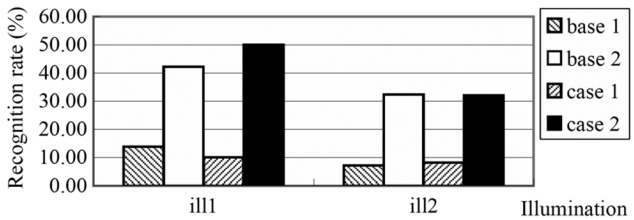


Fig. 18. Recognition rates under two different lighting conditions, denoted by ill1 and ill2 as depicted in Fig. 17, with and without lighting compensation by using the benchmark method [21] (base 1 and base 2) and the proposed algorithm (case 1 and case 2). Overall, the proposed method with lighting compensation can provide the best recognition results. Note that the recognition rates are computed from the averages of the face recognition rates under all different expressions. See the text for details.

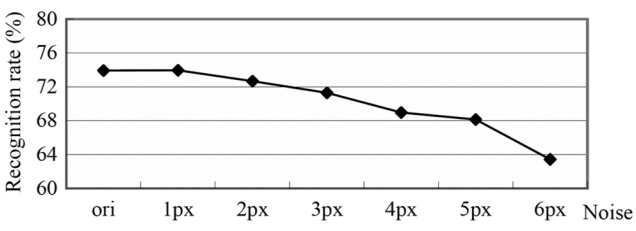


Fig. 19. Recognition rates under different levels of uniformly distributed random noises added to the feature points.

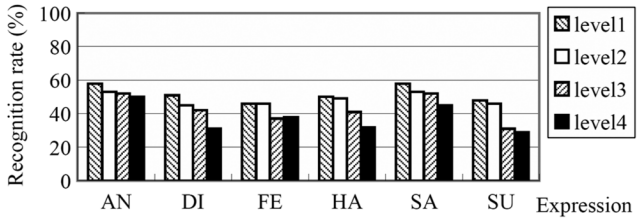


Fig. 20. Face recognition rates with an automatic feature detector [37]. The average recognition rate is dropped to 45.13%.

predefined feature points are extracted from face images, the optical flow for the expression variant face can be computed with the proposed constrained optical flow algorithm, which imposes the displacements at the facial feature points as hard constraints in the optical flow computation. Thus, the expression in the face image is normalized by warping with the estimated optical flow field. On the other hand, the computation time is also greatly reduced in our system by using a standard neutral face image in the optical flow computation and warping process. As shown in our experimental results, the proposed constrained optical flow warping algorithm significantly improves the recognition rate for face recognition from a single expressive face image when only one training neutral image for each subject is available.

The lighting condition is somehow restricted in our experiments. It is apparent that the optical flow computation may be influenced considerably when the illumination variation is large, which may influence the expression-normalized face image. The proposed method unified the geometry variation under different expressions and subjects, but the classification is based on the intensity comparison which assumes illumination similarity. In the future work, we will try to extend this work by considering the expression and lighting normalization altogether in a unified framework.

REFERENCES

[1] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis," *IEEE Trans. Neural Netw.*, vol. 13, no. 6, pp. 1450–1464, Nov. 2002.

[2] H. Cevikalp, M. Neamtu, M. Wilkes, and A. Barkana, "Discriminative common vectors for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 1, pp. 4–13, Jan. 2005.

[3] X. Chai, S. Shan, X. Chen, and W. Gao, "Local linear regression (LLR) for pose invariant face recognition," in *Proc. Int. Conf. Automatic Face and Gesture Recognition (FGR06)*.

[4] T. Cootes, C. Taylor, D. Cooper, and J. Graham, "Active shape models—Their training and application," *Comput. Vis. Image Understand.*, vol. 61, pp. 18–23, 1995.

[5] T. Cootes, G. J. Edwards, and C. Taylor, "Active appearance models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 681–685, Jun. 2001.

[6] I. A. Essa and A. Pentland, "A vision system for observing and extracting facial action parameters," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1994, pp. 76–83.

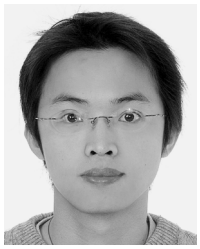
[7] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. New York: Academic, 1990.

[8] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under differing pose and lighting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 643–660, Jun. 2001.

[9] D. O. Gorodnichy, "Video-based framework for face recognition in video," in *Proc. Can. Conf. Computer and Robot Vision*, May 2005, pp. 330–338.

[10] Y. H. He, L. Zhao, and C. R. Zou, "Kernel discriminative common vectors for face recognition," in *Proc. Int. Conf. Machine Learning and Cybernetics*, Guangzhou, China, Aug. 18–21, 2005, pp. 4605–4610.

[11] B. K. P. Horn and B. G. Schunck, "Determine optical flow," *Artif. Intell.*, vol. 17, pp. 185–203, 1981.
[12] C. K. Hsieh and Y. C. Chen, "Kernel-based pose invariant face recognition," in *Proc. Int. Conf. Multimedia and Expo.*, Beijing, China, Jul. 2007.
[13] T.-K. Kim and J. Kittler, "Locally linear discriminant analysis for multimodally distributed classes for face recognition with a single model image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 318–327, Mar. 2005.
[14] Y.-H. Kim, A. M. Martinez, and A. C. Kak, "Robust motion estimation under varying illumination," *Image Vis. Comput.*, vol. 23, pp. 365–375, 2005.
[15] S. Z. Li and A. K. Jain, *Handbook of Face Recognition*. New York: Springer, 2005.
[16] X. Li, G. Mori, and H. Zhang, "Expression-invariant face recognition with expression classification," in *Proc. 3rd Can. Conf. Computer and Robot Vision*, Jun. 2006.
[17] S. C. Lin, M. J. Li, H. J. Zhang, and Q. S. Cheng, "Ranking prior likelihood distributions for Bayesian shape localization framework," in *Proc. Int. Conf. Computer Vision*, 2003.
[18] A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 228–233, Feb. 2001.
[19] A. M. Martinez, "Recognizing imprecisely localized, partially occluded and expression variant faces from a single sample per class," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 6, pp. 748–763, Jun. 2002.
[20] A. M. Martinez, "Matching expression variant faces," *Vis. Res.*, vol. 43, pp. 1047–1060, 2003.
[21] A. M. Martinez, "Recognizing expression variant faces from a single sample image per class," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Jun. 2003.
[22] S. Negahdaripour, "Revised definition of optical flow: Integration of radiometric and geometric cures for dynamic scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 9, pp. 961–979, Sep. 1998.
[23] U. Park, H. Chen, and A. Jain, "3D model-assisted face recognition in video," in *Proc. Can. Conf. Computer and Robot Vision*, May 2005, pp. 322–329.
[24] A. Rama and F. Tarres, "P2CA: A new face recognition scheme combining 2D and 3D information," in *Proc. Int. Conf. Image Processing*, Genoa, Italy, Sep. 2005.
[25] A. Rama and F. Tarres, "Partial LDA VS partial PCA," in *Proc. Int. Conf. Multimedia and Expo.*, Toronto, ON, Canada, Jul. 2006.
[26] M. Ramachandran, S. K. Zhou, D. Jhalani, and R. Chellappa, "A method for converting a smiling face to a neutral face with applications to face recognition," in *Proc. ICASSP*, Mar. 2005, pp. 18–23.
[27] R. Ramamoorthi and P. Hanrahan, "On the relationship between radiance and irradiance: Determining the illumination from images of a convex Lambertian object," *J. Opt. Soc. Amer.*, vol. 18, no. 10, pp. 2448–2459, 2001.
[28] A. Shashua and T. Riklin-Raviv, "The quotient image: Class-based re-rendering and recognition with varying illuminations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 129–139, Feb. 2001.
[29] C.-H. Teng, S.-H. Lai, Y.-S. Chen, and W.-H. Hsu, "Accurate optical flow computation under non-uniform brightness variations," *Comput. Vis. Image Understand.*, vol. 97, pp. 315–346, 2005.
[30] P.-H. Tsai and T. Jan, "Expression-invariant face recognition system using subspace model analysis," in *Proc. IEEE Conf. Systems, Man and Cybernetics*, Oct. 2005, vol. 2, pp. 1712–1717.
[31] M. A. Turk and A. P. Pentland, "Face recognition using Eigenfaces," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Maui, HI, Jun. 1991, pp. 586–591.
[32] Y. Yacoob and L. S. Davis, "Recognizing human facial expressions from long image sequences using optical flow," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 6, pp. 636–642, Jun. 1996.
[33] M.-H. Yang, "Kernel Eigenfaces vs. Kernel Fisherfaces: Face recognition using kernel methods," in *Proc. Int. Conf. Automatic Face and Gesture Recognition*, Washington, DC, May 2002, pp. 215–220.
[34] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3D facial expression database for facial behavior research," in *Proc. Int. Conf. Automatic Face and Gesture Recognition*, Apr. 2006, pp. 211–216.
[35] Y. Zhang and A. M. Martinez, "A weighted probabilistic approach to face recognition from multiple images and video sequences," *Image Vis. Comput.*, vol. 24, pp. 626–638, 2006.
[36] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, Dec. 2003.
[37] S. Milborrow and F. Nicolls, "Locating facial features with an extended active shape model," in *Proc. Eur. Conf. Computer Vision*, Marseille, France, 2008.



Chao-Kuei Hsieh received the B.S. degree in electrical engineering from the National Tsing Hua University, Hsinchu, Taiwan, in 2001. He is now pursuing the Ph.D. degree in the Department of Electrical Engineering, National Tsing Hua University.

His research interests include multimedia signal processing and pattern recognition.

Shang-Hong Lai (M'95) received the B.S. and M.S. degrees in electrical engineering from National Tsing Hua University, Hsinchu, Taiwan, and the Ph.D. degree in electrical and computer engineering from University of Florida, Gainesville, in 1986, 1988, and 1995, respectively.

He joined Siemens Corporate Research, Princeton, NJ, as a member of the technical staff in 1995. Since 1999, he became a faculty member in the Department of Computer Science, National Tsing Hua University, where he is currently a Professor. In 2004, he was a visiting scholar with Princeton University. His research interests include computer vision, visual computing, pattern recognition, medical imaging, and multimedia signal processing. He has authored more than 130 papers published in the related international journals and conferences. He holds ten U.S. patents for inventions related to computer vision and medical image analysis. He has been a member of program committee of several international conferences, including CVPR, ICCV, ECCV, ACCV, ICPR, and ICME.

Yung-Chang Chen (M'85–SM'90–F'05) received the B.S. and M.S. degrees in electrical engineering from the National Taiwan University, Taipei, Taiwan, in 1968 and 1970, respectively, and the Ph.D. (Dr.-Ing.) degree from the Technische Universitat Berlin, Berlin, Germany, in 1978.

In 1978, he joined the Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan. From 1980 to 1983, he was Chair of the Department of Electrical Engineering, National Central University, Chungli, Taiwan. From 1992 to 1994, he was Chair of the Department of Electrical Engineering, National Tsing Hua Univeristy. From 2002 to 2004, he was Dean of the College of Engineering and a Professor with the Department of Computer Science and Information Engineering, National Chung Cheng University, Chiayi, Taiwan. He is now a Professor with the Department of Electrical Engineering, National Hsing Hua University. His current research interests include multimedia signal processing, digital video processing, medical imaging, computer vision, and pattern recognition.

Dr. Chen serves as chair of the R.O.C. Image Processing and Pattern Recognition Society.