Design and Fusion of Pose-Invariant Face-Identification Experts

Tae-Kyun Kim and Josef Kittler, *Member, IEEE*

***Abstract*—We address the problem of pose-invariant face recognition based on a single model image. To cope with novel view face images, a model of the effect of pose changes on face appearance must be available. Face images at an arbitrary pose can be mapped to a reference pose by the model yielding view-invariant representation. Such a model typically relies on dense correspondences of different view face images, which are difficult to establish in practice. Errors in the correspondences seriously degrade the accuracy of any recognizer. Therefore, we assume only the minimal possible set of correspondences, given by the corresponding eye positions. We investigate a number of approaches to pose-invariant face recognition exploiting such a minimal set of facial features correspondences. Four different methods are proposed as pose-invariant face recognition "experts" and combined in a single framework of expert fusion. Each expert explicitly or implicitly realizes the three sequential functions jointly required to capture the nonlinear manifolds of face pose changes: representation, view transformation, and class discriminative feature extraction. Within this structure, the experts are designed for diversity. We compare a design in which the three stages are sequentially optimized with two methods which employ an overall single nonlinear function learnt from different view face images. We also propose an approach exploiting a three-dimensional face data. A lookup table storing facial feature correspondences between different pose images, found by 3-D face models, is constructed. The designed experts are different in their nature owing to different sources of information and architectures used. The proposed fusion architecture of the pose-invariant face experts achieves an impressive accuracy gain by virtue of the individual experts diversity. It is experimentally shown that the individual experts outperform the classical linear discriminant analysis (LDA) method on the XM2VTS face data set consisting of about 300 face classes. Further impressive performance gains are obtained by combining the outputs of the experts using different fusion strategies.**

***Index Terms*—Expert fusion, face image synthesis, linear discriminant analysis (LDA), multiple classifier system, pose-invariant face recognition.**

I. Introduction

FACE recognition has been competing with other biometric techniques such as fingerprint and iris recognition, with the understanding that it is less accurate but more user-friendly. The potential of face-recognition technologies is to identify humans without notice and at a distance. However, to realize this potential, it is essential to counteract the degradation in performance exhibited by face recognition systems for views different from the frontal pose. In the recent report [1] of the face-recognition vendor test (FRVT) 2002, which is the best known face competition, the verification rates (when the false accept rate (FAR) $= 0.01$) of all competing vendors for the faces which are rotated 45° right/left or 30° up/down, are less than about 45%. Most of them just achieved 20% or 30% verification accuracy for such a large rotation. This is a quite disappointing level of accuracy considering that all vendors showed more than 95% accuracy for the frontal faces. The pose data set has 87 identities without changes in illumination and expression. While it is true that, when more samples under different conditions are collected, better recognition performance is obtained, it is hard, or sometimes impossible, to acquire images of various views of the subjects, as in the case of passport photograph recognition.

In the recognition of faces imaged from nonfrontal views using a single model image, prior experience of the effect of view changes is essential. This can be obtained by learning from "prototype" or training faces, which can be either two-dimensional (2-D) or three-dimensional (3-D) data, and applied to new test faces to be recognized. Classically, a generic 3-D model of a human face has been used to synthesize face images from different view points [2], and approximate models, such as a cylinder, have also been applied to face recognition [3]. There are a number of methods [4]–[9] which have recently been developed to recognize a novel view face image based on statistical models acquired from prototype face images. Vetter and Poggio showed that new views can be synthesized from a single 2-D image in [4]. In their work, face images are first represented in a view subspace and the transformation matrices linking face images of different views are estimated. More recently, Blanz [7] utilized a 3-D morphable model and Li [6] applied kernel discriminant analysis and 3-D point distribution model for view-invariant face recognition. There are also other related studies [10]–[12], although they are in a slightly different field in that they require not a single but several model images for view-invariance modeling. In spite of the previous successes, all of the above methods have a strong drawback in requiring dense correspondence of facial features of different pose face images for image normalization. The step of feature detection or correspondence analysis, which is needed for separating the shape and texture components of face images in these methods, is usually difficult in itself [13], [14]. Errors in correspondences seriously degrade the face recognition accuracy of these methods, as shown in [7].

Among other relevant works, Graham and Allinson [15] applied a neural network for learning the view transfer function of the normalized face images with fixed eye positions. Talukder [16] also proposed a method for the eye registered face images, which involves finding a linear-view transfer matrix similarly to

Manuscript received December 20, 2003; revised November 30, 2005. This work was supported in part by the EU Project Biosecure and in part by the EPSRC Project 2D+3D=1D. This paper was recommended by Associate Editor E. Petajan.

T.-K. Kim is with the Department of Engineering, University of Cambridge, Cambridge, CB2 1PZ, U.K. (e-mail: tkk22@cam.ac.uk).

J. Kittler is with the Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, GU2 7XH, U.K. (e-mail: J.Kittler@surrey.ac.uk).

Digital Object Identifier 10.1109/TCSVT.2006.881197

1051-8215/$20.00 © 2006 IEEE

[4]. However, these studies lack a proper experimental validation. Moreover, the methods could further be improved by optimizing the three stages associated with the view transformation functions.

Discriminant analysis and multiple classifier fusion are two key elements of the proposed study. Linear discriminant analysis (LDA) is a well-known method [17], [18] of extracting class discriminative features of a piece of data. The method finds a linear transformation function which minimizes the scatter of data in the same class while maximizing the scatter of data in different classes. If the manifolds of image appearance changes are captured by the linear function from the training face classes, the function can be applied to new test face classes. However, the method fails in a nonlinear classification problem. To overcome the limitation of LDA, nonlinear frameworks of discriminant analysis have also been proposed in our former work [19], [20] as well as in the classical work [21]. Locally LDA (LLDA) exploits a set of piecewise linear functions [20] and generalized discriminant analysis (GDA) a kernel function [21] to learn nonlinear manifolds of data for the purpose of class-discriminative feature extraction.

Multiple classifier system is one of the subjects that have been intensively studied [22]. Generally, classifier selection and classifier fusion are the two types of combination [23]. While only one or a few local experts are nominated to make the decision in classifier selection, classifier fusion assumes that all classifiers are trained over the whole feature space, and thereby they are considered as complementary [24]–[27]. In the comparative studies of multiple expert fusion for personal identity verification, simple fusion strategies such as combination by averaging or majority vote were demonstrated to improve the verification performance [27]–[29]. Multiple classifier outputs are usually treated as classifier conditional posterior probabilities for the classes [30]. Under some assumptions, fusion often reduces to simple aggregation operators such as the product or average [27]. These simple fusion strategies do not require any additional training. Some trainable strategies such as decision templates [31] or the behavior knowledge space [26] method have also been proposed for person identity verification in [32]–[34].

In this paper, we propose methodology for the design and fusion of pose-invariant face identification experts based on the minimum information for face normalization. We assume that a single frontal image registered with reference to the eye positions is given as a model. We design four different types of robust experts for face identification at unknown views. Errors of the experts might be uncorrelated owing to dissimilarity of the sources of information and architectures used in the experts. A combining classifier is finally proposed for further accuracy improvement. All of the four experts can be approximately decomposed into the three basic sequential steps with the component functions: representation, view transformation, and class-wise discriminative feature extraction. Within this structure, the experts can be categorized into the methods based on statistical learning of face images captured at different views and the methods based on 3-D face models. In one of the four methods of learning, the three-stage process is empirically optimized by choosing the best combination of the component functions. Note that this three-stage process may be suboptimal as each step is separately trained. Two methods using nonlinear discriminant analysis techniques, which replace the three-stage functions, are also proposed. In addition, a computationally efficient approach based on a lookup table (LUT), which stores the correspondences of different pose images found by 3-D face models, is developed for complementing the methods based on statistical learning. All of these experts are quite different in their nature owing to different sources of information and architectures used. Thus, their fusion promises performance improvement. Importantly, the expert fusion exhibits quasi-monotonic behavior as the number of combined experts increases.

Section II briefly reviews conventional LDA as a core method for constructing the proposed experts. The basic structure of the proposed experts and the overall expert fusion architecture for pose-invariant face identification is given in Section III. Section IV is devoted to the detailed descriptions of the individual expert design. In Section V, the motivation for and strategies of expert fusion are presented. The experimental results and conclusions are drawn in Sections VI and VII, respectively.

II. LDA

LDA is a class-specific method in the sense that it represents data to make it useful for classification [17], [18]. Let $X = \{x_1, x_2, \dots x_M\}$ be a data set of given N-dimensional vectors of images. Each data point belongs to one of C object classes $\{X_1, \dots X_c, \dots X_C\}$. The between-class scatter matrix and the within-class scatter matrix are defined as

$$B = \sum_{c=1}^{C} M_c (m_c - m)(m_c - m)^T$$
$$W = \sum_{c=1}^{C} \sum_{x \in X_c} (x - m_c)(x - m_c)^T \qquad (1)$$

where m_c denotes the class mean and m is the global mean of the entire sample set X. The number of vectors in class X_c is denoted by M_c. LDA finds a matrix U maximizing the ratio of the determinant of the between-class scatter matrix to the determinant of the within-class scatter matrix as

$$U_{\text{opt}} = \max_{\arg U} \frac{|U^T B U|}{|U^T W U|} = [u_1, u_2, \dots, u_N]. \qquad (2)$$

The solution $\{u_i \mid i = 1, 2, \dots, N\}$ is a set of generalized eigenvectors of B and W, i.e., $Bu_i = \lambda_i W u_i$. Usually, principle component analysis (PCA) is performed first to avoid a singularity of the within-class scatter matrix commonly encountered in face recognition [18].

III. Structure of Experts and Expert Fusion

The fact that face data distribution of different poses is highly nonlinear motivates us to exploit the benefits of nonlinear architectures for the experts. Even though the proposed experts have different architectures and component functions, it is convenient to explain them by the proposed basis structure comprising the three sequential steps as shown in Fig. 1(a): representation, view transformation, and discriminative feature extraction.

First, a pair of input face images (frontal and rotated face image) is represented as low-dimensional feature vectors. The

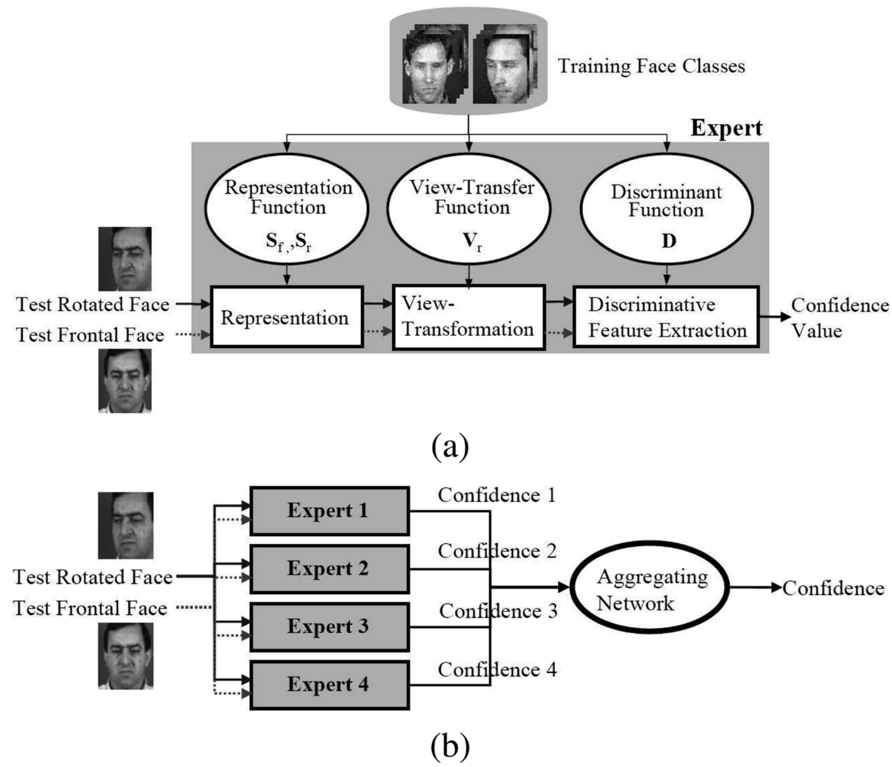


Fig. 1. (a) Basic component functions of the proposed view-invariant face recognition experts. (b) Overall structure of the multiple expert fusion system.

dimensionality of the vectors obtained by raster-scanning the original images is often so high that the subsequent component functions cannot be efficiently learned. A dimensionality reduction is performed based on statistical models of images. As different pose images have considerably different statistical characteristics, it is beneficial to exploit multiple representation functions, each of which covers a certain range of views. Using a view-specific representation function S, feature vector b of an image x is given as

$$\begin{aligned} b_{f,i} &= S_f(x_{f,i}) \\ b_{r,i} &= S_r(x_{r,i}) \end{aligned} \tag{3}$$

where f and r denote images within a certain range of frontal views and rotated views, respectively. Index i represents the ith face class.

In the view-transformation stage, different view images of the same classes are forced to have a similar representation. This is achieved by finding the respective transfer functions V_r defined as

$$\min_{\arg V_r} \sum_{i=1}^{C} |b_{f,i} - V_r(b_{r,i})|^2 \tag{4}$$

where C is the total number of the training face classes. The view-transformation function can be bi-directional, that is, the function can transform a frontal face to a rotated face or the rotated face to the frontal face. In the basic model proposed, the transformation to the frontal pose is adopted with the two benefits: low memory to store the representation vectors of all model face images and time-efficient recognition of a new test image. In this approach, the same number of feature vectors as the number of model face images are stored. Further, only a single view transformation of a test image of an arbitrary pose to the frontal pose is needed to compare it with all of the frontal model faces stored. Note that the transfer function is also view-specifically learned as denoted by V_r in (4).

The transformed feature vectors of all rotated images and the feature vectors of frontal images in the prototype set are the input for learning a class-discriminant function D. The discriminant function maximizes the class separability of the training data with the discriminative feature vectors d defined as

$$\begin{aligned} d_{f,i} &= D(b_{f,i}) \\ d_{r,i} &= D(V_r(b_{r,i})). \end{aligned} \tag{5}$$

With an analogy to LDA, the discriminant function D can be learned from the prototype face classes and applied to extract efficient discriminative features from face images of the new test classes. The expert finally produces a similarity value between a frontal and a rotated face image, based on a distance of the discriminative feature vectors d of the two images.

The overall structure of the proposed fusion of the experts is shown in Fig. 1(b). The four experts designed to enhance their diversity are aggregated taking their confidences into account. The multiple expert outputs are normalized to be comparable and treated as expert-conditional class posterior probabilities. As the purpose of this study is to examine the potential for improving the recognition accuracy by fusing different types of pose-invariant face experts, simple fixed fusion rules such as the product and sum are employed as the aggregating operator. Under the assumption of independence of the expert outputs, the fusion by product is optimal. Even though such an assumption may be quite strong, it has been noted that these simple operators usually work well.

IV. Expert Design

A. Three-Stage Linear Function Expert: PCA-Linear-Matrix (LM)-LDA

As the performance of the experts depends on the choice of the component functions for the proposed three-stage structure presented in the previous section, various combinations of linear and nonlinear functions obtained by statistical learning have been compared in our former report [35]. The study showed that the piecewise-linear combinatorial method PCA (as a function S)-LM (as a function V)- LDA (as a function D) is one of the most accurate classifiers. As it is also attractive in terms of computational costs, PCA-LM-LDA has been adopted as an expert in this study. In this approach, the three steps are separately optimized.

An eigen-subspace is chosen as a representation function among the various linear and nonlinear dimensionality reduction methods [36]–[41]. The eigen-subspace is constructed by PCA of the covariance matrix of the prototype face images resulting in a number of image-size eigenvectors [37]. $P_M = [p_1, \ldots, p_M]$ is the matrix containing the M eigenvectors corresponding to the M largest eigenvalues of the covariance matrix. Face images are represented as

$$\begin{aligned} b_{f,i} &= P_{f,M}^T (x_{f,i} - m_f) \\ b_{r,i} &= P_{r,M}^T (x_{r,i} - m_r) \end{aligned} \tag{6}$$

where $P_{f,M}$ and $P_{r,M}$ are the eigenvector matrices learned from a set of frontal face images and rotated face images of all classes, respectively, m_f and m_r denote the means of the frontal and rotated face images, respectively, and i denotes the ith face class.

Let $B_f = \{b_{f,1}, b_{f,2}, \ldots, b_{f,C}\} \in \mathcal{R}^{M\times C}$, $B_r = \{b_{r,1}, b_{r,2}, \ldots, b_{r,C}\} \in \mathcal{R}^{M\times C}$. M is the number of the eigenvectors used, thus being the dimensionality of the feature vectors b. C is the total number of the face classes. The $M \times M$ linear matrix L_r is defined as a view-transformation function such that $B_f \cong L_r B_r$. The element B_{fij}, ith element of the jth feature vector $b_{f,j}$ is represented as

$$B_{fij} = L_{ri1} B_{r1j} + \cdots + L_{riM} B_{rMj}. \tag{7}$$

This gives us C equations to solve for the M unknown parameters $L_{ri1}, \ldots, L_{riM}$. Similarly, for the full matrix, we have a linear regression problem which determines the $M \times M$ unknown parameters from the $C \times M$ given equations. The linear matrix can be calculated by

$$L_r = B_f B_r^T \left(B_r B_r^T\right)^{-1} \tag{8}$$

and the virtual frontal images generated by the view-transformed matrix are the columns of the matrix

$$P_{f,M}(L_r B_r) + m_f. \tag{9}$$

All nonfrontal images of the prototype face classes are transformed to the frontal version by (9). The virtual frontal images are the input for learning a discriminative function with the original frontal images of the prototype faces. LDA is applied as a learning method to minimize the volume of the same class faces and maximize the volume of the different class faces [18]. To avoid any problems arising from the within-class scatter matrix being singular, the feature vectors $b_{f,i}, L_r b_{r,i}, i = 1, \ldots, C$, are used as the inputs of LDA instead of images. Let the LDA transformation matrix learned be U and the global mean of the entire feature vectors be m. The final class discriminative vectors $d_{f,i}$ and $d_{r,i}$ of two face images $x_{f,i}$ and $x_{r,i}$ are given by

$$\begin{aligned} d_{f,i} &= U^T(b_{f,i} - m) = U^T\left(P_{f,M}^T(x_{f,i} - m_f) - m\right) \\ &= \left[U^T P_{f,M}^T\right] x_{f,i} + \left[-U^T P_{f,M}^T m_f - U^T m\right] \\ d_{r,i} &= U^T(L_r b_{r,i} - m) = U^T\left(P_{r,M}^T(x_{r,i} - m_r) - m\right) \\ &= \left[U^T L_r P_{r,M}^T\right] x_{r,i} + \left[-U^T L_r P_{r,M}^T m_r - U^T m\right]. \end{aligned} \tag{10}$$

The confidence value of the two face images to be in the same class is given as a reciprocal of the Euclidean distance $||d_{f,i} - d_{r,i}||$.

B. Monolithic Nonlinear Experts

1) LLDA: The method of discriminant analysis, LLDA [20], was proposed as a technique for designing pose-invariant face experts. It has a nonlinear optimization framework realized by a piecewise linear structure for the extraction of discriminative features from nonlinearly separated data. The single nonlinear structure of LLDA can be considered to replace the three-stage process of the PCA-LM-LDA method in the previous section. The LLDA method has benefits in terms of efficient and optimal learning by the single nonlinear optimization process compared with the PCA-LM-LDA, where each of the three steps is sequentially trained. Compared with the conventional nonlinear method based on kernel functions such as GDA, which will be also presented as an expert in this study, the method has the benefits of avoiding overfitting and low computational cost owing to its piecewise linear structure as shown in [19].

The LLDA method concurrently finds the set of locally linear functions, each of which is specific to a subset of input images. The input images are clustered into K subsets $k = 1, \ldots, K$. Each subset k represents a pose group (a set of face images belonging to a certain range of views) in this study, which has a different transformation function applied. Each face image belongs to the kth pose subset with a posterior probability $P(k|x)$. In the experiments in Section VI, the posterior probability will be simply hardened with a given pose label of face images, that is, if the face image x belongs to the k^*th pose group, then $P(k^*|x) = 1$ and $P(k|x) = 0$ for all of the other k's. The locally linear transformation $U_k = [u_{k1}, u_{k2}, \ldots, u_{kN}]$, $k = 1, \ldots, K$, is defined such that

$$d = \sum_{k=1}^{K} P(k|x) U_k^T (x - m_k) \tag{11}$$

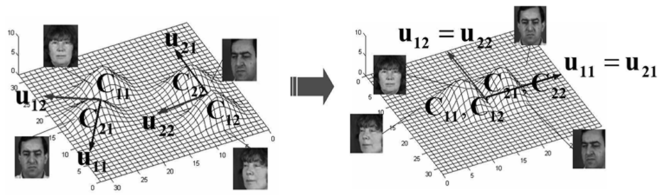


Fig. 2. LLDA for pose-invariant face classification: Left shows the original class data distributions of the two different pose groups and the pose-specific components found. The transformed data distributions are shown on the right. C_{ij} is the j-th pose cluster of the i-th class, u_{ij} is the j-th component of the i-th pose cluster. The pose-wise clustered face images in the original data space become class-wise clustered in the transformed space.

where m_k is the mean of the images in the kth pose group and d is the discriminative feature vector of an image x. The matrices U_k containing locally linear transformations are concurrently found so as to maximize the between-class covariance while minimizing the within-class covariance by analogy to the optimization concept of LDA explained in Section II. The concept of LLDA is illustrated in Fig. 2. Different view face images are class-wise clustered in the transformed space by the two sets of pose-specific transformations learned. Refer to previous work [19], [20] for the details of learning.

Let $x_{f,i}$ and $x_{r,i}$ be a frontal image and a rotated image of the ith face class, respectively, and suppose that the corresponding pose-specific linear transformation matrices of LLDA are U_f and U_r. The final output vectors d of the LLDA method are given as

$$\begin{aligned} d_{f,i} &= U_f^T(x_{f,i} - m_f) \\ d_{r,i} &= U_r^T(x_{r,i} - m_r) \end{aligned} \tag{12}$$

where m_f and m_r denote the means of the frontal and rotated images of the prototype faces, respectively. The confidence value of the two face images to be in the same class is given as a reciprocal of the Euclidean distance $\|d_{f,i} - d_{r,i}\|$.

Note that the locally linear transformations in (12) correspond to the combinations of the three-stage functions of the PCA-LM-LDA expert in (10) as

$$\begin{aligned} U_f &\equiv P_{f,M}U \\ U_r &\equiv P_{r,M}L_r^T U \end{aligned} \tag{13}$$

because the term $U^T m$ in (10) is common for all face images and thus can be eliminated in image comparison. The functions sequentially trained in the PCA-LM-LDA expert are efficiently represented by the functions learned in the single optimization framework of LLDA. The results of an experimental comparison of the experts, LLDA and PCA-LM-LDA, will be presented in Section VI.

2) GDA: The GDA [21] is a method designed for nonlinear classification based on a kernel function Φ which transforms the original space X to a new high-dimensional feature space Z such that $\Phi : X \rightarrow Z$. After transforming a data point into high-dimensional space, a linear classification function is sought in the transformed space similarly to LDA. In this study, the method is applied to the different pose face images of the prototype set, resulting in a single nonlinear transformation to extract class-discrimination features of the test face images.

The projection of a face image x is computed by the projection vectors $u_\Phi \in Z$ as

$$(u_\Phi)^T \Phi(x) = \sum_{c=1}^{C} \sum_{i=1}^{M_c} \alpha_{c,i} k(x_{c,i}, x) \tag{14}$$

where $\alpha_{c,i}$ are real weights learned during training and $x_{c,i}$ is the ith prototype face from class c. k is a kernel function. Here, an RBF kernel with an adjustable width was deployed to cope with nonlinear manifolds of multiview face data. For the details of learning the weights $\alpha_{c,i}$, refer to [21].

The major difference of the GDA expert from the previous three-stage linear method PCA-LM-LDA is again in that a single nonlinear transformation function U_Φ is applied to different view face images, $x_{f,i}$ and $x_{r,i}$, as

$$\begin{aligned} d_{f,i} &= U_\Phi^T \Phi(x_{f,i}) \\ d_{r,i} &= U_\Phi^T \Phi(x_{r,i}) \end{aligned} \tag{15}$$

where $d_{f,i}$ and $d_{r,i}$ are the discriminative feature vectors of $x_{f,i}$ and $x_{r,i}$, respectively. While different overall linear functions are applied to different view faces in both PCA-LM-LDA and LLDA experts [that is, the resulting overall functions are view-specific as shown in (13)], a single transformation function U_Φ is commonly applied to different view faces $x_{f,i}$ and $x_{r,i}$ by the GDA expert. The kernel function of GDA is flexible enough to capture highly nonlinear manifolds of the prototype face data so that the transformed prototype data is well class-wise clustered. To avoid overfitting of the GDA expert on the prototype set, a proper independent face data set has been exploited for evaluating the kernel parameters (see Section VI).

The confidence value of the expert is again inversely proportional to the Euclidean distance $\|d_{f,i} - d_{r,i}\|$.

C. Expert With a Pose Correction by 3-D Correspondence LUT

The three methods explained in the previous sections are purely based on statistical learning of images. They will be shown to be effective for capturing face view changes but a complementary benefit of using 3-D face models has also been investigated for further accuracy improvement. We propose a novel view-transformation of face images based on 3-D correspondence LUT (3D-LUT) which effectively replaces the representation and view-transformation function of the basic three-stage structure of the proposed pose-invariant experts. Conventional LDA is then applied to the images transformed by the 3D-LUT for discriminative feature extraction.

The method is motivated by the fact that, once a face image is texture mapped on a generic 3-D face model, images of arbitrary views can be synthesized. In our approach, the normal procedure of texture mapping, 3-D rotation, and rendering in computer graphics is replaced by a direct image transformation expressed in terms of the correspondence LUT, as illustrated in Fig. 3. The facial feature correspondences between different view images are found by using 3-D face models and stored in

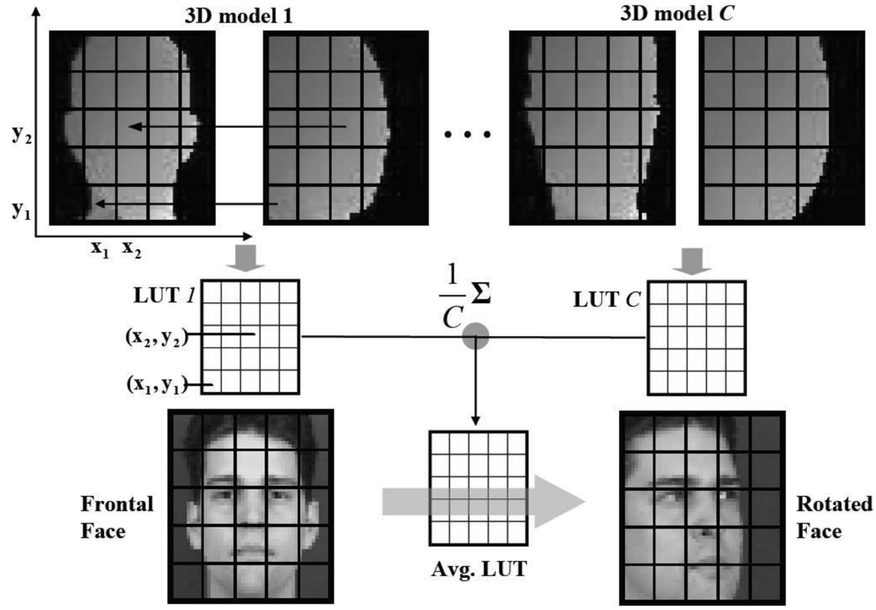


Fig. 3. Virtual view generation by using the 3D-LUT.

a general correspondence LUT. The view-transformation by the LUT is extremely efficient in time.

During the training, the parameters of 3-D rotation of face models and texture mapping are adjusted so that output face images at different views are normalized with reference to a fixed eye position. The image-sized LUT contains 2-D coordinates as an element which describes the correspondence of pixels of the frontal and rotated view face images as

$$(\overline{x}, \overline{y}) = \mathrm{LUT}_{r,i}(x, y) \tag{16}$$

where (x, y) is a point of the rotated view image of the ith face class and $(\overline{x}, \overline{y})$ is the corresponding point of the frontal view image. Such correspondences were sought by picking the same color pixels of the projected frontal and rotated images of the synthetic color-textured 3-D face models as shown in Fig. 3. The generic view-transformation function is simply constructed by averaging the correspondences of all face classes by

$$(\overline{x}, \overline{y}) = \overline{\mathrm{LUT}}_r(x, y) = \frac{1}{C}\sum_{i=1}^{C}\mathrm{LUT}_{r,i}(x, y). \tag{17}$$

where C is the number of the face classes. By using the average LUT, rotated face images are virtually generated from frontal face images. Pixel values of a virtual view face image $\overline{I}_f$ are obtained from those of the corresponding pixels of the frontal image I_f as

$$\overline{I}_f(x, y) = I_f(\overline{x}, \overline{y}) = I_f(\overline{\mathrm{LUT}}_r(x, y)). \tag{18}$$

The rotational direction from the frontal to an arbitrary angle is more beneficial in this method as more pixel information is kept in the frontal face images. Each pose group has an average correspondence LUT $\overline{\mathrm{LUT}}_r$. After transforming all frontal faces in the prototype data set to a rotated view r, LDA is applied to the pairs of the transformed images $\overline{I}_f$ and the original rotated images I_r of all classes. Let $\overline{x}_f$ and x_r denote vector representations of the images $\overline{I}_f$ and I_r, respectively. LDA is performed on the vectors $\overline{x}_f$ and x_r to learn a class-discrimination function, and final discriminative feature vectors d are obtained by

$$\begin{aligned} d_{f,i} &= U^T(\overline{x}_{f,i} - m) \\ d_{r,i} &= U^T(x_{r,i} - m) \end{aligned} \tag{19}$$

where U is the solution matrix of LDA and m is the mean vector of the entire vectors $\overline{x}_{f,i}$ and $x_{r,i}, i = 1, \ldots, C$. C is the number of the prototype face classes. The output of the expert is again a reciprocal of the Euclidean distance of the discriminative feature vectors $d_{f,i}$ and $d_{r,i}$.

V. Expert Fusion: Motivation and Strategies

Four different approaches to pose-invariant face identification have been proposed in the previous sections. PCA-LM-LDA, LLDA and GDA are based on statistical learning of 2-D appearance images of faces and the 3D-LUT method exploits 3-D facial models. PCA-LM-LDA and 3D-LUT explicitly generate virtual-view face images and exploit them for learning discriminative features, whereas LLDA and GDA compute view-robust representations of face input images. Some examples of the virtual-view face images obtained by PCA-LM-LDA and 3D-LUT are shown in Fig. 4. The characteristics of the two results are considerably different, and the errors in the transformations are hopefully uncorrelated. The errors in the view transformation generated by PCA-LM-LDA come from the blur of the transformed faces to the frontal views. The blurred images lose high-frequency components of the original face shapes. On the other hand, the rotation by 3D-LUT loses shape information mainly around face boundaries as shown in Fig. 4(c). While the generalization performance of the transformation of the PCA-LM-LDA method is degraded for nontrained face classes,

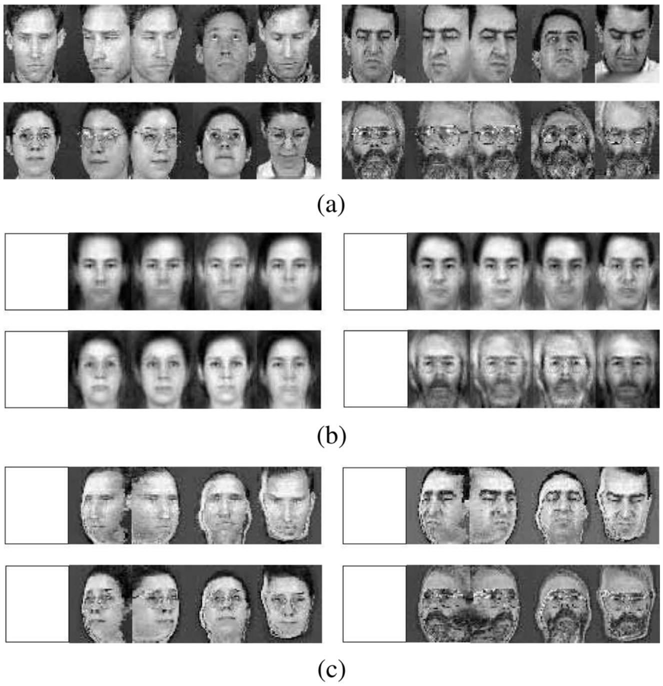


Fig. 4. Examples of the synthesized images of four face classes by the two methods. (a) Original five pose images. (b) Transformed face images to a frontal view by PCA-LM-LDA. (c) Transformed frontal face images to a rotated view by 3D-LUT.

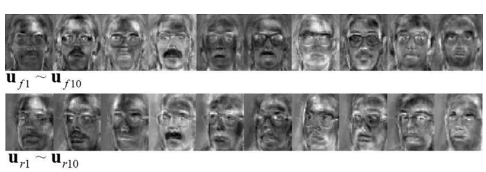


Fig. 5. Transformation functions of LLDA. Each row shows the local pose functions of LLDA. The corresponding local functions of the two pose groups, e.g., $u_{f,i}$ and $u_{r,i}$, are characterized by a rotation, ensuring view-invariant representation.

the rotation by 3D-LUT relatively well maintains its generalization performance. LLDA and GDA implicitly represent face images by similar representation vectors regardless of face poses. Fig. 5 shows some example transformation functions of LLDA. The local functions of LLDA, which are specific to a certain pose group, seem to provide discriminant features for different faces at the same view as the conventional LDA. Moreover, the method is likely to produce consistent features for the same face classes regardless of view angles, because all corresponding functions of the two different pose groups, for example, $u_{f,1}$ and $u_{r,1}$, are characterized by an appropriate rotation (see each pair of $u_{f,i}$ and $u_{r,i}$ in Fig. 5). On the contrary, GDA first maps images to high-dimensional vectors and deploys a common global function to extract effective features for classification. The GDA method possesses different characteristics from LLDA in terms of decision boundaries and, consequently, classification error probabilities. We experimentally found that all of the experts enhanced the performance of the conventional well-known method, PCA-LDA, which is also known as "fisherface" [18], and contributed to the

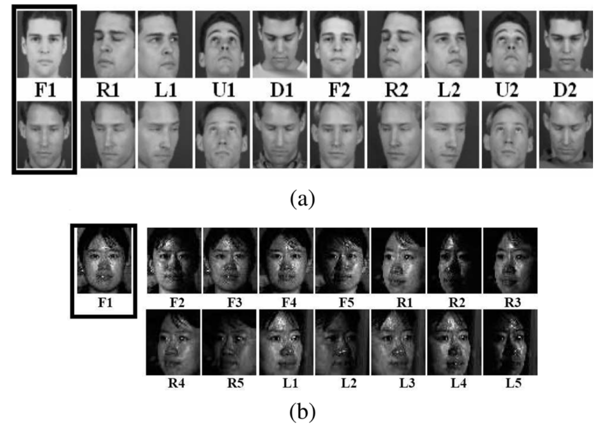


Fig. 6. Normalized data samples of (a) XM2VTS DB and (b) PIE DB.

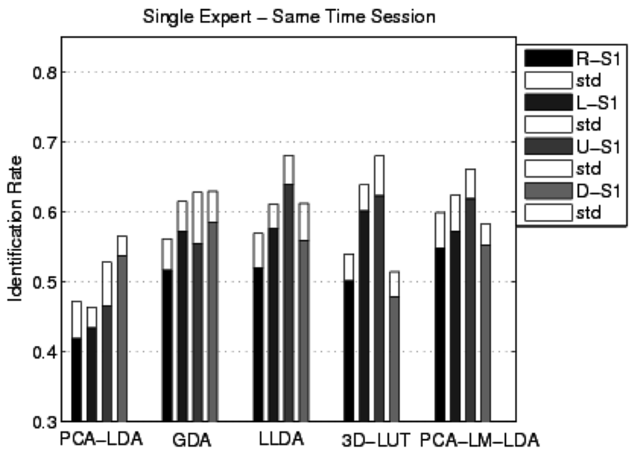
TABLE I
IDENTIFICATION RATES OF INDIVIDUAL EXPERTS. TENFOLD CROSS VALIDATION WAS PERFORMED. AVERAGE AND STANDARD DEVIATION OF THE METHODS ARE REPORTED

	PCA-LDA	GDA	LLDA	3D-LUT	PCA-LM-LDA
Mean	0.3566	0.4372	0.4541	0.4419	0.4587
Std	0.0386	0.0442	0.0394	0.0396	0.0403

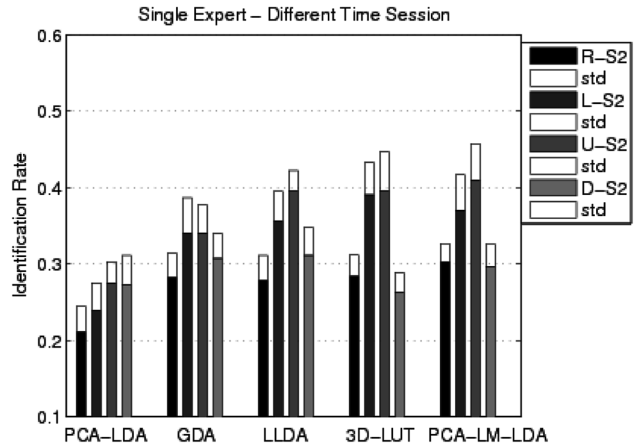
accuracy improvement of the expert fusion. The fact that the experts exploit quite different information sources and have distinct architectures promotes their diversity, which leads to accuracy improvement owing to their classification errors being uncorrelated.

Fusion at the confidence level is considered, where the matching scores reported by the individual experts are combined. We have tested the simple fixed fusion rules such as the sum, product, maximum, minimum, and median rules, as the purpose of this research is just to assess the viability of combining the pose-invariant face classifiers. Among the rules above, the weighted sum rule is also considered, where the weights reflect the accuracy of each expert achieved on an independent evaluation set. The use of any trained combiner instead of the fixed rules, provided a suitable sized evaluation set is available, would be an extension to our work. The confidence value $C_{ij}(x)$ of the base classifier j for class i is given as a reciprocal of the normalized Euclidean distance of the output vectors produced by each expert. The confidence value is scaled so that it is in the range of $[0, 1]$. The combining classifier $Q(x) = \{Q_i(x), i = 1, \ldots, C\}$ is then defined as follows:

$$\begin{aligned} Q_i &= \prod_j C_{ij}(x) \\ Q_i &= \sum_j C_{ij}(x) \\ Q_i &= \max_j C_{ij}(x) \\ Q_i &= \min_j C_{ij}(x) \\ Q_i &= \text{median}_j C_{ij}(x) \\ Q_i &= \sum_j w_j C_{ij}(x). \end{aligned} \tag{20}$$



(a)



(b)

Fig. 7. Identification rates of individual experts on XM2VTS DB for the (a) same time session and (b) different time sessions.

When the training set is large enough and overfitting of the base classifiers is avoided, the fixed combination rules are effective [27].

VI. Experimental Results and Discussions

A. Experimental Setup

We used the XM2VTS face data base [42] for the experiments. The face set consists of 2950 facial images of 295 persons with five pose variations (F, R, L, U, and D) and two different time sessions (S1 and S2) recorded at five-month intervals. This may be the largest public data set of subjects taken in different poses. The XM2VTS data set was manually annotated with pose labels of the face images. Each pose group exhibits a range of variations in personal pose. The images in all sessions were captured under the same illumination setting. The experiments were designed to study the effect of pose on the recognition accuracy, as well as the sensitivity to template aging. The images were normalized to 46×56 pixel resolution with a fixed eye position. The manually annotated eye positions were exploited. Some of the normalized data samples are shown in Fig. 6(a).

The data set was divided into three subsets in a random tenfold manner: 1250 images of 125 persons, 450 images of 45 persons, and 1250 face images of 125 persons for the training, evaluation, and testing, respectively. The training (or prototype) set was utilized to learn the transformation functions of the base experts whereas the evaluation set served to adjust the parameters of each expert such as the kernel parameters of GDA, the dimensionality of the output vectors, and scaling parameters of the individual experts for fusion. These parameters were chosen to achieve the best performance of each expert in terms of identification rate on the evaluation set. The recognition performance is reported as the identification rate on the test set. The frontal face F-S1 of the test set was selected as a gallery and the eight rotated face images of each class in the test set were exploited as queries. Tenfold cross validation was performed in all of the experiments. Note that the three sets for training, evaluation, and testing consist of different face classes. The experts are trained from the training face classes and are applied to new face classes in the test set, that is, the experts should show good generalization performance across different face identities. In order to achieve a reasonable generalization, the number of different subjects in the training set should be large. Here, it was heuristically set to 125. As a sufficient number of face classes were also required for the evaluation and testing, a data set containing a fairly large number of face classes was required. The XM2VTS data set just about met these requirements.

B. Accuracy Comparison of the Experts

The four proposed experts have been compared with the well-known conventional face recognition method, fisherface method (PCA-LDA), where the basis functions of LDA are learned from the eigen-features of the training set [18]. The dimensionality of the feature vectors at both PCA and LDA stages in the PCA-LDA method was chosen to yield its best recognition accuracy over the evaluation set. For the method of 3D-LUT, we used 108 SNU 3-D scanned facial models [43]. For the GDA, an RBF kernel with an adjustable width was deployed. The posterior probability in the LLDA method (see Section IV-B1) was simply hardened to get a crisp pose label for each face image.

The recognition accuracy of the single experts and the PCA-LDA method is shown in Table I and Fig. 7 for the separate experiments involving different poses and time sessions. Overall, the four experts outperformed the classical PCA-LDA method by about 10% on average. The four experts were shown to be roughly comparable. However, the LLDA and PCA-LM-LDA methods tended to be better than the other two in terms of both mean accuracy and standard deviation of the tenfold cross validation. The GDA method exhibited the largest variation over the ten experiments with the different combinations of the face classes in the training, evaluation, and test sets. From Fig. 7, all of the experts and the conventional method showed comparatively poor recognition rates for test data from a different time session. The proposed four experts showed 56.31% identification accuracy on average for the same time session and 33.28% on average for the different time

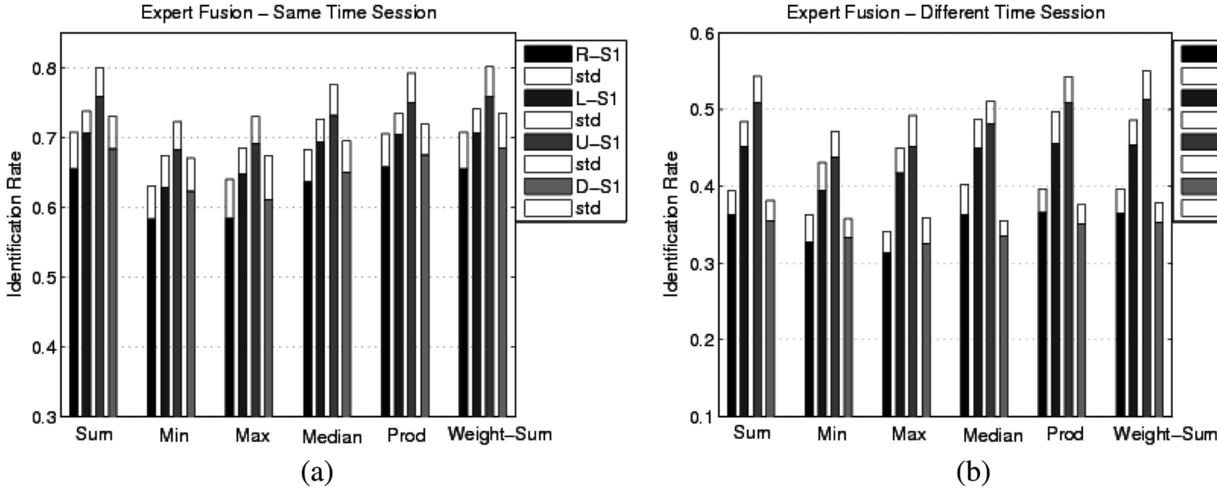


Fig. 8. Identification rates of the multiple expert system on XM2VTS DB for (a) the same time session and (b) a different time session.

TABLE II
IDENTIFICATION RATES OF THE FUSED MULTIPLE EXPERTS OBTAINED BY TENFOLD CROSS VALIDATION. AVERAGE AND STANDARD DEVIATION OF THE DIFFERENT COMBINING STRATEGIES ARE REPORTED

	Sum	Min	Max	Median	Prod	Weighted-Sum
Mean	0.5601	0.5010	0.5050	0.5424	0.5580	0.5607
Std	0.0372	0.0391	0.0413	0.0368	0.0374	0.0386

sessions. Both results exceed the corresponding identification accuracy of the classical PCA-LDA method by about 10%.

We look at the two methods LLDA and PCA-LM-LDA more closely as both experts are trained on the same sources of information and have the most similar architectures among the four experts, which are piecewise linear. Even though the PCA-LM-LDA method learns the representation, view transformation, and the discriminant function separately, and, thus, suboptimally for given a training set, it exhibited comparable performance to the LLDA method in the different pose cases. This might be explained by overfitting of the methods for given a training set. Further comparison of these two methods was carried out by using the PIE data set of 66 identities [44], which were equally divided into the training and test sets. As shown in Fig. 6(b), each class has 15 images (three poses $\times$ five illuminations). The frontal face F1 of the test set was selected as a gallery and all of the other images of the test set were used as queries. The identification rates of the PCA-LM-LDA and LLDA methods were 39% and 47.5%, respectively. The LLDA method well captured the effects of pose and illumination variations, thus yielding superior recognition accuracy. In spite of similarity of these two experts, they had still different classification error characteristics of the given test samples and were well suited for fusion.

C. Multiple Expert Fusion

Table II and Fig. 8 show the results of combining all four experts by the six different fusion rules. All six different gating rules improved the best individual expert and the conventional PCA-LDA method. The sum and the weighted-sum rules showed the best identification accuracy on average. In the weighted-sum rule, the weights were reflected by the accuracy

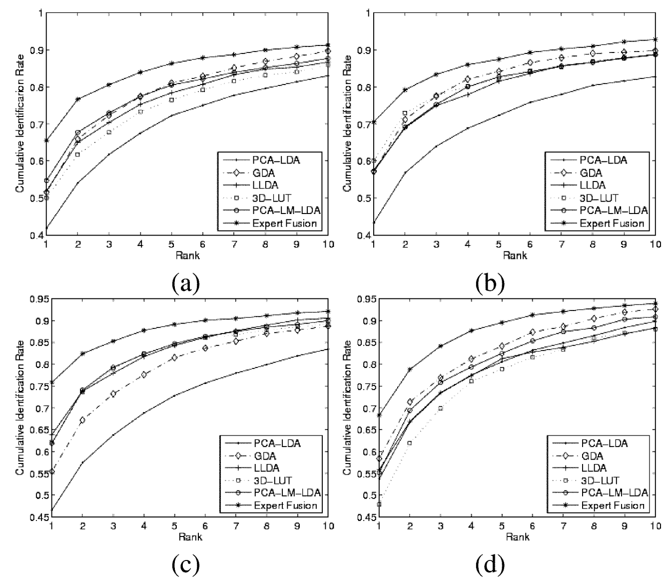
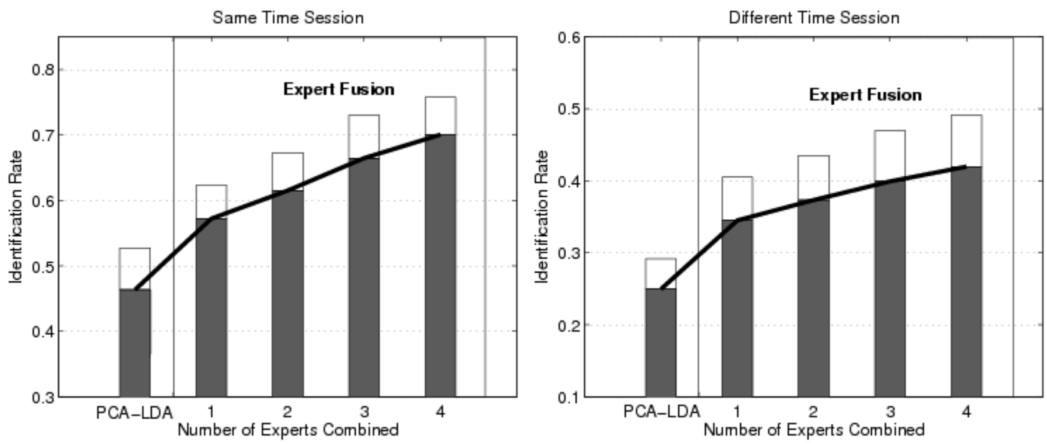


Fig. 9. Cumulative identification plots for (a) the right-rotated, (b) left-rotated, (c) up-rotated, and (d) down-rotated faces recorded in the same time session.

of the experts over the evaluation set. As all four experts exhibited similar accuracy, the accuracy of the weighted-sum and sum rules were similar. Note that the product rule also achieved comparable accuracy with the sum rule, which has about 10% improvement over the best proposed expert and 20% over the PCA-LDA method. The best multiple expert system showed 70.08% for the same time session and 42.06% for a different time session. The cumulative identification rates for each pose, R, L, U, and D, at the same time session are shown in Fig. 9. The proposed fusion of the experts by the sum rule showed constantly superior accuracy over all of the proposed experts and the conventional PCA-LDA method at all different pose cases. The proposed individual experts also consistently outperformed the PCA-LDA method in all cases except the down-rotated experiment. We also investigated the relationship between accuracy and the number of experts fused. We first found the best expert, PCA-LM-LDA, and then added the next best performing expert, and so on, in the order LLDA, 3D-LUT,



(a) (b)

Fig. 10. Identification rates of the multiple expert system for different number of the experts combined by the sum rule in (a) the same time session and (b) a different time session.

GDA. The combined results were obtained by the sum rule. The results are shown in Fig. 10. The identification rate improved quasi-linearly as the number of different experts increased, owing to different characteristics of the experts designed. The accuracy of the proposed multiple expert system expert fusion was significantly better than that of the conventional PCA-LDA method. It improved the classical method from 46% $\rightarrow$ 70% for the same time session and 24.9% $\rightarrow$ 42% for the different time session, respectively, on the 125 face classes rotated by more than 30° from a frontal pose.

VII. Conclusion and Future Work

We proposed a multiple face-recognition expert system based on different models of face view changes. Pose-invariant face identification experts are obtained by learning the statistics of face images or fitting 3-D face models. The proposed individual experts outperformed the classical PCA-LDA method. The fusion of the different experts yielded an impressive performance improvement owing to their different characteristics in terms of sources of information exploited and architectures used. We intend further to improve the performance of the proposed approach by exploiting dense correspondences of facial features in the future. The current performance was obtained with images registered using fixed eye positions, and this is a poor basis for face-image normalization. A more elaborate normalization of face images by using the dense correspondence information is expected to enhance the generalization performance of the pose correction methods, as it was shown to be effective in previous studies of pose-invariant face recognition [12].

Acknowledgment

The authors would like to thank H.-C. Kim, H.-S. Lee, and D.-J. Kim at POSTECH for their helpful discussions and motivation. The assistance of K.-A. Sohn, a former colleague in Samsung AIT, in implementing the algorithms based on 3-D face models is greatly appreciated. The authors would also like to acknowledge the Graphics and Media Laboratory, SNU, for providing the 3-D face models.

References

[1] P. J. Phillips, P. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and J. M. Bone, FRVT 2002: Evaluation Report [Online]. Available: http://www.frvt.org/FRVT2002/

[2] A. C. Aitchison and I. Craw, "Synthetic images of faces—An approach to model-based face recognition," in *Proc. Brit. Machine Vis. Conf.*, 1991, pp. 226–232.

[3] M. C. Lincoln and A. F. Clark, "Towards pose-independent face recognition," in *Proc. IEE Colloq. Visual Biometrics*, Mar. 2000, vol. 6, pp. 1–5.

[4] T. Vetter and T. Poggio, "Linear object classes and image synthesis from a single example image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 733–742, Jul. 1997.

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

[6] Y. Li, S. Gong, and H. Liddell, "Constructing facial identity surfaces in a nonlinear discriminating space," in *Proc. CVPR*, 2001, vol. 2, pp. 258–263.

[7] V. Blanz, S. Romdhani, and T. Vetter, "Face identification across different poses and illuminations with a 3-D morphable model," in *Proc. IEEE Int. Conf. Automatic Face Gesture Recognit.*, May 2002, pp. 192–197.

[8] B. Heisele, P. Ho, and T. Poggio, "Face recognition with support vector machines: Global versus component-based approach," in *Proc. ICCV*, 2001, vol. 2, pp. 688–694.

[9] D. Beymer and T. Poggio, "Face recognition from one example view," in *Proc. ICCV*, 1995, pp. 500–507.

[10] K. Okada and C. Malsburg, "Analysis and synthesis of human faces with pose variations by a parametric piecewise linear subspace method," in *Proc. CVPR*, 2001, vol. 1, pp. 761–768.

[11] R. Gross, I. Matthews, and S. Baker, "Eigen light-fields and face recognition across pose," in *Proc. Int. Conf. Automat. Face Gesture Recognit.*, 2002, pp. 3–9.

[12] ——, "Appearance-based face recognition and light-fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 4, pp. 449–465, Apr. 2004.

[13] L. Fan and K. K. Sung, "A combined feature-texture similarity measure for face alignment under varying pose," in *Proc. CVPR*, 2000, vol. 1, pp. 308–313.

[14] S. Baker and I. Matthews, "Equivalence and efficiency of image alignment algorithms," in *Proc. CVPR*, 2001, vol. 1, pp. 1090–1097.

[15] D. B. Graham and N. M. Allinson, "Automatic face representation and classification," in *Proc. Brit. Mach. Vis. Conf.*, 1998, pp. 64–73.

[16] A. Talukder and D. Casasent, "Pose-invariant recognition of face at unknown aspect views," in *Proc. IEEE Joint Conf. Neural Netw.*, 1999, vol. 5, pp. 3286–3290.

[17] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.

[18] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces versus Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[19] T.-K. Kim, J. Kittler, H.-C. Kim, and S.-C. Kee, "Discriminant analysis by multiple locally linear transformations," in *Proc. Brit. Mach. Vis. Conf.*, Norwich, U.K., 2003, pp. 123–132.

[20] T.-K. Kim and J. Kittler, "Locally linear discriminant analysis for multi-modally 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.

[21] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Computation*, vol. 12, pp. 2385–2404, 2000.

[22] J. Kittler and F. Roli, *Multiple Classifier Systems*. Berlin, Germany: Springer-Verlag, 2000–2001.

[23] K. Woods, W. P. Kegelmeyer, and K. Bowyer, "Combination of multiple classifiers using local accuracy estimates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 4, pp. 405–410, Apr. 1997.

[24] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, "Adaptive mixtures of local experts," *Neural Computation*, vol. 3, pp. 79–87, 1991.

[25] E. Alpaydin and M. I. Jordan, "Local linear perceptrons for classification," *IEEE Trans. Neural Netw.*, vol. 7, no. 3, pp. 788–792, Jun. 1996.

[26] L. Xu, A. Krzyzak, and C. Y. Suen, "Methods of combining multiple classifiers and their application to handwriting recognition," *IEEE Trans. Systems, Man, Cybernet.*, vol. 22, no. 3, pp. 418–435, Mar. 1992.

[27] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998.

[28] J. Czyz, J. Kittler, and L. Vandendorpe, "Combining face verification experts," in *Proc. Int. Conf. Pattern Recognit.*, 2002, vol. 2, pp. 20028–20031.

[29] W. Zhang, S. Shan, W. Gao, Y. Chang, B. Cao, and P. Yang, "Information fusion in face identification," in *Proc. Int. Conf. Pattern Recognit.*, 2004, vol. 3, pp. 950–953.

[30] R. P. W. Duin and D. M. J. Tax, "Classifier-conditional posterior probabilities," in *Proc. Adv. Pattern Recognit. (Lecture Notes in Computer Science)*, Berlin, Germany, 1998, vol. 1451, pp. 611–619.

[31] L. Kuncheva, J. C. Bezdek, and R. P. W. Duin, "Decision templates for multiple classifier fusion: An experimental comparison," *Pattern Recognit.*, vol. 34, pp. 299–314, 2001.

[32] J. Kittler, M. Ballette, J. Czyz, F. Roli, and L. Vandendorpe, "Enhancing the performance of personal identity authentication systems by fusion of face verification experts," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2002, vol. 2, pp. 581–584.

[33] S. Ben-Yacoub, Y. Abdeljaoued, and E. Mayoraz, "Fusion of face and speech data for person identity verification," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 1065–1074, Oct. 1999.

[34] L. Hong and A. Jain, "Integrating faces and fingerprints for personal identification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 12, pp. 1295–1307, Dec. 1998.

[35] T.-K. Kim, View-Transformation of Face Images in Kernel Space—Comparative Study HCI Lab., Samsung AIT, 2003, Technical Report.

[36] T.-K. Kim, H. Kim, W. Hwang, S.-C. Kee, and J. Kittler, "Independent component analysis in a facial local residue space," in *Proc. CVPR*, 2003, pp. 579–579.

[37] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cogn. Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.

[38] A. Pentland, B. Moghaddan, and T. Starner, "View-based and modular eigenspaces for face recognition," in *Proc. CVPR*, 1994, pp. 84–91.

[39] M.-H. Yang, "Kernel eigenfaces versus Kernel fisherfaces: Face recognition using kernel methods," in *Proc. IEEE Int. Conf. Automatic Face Gesture Recognit.*, 2002, pp. 215–220.

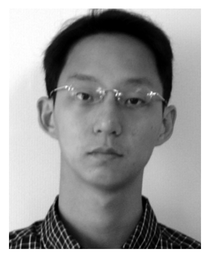
[40] S. Z. Li, J. Yan, X. Hou, Z. Li, and H. Zhang, "Learning low dimensional invariant signature of 3-D object under varying view and illumination from 2-D appearances," in *Proc. 8th IEEE Int. Conf. Comput. Vis.*, 2001, vol. 1, pp. 635–640.

[41] A. Hadid, O. Kouropteva, and M. Pietikainen, "Unsupervised learning using locally linear embedding: Experiments with face pose analysis," in *Proc. 16th Int. Conf. Pattern Recognit.*, 2002, vol. 1, pp. 111–114.

[42] K. Messer, J. Matas, J. Kittler, J. Lueettin, and G. Maitre, "XM2VTSDB: The extended M2VTS database," in *Prof. Audio and Video-Based Biometric Person Authentication*, 1999, pp. 72–77.

[43] B. Choe, H. Lee, and H.-S. Ko, "Performance-driven muscle-based facial animation," *J. Visual. Comput. Animation*, vol. 12, pp. 67–79, 2001.

[44] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression (PIE) database," in *Proc. IEEE Int. Conf. Automatic Face and Gesture Recognit.*, 2002, pp. 53–58.



Tae-Kyun Kim received the B.Sc. and M.Sc. degrees from the Department of Electrical Engineering and Computer Science, Korea Advanced Institute of Science and Technology (KAIST), Seoul, in 1998 and 2000, respectively. He is currently working toward the Ph.D. degree at the Machine Intelligence Laboratory, University of Cambridge, Cambridge, U.K.

He was a Research Staff Member with the Samsung Advanced Institute of Technology, Korea, during 2000–2004. His research interests include computer vision, statistical pattern classification, and machine learning. He regularly reviews for the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE and was the Korea delegate of MPEG-7. The joint proposal of face descriptor of Samsung and NEC, for which he developed the main algorithms, was accepted as the international standard of ISO/IEC JTC1/SC29/WG11.

Josef Kittler (M'92) is Professor of Machine Intelligence and Director of the Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, U.K. He has worked on various theoretical aspects of pattern recognition and image analysis and on many applications including personal identity authentication, automatic inspection, target detection, detection of microcalcifications in digital mammograms, video coding and retrieval, remote sensing, robot vision, speech recognition, and document processing. He has coauthored a book *Pattern Recognition: A Statistical Approach* (Prentice-Hall, 1982) and published more than 500 papers. He is a member of the Editorial Boards of *Image and Vision Computing, Pattern Recognition Letters, Pattern Recognition and Artificial Intelligence, Pattern Analysis and Applications,* and *Machine Vision and Applications.*