Representation Plurality and Fusion for 3-D Face Recognition

Berk Gökberk, Helin Dutağacı, Aydın Ulaş, Lale Akarun, *Senior Member, IEEE*, and Bülent Sankur

***Abstract*—In this paper, we present an extensive study of 3-D face recognition algorithms and examine the benefits of various score-, rank-, and decision-level fusion rules. We investigate face recognizers from two perspectives: the data representation techniques used and the feature extraction algorithms that match best each representation type. We also consider novel applications of various feature extraction techniques such as discrete Fourier transform, discrete cosine transform, nonnegative matrix factorization, and principal curvature directions to the shape modality. We discuss and compare various classifier combination methods such as fixed rules and voting- and rank-based fusion schemes. We also present a dynamic confidence estimation algorithm to boost fusion performance. In identification experiments performed on FRGC v1.0 and FRGC v2.0 face databases, we have tried to find the answers to the following questions: 1) the relative importance of the face representation techniques vis-à-vis the types of features extracted; 2) the impact of the gallery size; 3) the conditions, under which subspace methods are preferable, and the compression factor; 4) the most advantageous fusion level and fusion methods; 5) the role of confidence votes in improving fusion and the style of selecting experts in the fusion; and 6) the consistency of the conclusions across different databases.**

***Index Terms*—Classifier selection, face representation, feature extraction, fusion, independent component analysis (ICA), nonnegative matrix factorization (NMF), 3-D face recognition.**

I. Introduction

AUTOMATIC identification and verification of humans using facial information continue to be an active research area, particularly with the increasing security concerns in the last decade. An overwhelming majority of face recognizers are focused on 2-D intensity images. Despite significant progress in 2-D intensity-based face recognizers, there are still considerable challenges in uncontrolled environments due to the handicaps of pose, illumination, and expression variations as well as occlusion by accessories. With the recent availability of accurate and affordable 3-D sensors, which are capable of sensing both 3-D face shape and texture, it is widely believed that some of the inherent problems of 2-D intensity-based recognizers can be surmounted. It is envisioned that 3-D face recognition can both play a complementary role in 2-D intensity-based recognition and also as a standalone purely 3-D system. Naturally, 3-D face recognition systems are not immune from all handicaps of intensity-based systems, as occlusion and expression problems persist. In this paper, we use the term 3-D face recognition to specifically refer to the 3-D-to-3-D matching problem, where both the probe and gallery faces contain 3-D data. There are also other types of algorithms using 3-D models to solve 2-D-to-2-D matching problems, and these are outside the scope of this paper.

The 3-D face recognition algorithms in the literature have the following commonality of structural modules: 1) 3-D face detection; 2) facial feature localization; 3) face normalization/ alignment and registration; 4) facial feature extraction; and 5) face-matching algorithm and decision unit. The first stage, i.e., the 3-D face detection in cluttered environments or under occlusion, is studied in few papers in [1]. This is mostly due to the fact that the 3-D face capture has to be done in a limited depth range with the present technology, and therefore, the background clutter typical nuisance of 2-D face intensity images is a minor issue. Thus, for most of the systems, the face detection module is reduced to locating and cropping the face region from the rest of the body [2], [3]. Automatic localization of facial landmarks is instrumental for the accurate registration of 3-D faces, and the registration step itself proves to be of paramount importance for the subsequent recognition algorithm. Most of the 3-D face recognition algorithms in the open literature require prior registration, or the registration process is an integral part of the recognition algorithm itself. There are very few papers where rotation-invariant features are used to match nonregistered 3-D faces [4]–[6].

Some other schemes employ 3-D data concomitantly with the 2-D intensity data; hence, their registration must solve the 2-D-to-3-D registration problem. Their alignment is followed by separate 3-D and 2-D facial feature extraction and then training stage of the face matchers. There exists in the literature a plethora of facial features [7], [8] and several combinations to build classifiers.

In this paper, we address the issues of 3-D face representation, facial features, and fusion schemes. First, we consider the crucial issue of assessing the role of different feature sets and classifier mechanisms in a fusion setting. Notice that the choice of informative facial features and their fusion at the score, rank, or decision level are still open problems. The coupling of a specific feature set and a specific classifier is denoted in this paper as an individual face expert. We then analyze the benefits of consultation between these experts vis-à-vis the performance of single experts. Second, we emphasize the relevance of diverse face representation methods and introduce

Manuscript received May 21, 2006; revised December 21, 2006 and June 4, 2007. This work was supported in part by Boğaziçi University under Research Grant 03A203, by the FP6 Biosecure NoE Project, and by TÜBİTAK under Research Grants 103E038 and 104E080. This paper was recommended by Associate Editor K. Bowyer.

B. Gökberk is with Philips Research Laboratories, Eindhoven, The Netherlands.

H. Dutağacı and B. Sankur are with the Department of Electrical and Electronics Engineering, Boğaziçi University, 80815 Istanbul, Turkey.

A. Ulaş and L. Akarun are with the Department of Computer Engineering, Boğaziçi University, 34342 Istanbul, Turkey.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMCB.2007.908865

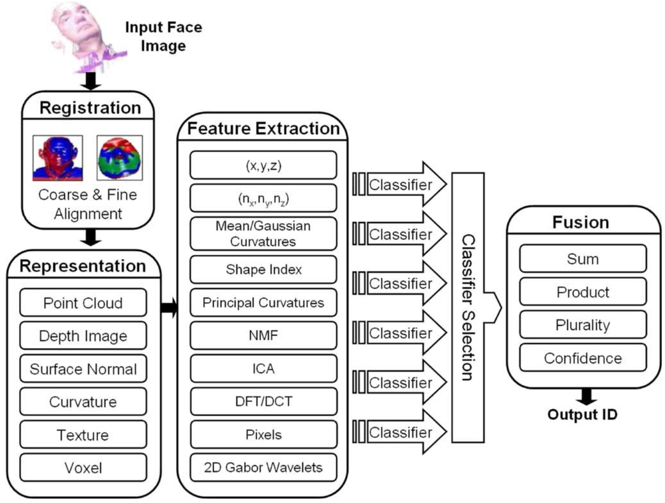


Fig. 1. Overall structure of the proposed face recognition system.

classifier-ensemble-building techniques. Third, some novel 3-D facial feature extraction techniques are reported. The study of these problems results in an extensive comparison study. Notice that both face detection and face registration problems are outside the scope of this paper. We assume, therefore, that the face has been localized, and we employ one of the most frequently used facial-registration methods, namely, the iterative closest point (ICP) algorithm. The overall structure of the proposed scheme is shown in Fig. 1.

The paper is organized as follows. Previous work on 3-D face recognition is presented in Section II. In Section III, coarse face alignment and fine registration methods are explained. face representation schemes and feature extraction modules are introduced in Sections IV and V, respectively. We give a brief overview of the score-, rank-, and decision-level fusion techniques in Section VI. Experimental results are provided in Sections VII and VIII. We conclude in Section IX.

II. Previous Work on 3-D Face Recognition

We first review the major 3-D face recognition approaches in the literature from the perspective of their role in a fusion scheme. At the data level, we discuss various face representation techniques; at the feature level, we point out to the diversity of choices; finally, at the classification stage, we study the fusion algorithms. A recent thorough survey of 3-D face recognition algorithms can be found in [9].

1) Representation Techniques: The most popular approach in 3-D face recognition systems has been to convert the 3-D point-cloud information into 2-D depth images (range images). This conversion operation is needed because the 3-D data may not correspond to a regular grid. While the 2-D data are more familiar to work with, the loss of intrinsic face information due to resampling and mapping to a regular grid must be accounted for. When more than one point is mapped to a cell in a 2-D grid, these points are undersampled during a conversion to 2-D. A case in point is the sloping parts of the face, which suffer due to the foreshortening effect in the 3-D to 2-D conversion. Some of these sloping parts may incorporate interperson differences like the slopes of nostrils. Once the depth image is formed, one can treat the 3-D face recognition problem as simply a 2-D image-matching problem.

Pan *et al.* [10] design a pose-invariant recognition system by projecting the preregistered 3-D point-cloud data to a plane parallel to the face plane. They achieve pose invariance via a variant of the ICP-based registration. Their projection flattens out the facial surface, which is where their algorithm differs from other depth-image-based techniques. Their principal component analysis (PCA)-based identification method outperforms other depth-image-based approaches on the FRGC v1.0 database.

An approach for matching range images, using the original measured data and not their subspace projection, is discussed in [11]. In that work, Russ *et al.* apply the partial-shape Hausdorff-distance metric to range images. The motivation behind using the Hausdorff distance is its partial invariance to inconsistencies such as noise, holes, and occlusions in the 3-D facial data. Their classification experiments conducted on the FRGC v1.0 database show the superiority of the proposed scheme to the standard PCA-based matching algorithm.

As an alternative to depth images, it is also possible to construct 2-D images that represent other properties of 3-D data, such as surface curvature and surface normals. Abate *et al.* [12] generate normal maps, which store three-variate mesh normals in lieu of the red, green, and blue (RGB) components. The difference between the normal maps of the two images is calculated in terms of three difference-angle histograms. The authors, however, do not report comparative-analysis results.

Many 3-D acquisition systems provide 3-D point clouds as raw data, possibly coupled with 2-D texture information. Thus, for many 3-D face recognition systems, point-cloud or point-set data are the default input data representation [13]–[15]. The point-cloud representation of a probe face is subjected to the ICP method for registration, usually to each of the point clouds in the gallery. The quality of the ICP alignment is generally sufficiently good to allow for pointwise matching of two face point clouds. In [16] and [17], all the point sets of the probe and gallery faces are registered to an average face via the ICP in order to align the faces to a common reference frame and to establish dense correspondences. Then, the features are extracted from thus aligned point sets.

There are several alternatives to the ICP-based matchers. For instance, Koudelka *et al.* [18] first locate automatically several facial landmarks such as nose tip, sellion, inner eye corners, and mouth center and then sample 150 random points in their neighborhood. The matching of two facial surfaces is then accomplished via a mixture of ICP and Hausdorff algorithms. The use of the Hausdorff measure is beneficial if there are incomplete or missing data in one of the facial surfaces. The authors show the feasibility of their method on the FRGC v1.0 database.

One shortcoming of the ICP algorithm is that it can only handle rigid transformations. However, human faces generally exhibit nonrigid deformations under expression variations. Therefore, nonrigid registration algorithms could be beneficial in establishing the correspondence between facial surfaces. For this purpose, İrfanoğlu *et al.* [17] use the thin-plate-spline (TPS)-warping algorithm. First, several facial landmarks are located automatically, and a given face image is then warped to an average face model (AFM) using TPS. A similar idea is also proposed in [19] where a generic face model is fitted to a given face, and the related displacement information forms a separate deformation image. Finally, the biometric signature is obtained from the wavelet analysis of this deformation image. Although warping-based registration algorithms may have a potential of establishing better correspondences around the dynamic facial regions, they may have the side effect of suppressing characteristic differences between faces [7]. In order to avoid this side effect, Lu and Jain [20] have suggested the use of person-specific deformable models. The deformations are learned from a small group of subjects. Then, the learned deformation model is transferred to the 3-D neutral model of each subject in the database via TPS. At the matching stage, the person-specific deformable models are fitted to the test face using a modified ICP algorithm where deformation parameters are updated in an iterative way.

Besides ICP, there are other schemes where the registration [21] or correspondence matching process [4], [5] is inherent to the recognition algorithm. Mian *et al.* [4], [5] used rotation-invariant tensors that are constructed in locally defined coordinate bases to represent the 3-D faces. At the recognition stage, the best matching pairs of features, i.e., the correspondences, between the template and test images are found either by exhaustive matching [5] or via a 4-D hash table [4]. Bronstein *et al.* [21] proposed an expression-invariant face recognition algorithm, where one 3-D face is embedded onto another face by multidimensional scaling (MDS). The MDS is used to establish intrinsic geometric correspondence between two similar but deformed surfaces.

Recently, Samir *et al.* [6] represented a facial surface as a collection of planar curves derived from the level sets or (from) geodesic curves that are centered at the nose tip of the face. The recognition performance on the FRGC v1.0 face database is 90.4% with three gallery images per subject. The second type of representation based on geodesic curves is invariant to rotation, but the authors do not report the results with this technique.

2) Fusion Techniques: Pan and Wu [14] present a 3-D face recognition system that combines profile and surface matchers. The three profile experts use one vertical and two horizontal profile measurements. The surface expert makes use of a weighted ICP-based surface matcher. The similarity scores from these four matchers are combined by the sum rule. Obviously, their system is based on shape information only. Their recognition performance on the 3DRMA [22] database having 120 persons shows that the surface matcher, which obtains 8.79% error rate, is better than the profile matchers, and the fusion of the four experts reduces the error rate to 7.93%.

In [16], Gökberk *et al.* have briefly discussed shape-only features such as point-cloud-, surface-normal-, depth-image-, and profile-based shape information. In this first round of fusion experiments, they have observed that some fusion schemes outperform the best individual classifier. They experimented with various combination methods such as fixed rules at score level (sum/product), rank-based combination rule (Borda count), and abstract-level voting method (plurality voting). In addition, they have proposed a two-level serial-fusion scheme where the first level functions as a prescreener, whereas the second level uses linear discriminant analysis (LDA) to better separate the gallery images. Their experimental results on the 3DRMA database show that two-tier serial fusion is most beneficial.

The work presented in [16] can be considered as an initial attempt to discover the usefulness of various fusion schemes for 3-D face recognition problem. In the present paper, we significantly extend the methodology in [16] 1) by the inclusion of a much broader range of different face experts and 2) by covering a wider spectrum of possible fusion rules.

Another type of shape-based expert fusion is proposed in [23]. This approach is essentially a multiregion approach where different facial-region pairs from the gallery and probe images are matched, and the matching scores are combined with the product rule. The local experts compute the surface similarities of three overlapping regions around the nose by using the ICP algorithm, and their registration errors by these three surface matches are then combined. The local experts choose regions specifically around the nose for expression invariance. The recognition experiments conducted on the FRGC v2.0 database show that the proposed multiregion approach obtains 91.9% classification rate in multiple-probe experiments, which is better than the holistic PCA (70.7%)- and the ICP (78.1%)-based algorithms.

A feature-level fusion scheme is presented in [24] where global shape features are concatenated with the local features. The dimensionality of the concatenated vector is reduced by the PCA method.

Fusion techniques are frequently used when both shape and texture modalities are available. The standard approach is to design separate classifiers for each individual modality and to combine them at the score, rank, or decision level. A typical example of this approach is given in [25], where Chang *et al.* have used PCA-based matchers for shape (depth image) and texture modalities and fused their match scores by a weighted sum rule. The experimental results obtained on the UND database containing 198 subjects reveal that fusing the texture and shape modalities achieves 97% identification rate, whereas individual modalities have 96% and 91% identification rates, respectively.

Ben Abdelkader and Griffin [26] use a local-feature-analysis (LFA) technique instead of the classical PCA to extract features from both shape and texture modalities. This classifier combines texture and shape information with the sum rule. Another interesting variant in this paper is the data-level fusion. The depth-image pixels are concatenated to the texture-image pixels to form a single vector. LDA is then applied to the concatenated feature vectors to extract features. The authors report 100% and 98.58% accuracies for the LFA- and LDA-based fusion methods, respectively, for a face database of 185 persons. These accuracies improve the best single modality (texture) rates by 0.24% and 1.36% for the LFA and LDA methods, respectively.

Mian *et al.* [5] propose the use of local textural and shape features together in order to cope with the variations caused by expressions, illumination, pose, occlusions, and makeup. The textural features are based on scale-invariant feature transform. Tensors constructed in locally defined coordinate bases are used as 3-D descriptors. The two schemes are fused at score level with confidence-weighted sum rule. They have tested their algorithm with FRGC v2.0. With 3-D local features, they have achieved identification performances of 89.5% and 73.0% for probes with neutral and nonneutral expressions, respectively. These figures improved to 95.5% and 81.0% with fusion of shape and texture modalities.

A prominent example of fusion of shape- and texture-based matching schemes is presented in [27]. Wang and Chua [27] select 2-D Gabor wavelet features as local descriptors for the texture modality, and they use point signatures as local 3-D shape descriptors. These feature-based representations are matched separately using the structural Hausdorff distance, and then their similarity scores are fused at the score level by using a weighted sum rule. These authors had previously used 3-D Gabor features instead of point signatures as local shape descriptors in [28] in the same setting.

Although most of the studies use fusion of different modalities at the decision stage, it is also possible to combine the modalities before the decision phase. A typical example is given in [15], where shape and texture information is merged at the point-cloud level, thus producing 4-D point features. A variant of the ICP method is then employed to determine the combined similarity of textured 3-D facial shapes.

Similar to the work presented in [16], a two-tier combination idea was also used in [13] for the 2-D texture images. Here, the ICP-based surface matcher eliminates the unlikely classes at the first round and also at the second round; LDA analysis is performed on the texture information to finalize the identification.

A different algorithm that uses feature fusion via hierarchical graph matching (HGM) is presented in [29]. HGM has the role of an elastic graph that stores local features in its nodes and structural information in its edges. HGM is fitted to both the texture-image and shape features since the shape image is registered to the texture image. The scores produced from texture and shape HGMs are then fused by a weighted sum rule. Experimental results obtained on the FRGC v2.0 database show that, although texture modality significantly outperforms shape modality, the integration of scores increases the performance further.

In this paper, we treat the design of a 3-D face recognizer from a more general fusion perspective. Effectively, we develop a fairly complete set of individual face classifiers that result from feasible Cartesian products of face representations and of feature extraction techniques. This is in contrast to the tradition in the literature of selecting two complementary experts—one from shape and the other from texture domain. Another novel aspect of this paper is to design a scheme to determine the experts that should have a voice in the consultation session, rather than inviting every expert indiscriminately.

III. Face Registration via an AFM

The registration step is used to transform all faces into a common coordinate system and then to establish dense point-to-point correspondences. We use a two-tier scheme consisting of a coarse adjustment followed by a refined registration step. The coarse registration uses seven manually located facial landmarks and then applies the Procrustes algorithm for alignment. These fiducial landmarks are the four inner and outer eye corners, the nose tip, and the two mouth corners. An alternative landmarking scheme would have been to use the automatic 2-D/3-D landmarking method being developed in [30] and [31]. In this paper, we choose to use manual landmarking to avoid any errors originating from automatic landmarking and to focus on the performance of individual experts and their fusion.

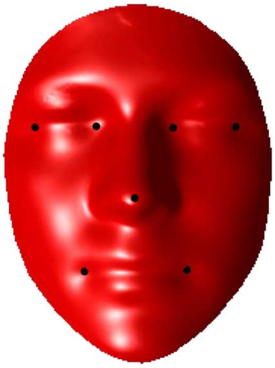


Fig. 2. AFM and its seven landmarks.

We first construct an AFM using a training set of face scans. We average the coordinates of the landmark points, apply the Procrustes analysis to the training faces, and then reaverage the transformed coordinates to obtain the AFM. Fig. 2 shows the computed AFM and its landmark positions. Once the AFM is available, any gallery or probe face will be transformed to the AFM via Procrustes analysis in the coarse-alignment step. It should be noted that Procrustes analysis rescales the faces in order to find the best transformation. This output of the coarse alignment, although essential, is however not adequate for our recognition goals; hence, it must be followed by a second tier of the registration process. To this purpose, we invoke the ICP algorithm to improve the estimates of translation and rotation parameters. This rigid registration algorithm retransforms the Procrustes-aligned face scan on a finer scale to the AFM. Using the output of the ICP algorithm, we can finally establish dense point-to-point correspondences between each 3-D point in the AFM with its nearest point in the face image. Therefore, after the three operations of coarse alignment (Procrustes), fine alignment (ICP), and nearest neighbor selection, each image contains exactly the same number of 3-D points registered to those in the AFM.

Note that the focus of this paper is the design of 3-D face classifiers, and ancillary problems such as face registration and detection are of secondary interest.

IV. Types of Face Representation

There are several alternatives for face representation, with corresponding extracted features. For all feature types, we start from the registered 3-D coordinate data coming from the preprocessing stage. We consider five different representation schemes for recognition purposes: point cloud, surface normals, curvature-based representation, depth image, and 3-D voxel representation. In the following sections, we briefly describe the construction of each representation.

A. Point-Cloud Representation

The point cloud is the set of the 3-D coordinates $\{x, y, z\}$ of the points of a face object. A face with N samples is simply represented in terms of three coordinate vectors X, Y, and Z of length N. All correspondences among points of different faces are determined at the registration step.

Although the ensemble of face points encodes the variations among different faces, there is a very loose neighborhood

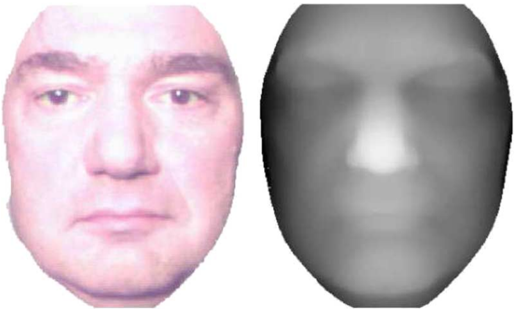


Fig. 3. Texture and depth images of a sample subject.

information in the point-cloud representation due to the 1-D vector structure of the coordinates. The simplest scheme is to use the coordinates themselves as features and to calculate the sum of Euclidean distances between the corresponding points of two faces. We also employ subspace-based techniques on the point-cloud data as described in Section V.

B. Depth Image

One of the most conventional ways to represent face data is the depth image where the z-coordinates of the face points are mapped on a regular $x{-}y$ grid by using linear interpolation. The depth image has the form of a 2-D function $I(x, y)$, which is similar to an intensity image (Fig. 3, right image). Thus, many techniques applicable to intensity images for classifying facial-appearance variations can be directly used for depth images to bring forth facial-landscape differences among subjects. The classical dimensionality-reduction techniques such as the PCA, LDA, and independent component analysis (ICA) have been previously applied to depth images [8], [16], [32]. In Section V, we consider a number of feature extraction techniques applicable to depth images.

C. 3-D Voxel Representation

The initial point-cloud data can be converted to a voxel structure, denoted as $V_{\mathrm{d}}(x, y, z)$, by imposing a lattice. The first step of the voxel conversion procedure is to define an $N \times N \times N$ grid box in such a way that the barycenter of the point cloud coincides with the center of the box. Then, we define a binary voxel occupancy function $V(x, y, z)$ on this grid. This is simply an indicator function: If, in a cell at location (x, y, z), there do not exist any points of the cloud, $V(x, y, z)$ is set to zero. If there are one or more points in that cell, then the binary function at that voxel location assumes the value of one. Therefore, all cells on the face have the value of one, and the rest of the cells in the space are set to zero, which, in effect, defines a 3-D shell. Fig. 4 shows a sample point cloud and the corresponding binary voxels.

We have found it advantageous to convert the binary voxel data into a continuous form via the distance transformation. We apply 3-D distance transform to the binary function $V(x, y, z)$ to fill the voxel grid and obtain $V_{\mathrm{d}}(x, y, z)$. The distance transform is defined as the smallest Manhattan distance of a voxel point to the binary surface. This function gets a value of zero on the face surface, and it increases as we get further away from the surface. By using the distance transform, we distribute the shape information of the surface throughout the 3-D space and obtain a smoother representation compared to the binary voxel description. Fig. 5 shows slices from the voxel representation based on the distance transform.

D. Surface Normals and Curvature-Based Representation

In the 3-D free-form object recognition community, a plethora of local surface descriptors exists [33]–[38]. Among them, surface normals and curvature-based descriptors are the most popular ones. One can consider the human face surface as an instance of a free-form object and can use these local descriptors to represent face information. In the surface-normal-based representation, each point of the facial point-cloud data is described by its 3-D (n_x, n_y, n_z) unit normal vector.

Curvature-related descriptors are attractive since they are invariant to rotations, and therefore, they are frequently used in segmenting 3-D surfaces [1]. Surface normals are features inspired by differential geometry of surfaces, and they actually encode the rate of change of the surface over local patches. The first curvature-based descriptor we use relies on principal directions, which correspond to the maximum and minimum curvatures. Each principal direction is a three-vector in the global coordinate system. We also use two of their derivatives, i.e., the mean (H) and Gaussian (K) curvatures extracted from each facial surface point. These will be referred to as mean curvature- and Gaussian curvature-based representations.

The second local descriptor scheme uses the shape-index values. The concept of shape index was originally proposed by Koenderink and van Doorn [39]. For each point p on the surface, the shape index $S_i(p)$ is defined as

$$S_i(p) = \frac{1}{2} - \frac{1}{\pi} \arctan \frac{\kappa_1(p) + \kappa_2(p)}{\kappa_1(p) - \kappa_2(p)} \tag{1}$$

where $\kappa_1(p)$ and $\kappa_2(p)$ are the principal curvatures at point p, with $\kappa_1(p) \geq \kappa_2(p)$. Shape-index values are scalars in the range $[0, \ldots, 1]$ with the exception of planar surfaces. In the shape-index-based representation, each vertex is characterized by its shape-index value.

E. 2-D Intensity Images

Each 3-D face has its concomitant 2-D color-texture (RGB) information, which is also densely registered to its corresponding shape image. We make use solely of the gray-level information after histogram equalization in order to mitigate any global illumination artifacts.

V. Facial Feature Extraction Methods

We have explored a fairly exhaustive set of features that extract discriminative information from 3-D faces. Some of these features appear in more than one guise. For example, discrete Fourier transform (DFT) was applied on the voxel representation and on the depth image. Similarly, ICA was applied to the point-cloud and depth-field representations of 3-D faces. We assume that the 2-D data (e.g., depth and intensity images) have size $N_1 \times N_2$, the point clouds have size $N \times 3$, and the 3-D voxel data have size $N \times N \times N$.

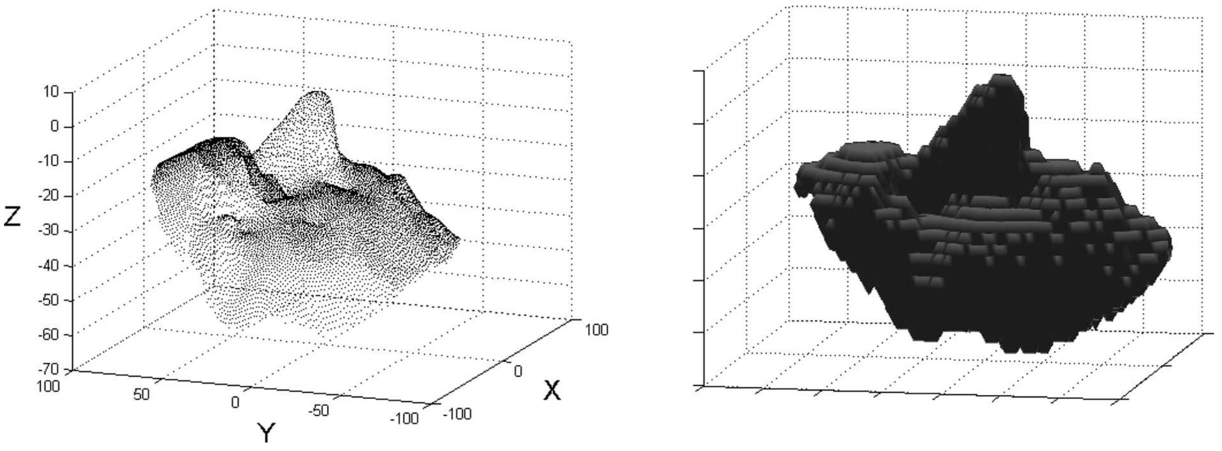


Fig. 4. Point cloud and its binary voxel representation.

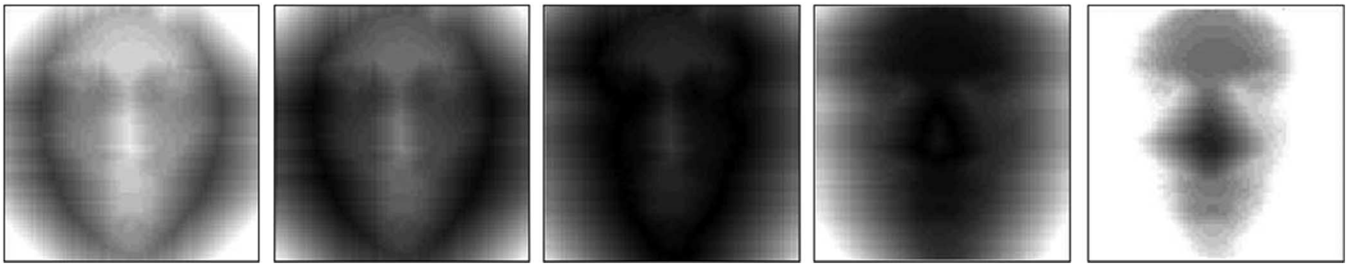


Fig. 5. Slices from the voxel representation based on the distance transform. The slices are parallel to the frontal view of the face and ordered from back of the volume to the front.

A. DFT/Discrete Cosine Transform (DCT)

DFT and DCT are prototypical of model-driven features, and they have been the workhorse of data representation and classification studies. Their low-frequency coefficient sets have excellent representation property, particularly for highly correlated data, where their performances approach that of the Karhunen–Loeve transform. We have employed DFT-based features for both depth images and 3-D voxel data since these two representations provide neighborhood information. In the point-cloud representation, the 1-D vector structure only provides point-to-point neighborhood. On the other hand, DFT/DCT coefficients reflect the spatial dependence of points.

Specifically, for a depth image $I(x, y)$, we calculate its $N_1 \times N_2$-point DFT and extract $K \times K$ low-frequency coefficients to form a feature vector of size $2K^2 - 1$ by concatenating the real and imaginary parts of the coefficients. Likewise, we compute the global DCT. However, in this case, we obtain a feature vector of size K^2 since the DCT coefficients are real. For faces represented in terms of voxels, we compute the 3-D DFT of the distance transform $V_{\mathrm{d}}(x, y, z)$. The feature vector of size $2K^3 - 1$ is obtained by concatenating the low-pass $K \times K \times K$ real and imaginary terms.

Faces have typically slowly varying surface shapes, which means that there exists a rapid power differential in the DFT/DCT coefficients with an increasing frequency. We only select the $K \times K$ ($K \times K \times K$ for the 3-D voxel data) low-pass coefficients, where K is not larger than ten. While the energetic coefficients at DC and at very low frequencies represent the gross structure, a portion of the higher frequency coefficients carries the shape-difference information between individuals. These coefficients, which are important for face classification, tend to be overshadowed by the heavyweight coefficients. This problem can be remedied by the QR-decomposition technique. We thus apply QR decomposition to these feature vectors: $\mathbf{F} = \mathbf{QR}$, where $\mathbf{F}$ is the matrix consisting of feature vectors if we have only one training sample per individual. For the case of more than one sample per individual, $\mathbf{F}$ contains the difference of the feature vector of each subject to its class mean. In this case, the QR decomposition corresponds to a variant of linear discriminant analysis, where $\mathbf{F}$ corresponds to the within-class scatter matrix. $\mathbf{R}$ is the upper triangular matrix obtained from QR decomposition of the training features. In effect, we transform all the feature vectors in both training and test sets by multiplying them with the inverse of $\mathbf{R}$, so that a feature vector f is mapped to $f^T \leftarrow f^T \mathbf{R}^{-1}$. Finally, the transformed test and training feature vectors are compared using the cosine distance.

B. Independent Component Analysis

The ICA and nonnegative-matrix-factorization (NMF) features are the typical examples of data-driven features, which have recently become quite popular. We test the potential of the ICA scheme as a discriminative feature for 3-D face data. We extract the ICA coefficients from either the 3-D point cloud or the depth-image representation.

The ICA is a statistical technique based on a signal model where the observations are treated as mixtures of unobserved sources. There are two different architectures for ICA, which are called ICA1 and ICA2. In ICA1, the basis images are independent, and in ICA2, the mixing coefficients are also

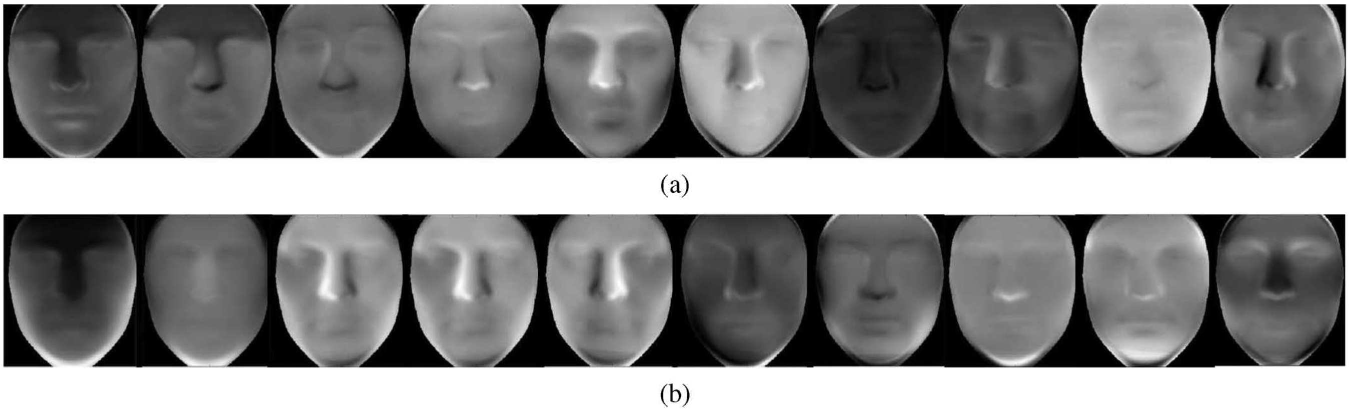


Fig. 6. First ten basis faces obtained from (a) PCA and (b) ICA methods.

independent. We utilize the ICA2 architecture, where each point in a face is considered as a mixture of independent coefficients. If $\mathbf{X}$ is a data matrix incorporating the measured variables, then it can be split as follows: $\mathbf{X} = \mathbf{AS}$, where $\mathbf{A}$ is the mixing matrix, and $\mathbf{S}$ contains the independent coefficients. The columns of $\mathbf{A}$ form a basis for the database, whereas the columns of $\mathbf{S}$ provide the ICA features for the corresponding faces residing in the columns of the data matrix $\mathbf{X}$.

For the point cloud, all x-, y-, and z-coordinates of a face are concatenated to a single vector. Its dimensionality is then reduced via PCA. The columns of the data matrix $\mathbf{X}$ for the ICA analysis are constituted of the first K PCA coefficients of the faces. Then, the FastICA algorithm described by Hyvarinen and Oja [40] is applied to obtain the basis $\mathbf{A}$ and the independent coefficients $\mathbf{S}$. Finally, we apply the QR-decomposition technique to the ICA-based features to reweight the elements of the feature vector according to their discriminative power.

The ICA analysis for depth images follows a similar procedure. The columns of a depth image are concatenated to form a single 1-D vector (one for each face). These data are subjected to PCA reduction, ICA decomposition, and QR normalization.

Fig. 6(a) shows the first ten basis functions derived from PCA, whereas Fig. 6(b) shows the ten independent face components. PCA only captures the second-order variations of the general face geometry, whereas one can observe that the ICA basis images resemble the individual faces in the database.

TABLE I
REPRESENTATIONS, FEATURES, AND ACRONYMS FOR FACE EXPERTS

Representations	Features	Acronym
Point Clouds	(x,y,z) coordinates	PC-XYZ
	ICA coefficients	PC-ICA
	NMF coefficients	PC-NMF
Surface Normals	(nx,ny,nz) unit normals	SN
Depth Images	Pixels	DI-PIXEL
	DCT coefficients	DI-DCT
	DFT coefficients	DI-DFT
	ICA coefficients	DI-ICA
	NMF coefficients	DI-NMF
Curvature	Shape-index (SI)	CURV-SI
	Principal directions (PD)	CURV-PD
	Mean curvature (H)	CURV-H
	Gaussian curvature (K)	CURV-K
Voxel	3D DFT coefficients	VOXEL-DFT
Texture Images	Pixels	TEX-PIXEL
	2D Gabor wavelet coefficients	TEX-GABOR

C. Nonnegative Matrix Factorization

NMF [41] is another matrix-factorization technique with the added constraint that each factor matrix has only nonnegative coefficients. It has been observed that avoiding the artificiality of negative coefficients enhances the physical significance of the component sources. In fact, each source resembles a part of the object leading to a part-based description. For example, when we use NMF decomposition of 2-D intensity faces, we observe that the basis vectors are found to reflect the local features of faces. We explore the recognition capability of both ICA- and NMF-based features of 3-D faces in a standalone mode as well as in fusion scenarios.

Given a nonnegative data matrix $\mathbf{X}$ of size $M \times L$, we obtain two nonnegative matrices $\mathbf{W}$ and $\mathbf{H}$ such that $\mathbf{X} \approx \mathbf{WH}$, where $\mathbf{W}$ is of size $M \times K$, and $\mathbf{H}$ of size $K \times L$. Since we force the two matrices to be nonnegative, we can only reconstruct $\mathbf{X}$ approximately from their product. The columns of $\mathbf{W}$ can be regarded as basis vectors, and the columns of $\mathbf{H}$ are utilized as feature vectors of the corresponding faces.

Parallel to the preprocessing stage of ICA decomposition, we first apply PCA to reduce the dimensionality of the raw data (depth or point-cloud information) and place the first M PCA coefficients of each face into the columns of the data matrix. We add a constant to the PCA coefficients to obtain a nonnegative data matrix. The nonnegative factors $\mathbf{W}$ and $\mathbf{H}$ are obtained using the multiplicative update rules described in [42]. Then, the QR decomposition is applied to the NMF-based features as described in Section V-A.

D. 2-D Texture-Image Features

We have extracted two different features from the texture images. The first one is the simple pixel-based approach that codes the image as a vector of grayscale intensity values. The second approach is based on the well-known 2-D Gabor wavelet-based approach [28], [43]–[45]. Since the intensity images are aligned in the registration phase, we do not need to employ a time-consuming elastic-bunch graph-based localization algorithm. Instead, we place a rectangular grid of size

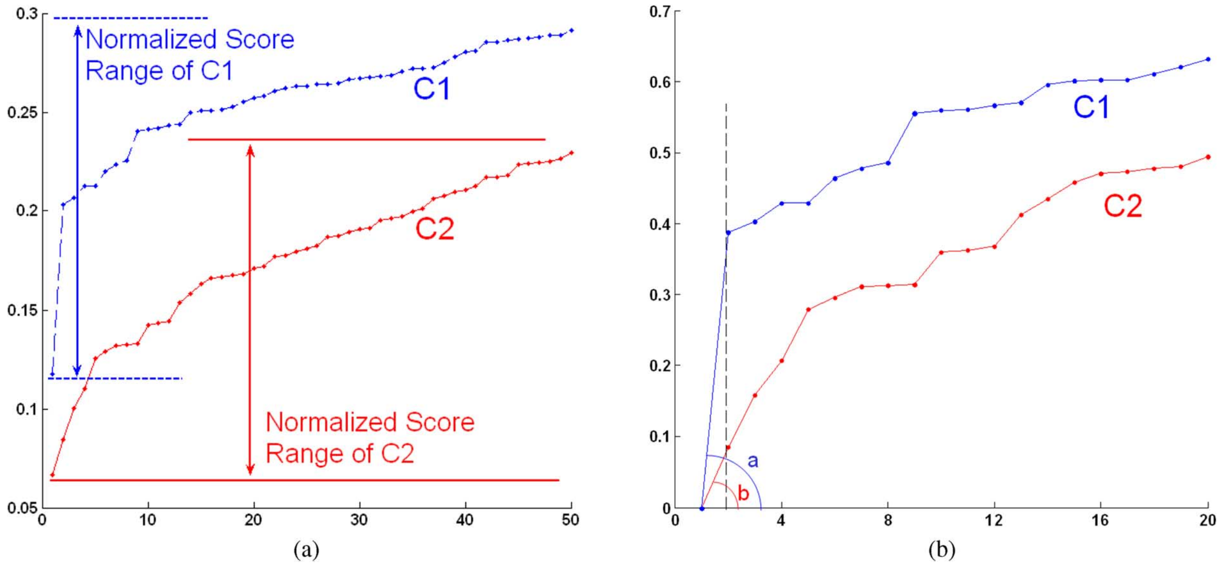


Fig. 7. Illustrative example of the estimation of confidences for the top-ranked classes. (a) Normalized scores (distances) of a test example for each class in the training set (in increasing order). Classifiers 1 and 2 have different score ranges (denoted by double arrows). (b) Renormalized distances calculated from (2). Slopes a and b denote the estimated confidences for the top-ranked class for classifiers 1 and 2, respectively.

$M \times N$ and obtain Gabor coefficients only at the grid points. Five different frequencies and eight equally spaced orientations are used to generate Gabor kernels. Therefore, in the Gabor-based representation scheme, we obtain a feature vector of dimensionality $M \times N \times 40$. It is also possible to apply feature transformation methods to the raw Gabor features. Indeed, we find it useful to apply LDA to the Gabor features if sufficient training data are available. Therefore, we make use of two different schemes for the texture-based Gabor classifier: the one that uses raw Gabor magnitudes, and the one which extracts LDA coefficients.

Table I shows all of the representation techniques and their corresponding feature extraction methods. As a result of the feasible Cartesian product of representations and features, we create 16 different face experts. Their acronyms, as given in Table I, will be used to represent the corresponding experts in the rest of the paper.

VI. Fusion Methods

A survey of classifier-fusion techniques used in 3-D face recognition community (Section II) reveals the dominance of fixed combination rules such as sum and product rules [46]. The preference for these methods stems from the following facts: 1) they are simple yet effective; 2) the number of training samples per subject is very limited in face recognition applications, and this precludes more advanced classifier combination methods such as stacking [47], mixture of experts [48], bagging [49], and AdaBoost [50]; and 3) integration at score, rank, or decision level is flexible, in that a new expert's opinion can be easily incorporated without affecting the existing experts. To the best of our knowledge, [7] and [16] are the only two principled studies of the tradeoffs of various fusion algorithms for the 3-D face recognition problem.

In the sequel, we briefly review a number of fusion methods applicable to 3-D face recognition systems. Abstract-level fusion algorithms are used to fuse the individual experts that produce only class labels. In this category, plurality voting (PLUR) is the most commonly used one, which just outputs the class label having the highest vote. If the classifiers produce a ranked list of class labels, then rank-level fusion schemes can naturally be used. In this second category, the Borda-count method is often used. The Borda scheme calculates the combined ranking by summing the class ranks as assigned by the individual classifiers. The fused opinion is then simply the class having the smallest total rank sum. The third fusion category is commonly referred to as the score- or measurement-level fusion since, given a test pattern, individual classifiers produce class similarity scores. These scores can be combined by using simple arithmetic rules such as sum, product, min, max, and median rules after score normalization [46].

We have devised a method to improve the score- and abstract-level combination methods using the estimated confidences of individual classifiers. The confidences can be attributed according to the similarity scores that are reported by the classifiers. A caveat is that it may be misleading to use normalized scores of the top-ranking class labels due to noncommensurate score normalizations. Score-normalization techniques, such as the min–max method, use only the training set, and their generalization ability is not optimal. A sample case is shown in the piecewise linear curves in Fig. 7(a), where the x-axis denotes the rank, and the y-axis denotes the distance scores in increasing order, for a given test pattern. Reading off from the graph at the first rank $(x = 1)$, we see that the nearest class found by the second classifier has a distance of 0.06, and the nearest distance found by the first classifier is 0.13. Accordingly, the second classifier seems to be more confident than the first one. However, in this particular case, this is wrong since the score range of classifier 1 (0.13–0.9) is different from that of classifier 2 (0.06–0.8), and the second classifier generally gives lower score values. This pitfall is due to the insufficient training data in estimating the score-normalization parameters. Obviously, if the classifier score ranges were the same, then the scores could be used right away as confidences.

To compensate for range disparity, we propose to use a differential confidence measure, i.e., the relative distance between the first two nearest neighbors of the classifier. The procedure is as follows. Given a probe image, we generate $d = [d_1, d_2, \ldots, d_N]$ as the vector of ranked distances to the N classes. Here, d_1 is the distance between the test sample and its nearest class in the training set, whereas d_N is that of the least similar one. An obvious score range for a classifier is $r = d_N - d_1$, whereas we prefer the more robust median estimate $r = \text{Med}(d_1, d_2, \ldots, d_N) - d_1$. The score normalization is then effected via the following:

$$d'_i = \frac{(d_i - d_1)}{\text{Med}(d_1, d_2, \ldots, d_N) - d_1}, \qquad i = 2, \ldots, N. \tag{2}$$

Finally, the classifier confidence score is declared as simply d'_2, as shown in Fig. 7(b). One can interpret the d'_2 corrected confidence as the slope of the d' curve at the x-intercept. After this correction, the classifier 1 in the example of Fig. 7 becomes the higher confidence classifier.

In [51], a similar idea is used to weight the scores during the SUM rule-based fusion phase. However, in our approach, once we compute the confidences, we do not use score values since they usually have different ranges due to the suboptimal score normalization. We propose to use an improved version of the plurality voting—the modified plurality voting (MOD-PLUR): Whenever there are ties, we select the class label having the highest average confidence value among the equiprobable classes.

For comparison purposes, we keep track of the simpler approach where the class with the greatest confidence among the top-ranked classes is selected. This second approach will be referred to as the highest confidence (HC) fusion.

VII. Experimental Results

A. Face Database and Experimental Protocols

We have tested our algorithms on two databases, namely, the FRGC v1.0 and the FRGC v2.0 [51], [52]. For both databases, texture information is stored as RGB values with 480 × 640 resolution. Shape data contain between 30 000 and 40 000 3-D coordinates. Although the quality of the scanned data is high, there are two types of noise affecting 3-D faces: small protrusions and impulse noise-like jumps.

After the preprocessing operations of alignment and cropping, the original 3-D point clouds with varying number of samples are reduced to a fixed number of registered 16 560 points, and the correspondences are established. Similarly, depth and texture image resolutions are reduced to 281 × 321.

The original FRGC v1.0 database contains 943 3-D scans of 275 subjects. We had to use a subset of the original database since 75 subjects had only one scan, and 14 3-D scans were badly registered with the texture data. Thus, the part of the database involved in our experiments contained 854 2-D and 3-D scans of 195 subjects. Each subject had at least two and at most eight 3-D scans. The FRGC v1.0 database consists mostly of frontal faces and does not exhibit significant expression variations. However, some scans have slight in-depth pose variations and different expressions.

TABLE II
Experimental Configurations for the FRGC v1.0 Data Set

	Training samples per subject	Number of Subjects	Training scans	Test scans	Fold count
E_1	1	195	195	659	2
E_2	2	164	328	464	3
E_3	3	118	354	300	4
E_4	4	85	340	182	5

In the FRGC v2.0 database, there are 4007 face scans of 465 subjects. The subjects in the FRGC v1.0 database are included in this database, too. We have eliminated 55 subjects since they had only one scan. Thus, we have used 3952 scans of 410 subjects in our identification experiments. The FRGC v2.0 database, as opposed to v1.0, contains face scans with significant expression variations and presents a grander recognition challenge.

For FRGC v1.0, we have designed four different experimental configurations, as shown in Table II. Each configuration contains a different amount of training samples per subject. The subscript i in experiment E_i denotes the number of training samples per subject in that experiment. The reason for the different populations is that in the FRGC v1.0 database, 195 subjects have more than two 3-D scans, 164 subjects have more than three scans, etc. Thus, E_1 is designed so that every subject possesses only one image in the training set, whereas the rest ($854 - 195 = 659$ images) are placed in the test set. When there are two or more images per subject, one can assign the role of training and test samples in a round-robin fashion. For example, if there are n images per subject, then there exist n different ways of selecting the probe (test image) and the gallery (remaining $n - 1$ training images). We call each such assignment a “fold” and report performance results as the average of these folds. The number of folds is shown in the last column of Table II. The most difficult experiment is obviously E_1; while there exists a single training image per person, both the enrolled population and the number of test images are largest.

In the experiments with FRGC v2.0, we have only considered the case where there is only one gallery image in the database. Since we have called the corresponding experiment protocol E_1 for the FRGC v1.0 data set, we call this protocol E'_1. However, the two experimental protocols have an important difference: we have used the FRGC v1.0 to train our subspace-based methods such as the ICA and NMF and used the class information in FRGC v1.0 to estimate the LDA and QR-normalization parameters. Then, these parameters and the basis images were fixed and were used to calculate the feature vectors of the gallery images as well as the probe images of FRGC v2.0. We have chosen the earliest scan of each subject as the gallery image. All the 410 gallery images are neutral, i.e., they do not have facial expressions. All the other scans are used as test images: Thus, we have 3542 test images. Some 1984 of the test images are neutral faces, and the remaining 1558 faces exhibit expression variations.

TABLE III
BASE EXPERTS' AVERAGE RANK-1 IDENTIFICATION PERFORMANCES (IN PERCENT) ON THE FRGC V1.0 DATABASE

Experts	Dimensionality	Distance Measure	E_1	E_2	E_3	E_4
PC-XYZ	$16,560 \times 3 = 49,680$	$\sum L2$	87.71 ± 0.00	94.68 ± 0.45	97.92 ± 0.57	98.90 ± 1.29
PC-ICA	90	COS	85.66 ± 1.18	**98.71** ± 0.22	**99.67** ± 0.47	**99.89** ± 0.25
PC-NMF	90	COS	85.13 ± 1.29	**97.77** ± 0.12	**99.25** ± 0.63	**100.00** ± 0.00
SN	$16,560 \times 3 = 49,680$	$\sum L2$	**89.07** ± 1.50	96.84 ± 0.33	98.92 ± 0.42	99.45 ± 0.67
DI-PIXEL	$281 \times 321 = 90,201$	L2	55.99 ± 1.50	70.19 ± 1.62	79.75 ± 1.20	87.69 ± 2.92
DI-DCT	49	COS	78.53 ± 2.47	**97.63** ± 0.57	**99.58** ± 0.50	**99.78** ± 0.30
DI-DFT	49	COS	75.95 ± 2.25	97.13 ± 1.11	99.08 ± 0.42	99.56 ± 0.46
DI-ICA	80	COS	72.46 ± 0.97	96.55 ± 0.78	98.92 ± 0.50	99.01 ± 0.60
DI-NMF	70	COS	71.55 ± 0.54	95.83 ± 0.33	98.67 ± 0.90	99.67 ± 0.30
CURV-SI	16,560	L1	**90.06** ± 1.18	96.55 ± 0.57	98.67 ± 0.27	99.34 ± 0.60
CURV-PD	$16,560 \times 3 \times 2 = 99,360$	$\sum(COS + COS)$	**91.88** ± 0.54	97.13 ± 0.76	99.08 ± 0.88	99.45 ± 0.67
CURV-H	16,560	L1	87.41 ± 1.29	95.69 ± 0.78	98.50 ± 0.43	98.90 ± 0.39
CURV-K	16,560	L1	84.37 ± 1.72	93.89 ± 0.12	97.25 ± 0.88	98.46 ± 0.46
VOXEL-DFT	53	COS	64.26 ± 0.97	91.16 ± 1.35	97.92 ± 1.52	99.34 ± 0.46
TEX-PIXEL	$281 \times 321 = 90,201$	L2	64.04 ± 0.43	77.16 ± 1.41	84.33 ± 1.47	92.53 ± 1.97
TEX-GABOR	$887 \times 40 = 35,480$	L1	74.73 ± 1.61	87.36 ± 1.47	91.92 ± 1.77	96.26 ± 1.06

B. Comparative Analysis of Individual Face Experts

Table III shows the rank-1 correct classification rates of the face experts for the four experimental setups conducted on FRGC v1.0, where the boldface figures denote the top three competitors in that category. Each expert relies on a different feature extraction method and thus has different input-feature-vector size. For each method, we provide feature dimensionality and the distance measure used. Note the wide disparity in feature dimensionality. For example, transform-domain features have a compression ratio of 1 : 350 vis-à-vis the raw point-cloud data. Moreover, a distance measure appropriate for each feature type was determined experimentally from the L1, L2, and COS set as given in Table III. For the multidimensional features, i.e., surface normals, principal directions, and point-cloud coordinates, the distances are simply calculated for each dimension separately and then summed. For these cases, the distance measure has the additional symbol Σ. One exception is the curvature principal directions (CURV-PD) method since this feature consists of two "three vectors." Therefore, the distance is calculated by the sum of two three-vector differences. By inspecting the results obtained on the FRGC v1.0 database, we find it useful to state the following comments.

1) Not surprisingly, there is a jump in performance between the single-gallery case E_1 and the experiments with at least two training images per subject. In fact, almost half of the face experts attain a nearly perfect classification whenever at least two training face samples are provided.
2) For the single-gallery experiment E_1, the top three experts are all related to surface curvatures [CURV-PD, curvature shape index method (CURV-SI), and surface normals method (SN)]. We consider the surface normals to form an indirect expression of surface curvature. In the multigallery experiments (E_2, E_3, and E_4), the subspace techniques PC-NMF, PC-ICA, and depth image-based discrete cosine transform method (DI-DCT) outperform others. This shift from surface to subspace experts as more data become available is intriguing but can be explained as follows. The subspace techniques use all of the available data collectively to extract features, whereas the surface techniques still rely on individual but multiple comparisons. For example, the PC-NMF technique, which has the perfect score in E_3, falls to a mediocre position in E_1 with a score of only 85%. We conjecture that the subspace techniques achieve their full potential when adequate training data are supplied to construct their feature subspaces. The subspace techniques need more training samples to model the within-class variability through the analysis into the basis faces and the corresponding coefficients. The final QR-normalization step in the subspace-based techniques also requires at least two training samples per subject in order to reweight the features according to class separability.
3) One important observation is that the discrimination abilities of surface-based descriptors i.e., CURV-PD, CURV-SI, SN, and PC-XYZ, are better than others. Another observation is that the 3-D directions of the minimum and maximum curvatures carry better discriminatory information as compared to the scalar version of the curvature information, i.e., magnitude alone of the mean or Gaussian curvatures. For example, principal direction-based CURV-PD expert improves the identification accuracy by almost 2% when compared to shape-index-based CURV-SI classifier.
4) Inspection of the performance scores in Table III reveals that face recognition experts using similar face representation methods achieve similar scores. Thus, it is not the feature type per se that is the determining factor but the underlying information representation. In fact, we may group the face experts in the order of decreasing discrimination power as follows: curvature-, point-cloud-, depth-image-, texture-, and voxel-based. Once the representation type is chosen, the performance variations due to features become relatively small. Hence, we should shift the focus from the choice of feature to the choice of representation. To give an example, consider ICA- versus NMF-based features for experiment E_1.

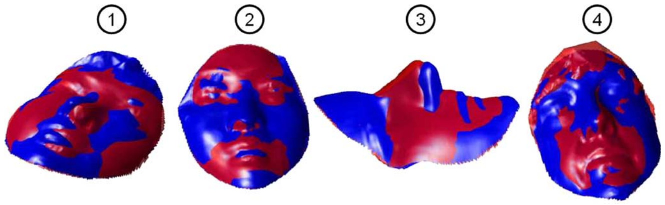


Fig. 8. Misclassified faces in experiment E_1. (1) Pan variation. (2) Incorrectly normalized faces. (3) Tilt variation. (4) Errors due to hair region.

TABLE IV
BASE EXPERTS' RANK-1 IDENTIFICATION PERFORMANCES (IN PERCENT) FOR E_1' EXPERIMENT ON THE FRGC V2.0 DATABASE

Method	Dimensionality	Rank-1 Accuracy
PC-XYZ	70	80.07
PC-ICA	300	88.31
PC-NMF	200	86.34
SN	50	83.79
DI-PIXEL	600	57.82
DI-DCT	169	76.14
DI-DFT	127	73.97
DI-ICA	450	67.25
DI-NMF	300	62.68
CURV-SI	80	75.30
CURV-PD	85	80.35
CURV-H	80	72.56
CURV-K	80	70.78
VOXEL-DFT	127	72.67
TEX-PIXEL	80	69.65
TEX-GABOR	120	80.27

The depth-image-based classifiers DI-ICA and DI-NMF obtain 72% average performance rate. On the other hand, with the point-cloud representation, PC-ICA and PC-NMF achieve 85% average recognition rate. Hence, it is the representation (depth versus point cloud), rather than the feature extraction tool, that is the determining factor (ICA versus NMF).

5) In terms of the usefulness of shape and texture modalities, we observe the clear superiority of the shape-based face classifiers. In this paper, we used Gabor-wavelet-based 2-D recognition algorithm since Gabor-based algorithms are widely used in the literature [28], [43]–[45]. However, the 2-D texture-based Gabor method can only attain the appallingly low 74.73% correct classification rate in E_1.

Fig. 8 shows four sample face images misclassified by all of the 16 face experts in experiment E_1. Blue face (lighter) is the gallery image, whereas the red (darker) face is the misclassified probe image. Errors generally stem from incorrect registration of faces. Pose discrepancies along both vertical and horizontal axes are visible in the first and third images. Another source of error is particularly visible in the forehead regions due to the presence of hair (see the first and fourth images).

Table IV shows the individual rank-1 correct classification rates on FRGC v2.0 with the single-gallery-image setup. As stated before, the FRGC v1.0 database has been used to tune the parameters of the subspace-based methods, the QR normalization, and the linear discriminant functions. In the FRGC v2.0 experiments, we apply the LDA to the features of PC-XYZ, SN, CURV-K, CURV-H, CURV-SI, CURV-PD, pixel-based texture method (TEX-PIXEL), and Gabor-based texture method (TEX-GABOR) methods. For the DI-PIXEL approach, the PCA feature extraction algorithm is employed to compute the feature coefficients. Although we apply LDA and PCA to the raw feature vectors in the v2.0 database, we keep the names of 3-D face experts as used in the v1.0 experiments in order to avoid confusion. Our experimental results show that, with the help of FRGC v1.0 training set, it is possible to significantly improve the identification rates of these methods when compared to using their raw features only. The second column of Table IV displays the feature dimensionality of each method. For the methods that use LDA or PCA, dissimilarities between feature vectors in the transformed subspace are calculated using the cosine distance. For DFT-, DCT-, ICA-, and NMF-based methods, we have increased the dimensionality in subspace-based techniques (when compared to the FRGC v1.0 experiments) since we need more features to discriminate between the subjects in a larger database. The individual performances in FRGC v2.0 can be interpreted in relation to the results obtained with FRGC v1.0 as follows.

1) Point-cloud-based PC-ICA and PC-NMF methods perform best, yielding 88.31% and 86.34% rank-1 identification accuracies, respectively. Since we have built the subspace models using FRGC v1.0, we had enough data to construct the subspaces.
2) In general, point-cloud-based methods perform better than depth-image-based methods. The best depth-image-based method, namely, the DI-DCT, reaches 76.14% identification rate, whereas all of the point-cloud approaches attain identification rates greater than 80%.
3) The best two surface-descriptor-based approaches, the SN and the CURV-PD, attain 83.79% and 80.35% recognition rates, respectively.
4) Contrary to the situation in the FRGC v1.0 experiments, the Gabor-based texture classifier now attains similar identification rates when compared with the shape-based classifiers. This improvement is due to LDA transformation that uses a sufficiently large training set from the FRGC v1.0 database.

For the FRGC v2.0 simulations, we also implemented $1:1$ ICP-baseline matcher for comparative analysis with the PC-XYZ algorithm. Given a probe scan, the ICP-baseline matcher tries to register that probe face to all of the gallery scans, and it outputs the class having the smallest registration error. As explained in Section III, our PC-XYZ algorithm essentially follows the same procedure, but with the help of a single AFM, thus speeding up the registration process. In order to see the effects of the AFM-based rapid-registration algorithm on the identification performance, we carried out identification simulations for the ICP-baseline algorithm on the v2.0 database. In order to decrease the time complexity, we have used the cropped central facial regions (see the AFM in Fig. 2 for the illustration of the cropped regions used in our simulations) and reduced the point-cloud size by a factor of four by taking every other column and row.[1] The ICP-baseline matcher obtains 78.43%

[1] The 3-D point clouds are stored in matrices of size 480 × 640 in the FRGC file format.

TABLE V
RANK-1 IDENTIFICATION PERFORMANCES (IN PERCENT) OF THE FUSION METHODS ON THE FRGC v1.0 DATABASE

	E_1	E_2	E_3	E_4
Best Individual	91.88	98.77	99.67	100.00
MIN	88.54 (-3.34)	95.62 (-3.09)	97.75 (-1.92)	98.68 (-1.32)
MAX	61.15 (-30.73)	84.70 (-14.01)	89.00 (-10.67)	93.85 (-6.15)
MEDIAN	83.08 (-8.80)	95.19 (-3.52)	98.00 (-1.67)	99.34 (-0.66)
BORDA(C:All)	88.31 (-3.57)	97.34 (-1.37)	99.33 (-0.34)	99.67 (-0.33)
SUM	92.03 (0.15)	99.21 (0.50)	99.75 (0.08)	**100.00 (0.00)**
PROD	72.23 (-19.65)	99.35 (0.64)	**99.83 (0.16)**	**100.00 (0.00)**
PLUR	93.40 (1.52)	**99.50 (0.79)**	**99.83 (0.16)**	**100.00 (0.00)**
HC	93.32 (1.44)	98.78 (0.07)	99.42 (-0.25)	**100.00 (0.00)**
MOD-PLUR.	**93.63 (1.75)**	99.43 (0.72)	**99.83 (0.16)**	**100.00 (0.00)**

correct classification rate, which is 1.64% worse than the PC-XYZ performance (see Table IV, second column). A similar ICP-baseline identification performance (78.1%) is reported in the work of Chang *et al.* [23] who use a very similar experimental setting on the FRGC v2.0 database (see Table XI for the details of their experimental setup). It should be noted that the PC-XYZ algorithm uses LDA feature extraction which may explain this performance improvement. Our AFM-based registration scheme orders the 3-D features, which turns out to be beneficial since it vectorizes unordered features. This enables the application of any statistical feature extraction algorithm such as the LDA on the ordered 3-D feature vectors, which would not be possible in the traditional 1 : 1 ICP-based matching.

C. *Comparative Analysis of Fusion Methods*

In this section, we discuss the impact of the decision fusion methods in improving the recognition performance. Table V presents the rank-1 correct classification accuracies of different fusion techniques for the four experimental configurations: E_1, E_2, E_3, and E_4. In each fusion method, all of the 16 base face experts listed in Table III are combined. In the columns of Table V, the numbers in parentheses denote the classification rate improvement (or loss) with respect to the best individual face expert in that experimental category. For example, in experiment E_1, the best face expert (CURV-PD) obtains 91.88% classification rate, and all improvement figures in the E_1 column of Table V are calculated with respect to this baseline. To facilitate comparisons, these base expert accuracies are shown in the second row of the table with legend "Best Individual." For each experiment, the best fusion accuracies are highlighted in boldface. As expected, the fusion gains diminish from harder toward simpler experiments, i.e., from E_1 to E_4. Therefore, it makes sense to analyze the advantage of fusion methods, particularly for the most challenging experiment—the E_1.

One can generally observe that fusion may cause significant classifier performance losses with respect to the best expert's if the fusion method is not judiciously chosen. On the other hand, the contribution of fusion methods remains modest even in the most difficult experiment. More specifically, the sum rule, which is the most widely used fusion technique in the 3-D face community, reports a slight improvement of 0.15%. The product rule performs very badly in E_1, but contributes positively in other experiments, pointing out again to the singularity of the single-gallery experiment. This performance degradation in E_1 is due to the insufficient amount of training samples to estimate the score range with the min–max technique. If the training set is small, the estimated normalization parameters do not generalize well in the identification phase. This problem does not occur in experiments E_2, E_3, and E_4 where sufficient training data exist.

Min, max, median, and Borda-count methods do not surpass the accuracy of the best face expert in the respective experiments. A few words of comment are in order for the Borda-count method. Should one use all the possible ranks from one up to the subject size or the top-ranking ones? We have observed that combining the ranks of the top three achieves better results as compared to combining, for example, those of 195 subjects or any other subset. The fusion loss, as in Table V, becomes, with this top-three Borda, a fusion gain of 1.97, 0.64, 0.16, and 0.00 points for the four experiments E_1, E_2, E_3, and E_4, respectively. Apparently, the most important information is in the top-ranking face experts.

Plurality voting, despite its simplicity, performs very well. For example, it improves the best expert's classification rate by 1.52% in E_1. The modified plurality (MOD-PLUR), presented in Section VI, performs better than its classical version (cf. MOD-PLUR with PLUR). The advantage of using confidence-aided fusion becomes more evident when we compare it with the performance of the HC fusion rule. Essentially, the HC rule is a classifier-selection method similar to the frequently used MIN fusion rule. The only difference between the HC and MIN rules is that the HC rule uses confidences to reach a decision, whereas the MIN rule uses normalized scores to select the final class label. For experiment E_1, the MIN rule has 88.54% identification rate. However, the HC rule obtains 93.32% identification rate which is better than the best face expert in the ensemble by 1.44%. This observation is very important and shows the superiority of the confidence-assisted fusion scheme. To recap, we recommend the use of the relative distance between the first and the second class candidates, if correctly normalized score measurements are not available, which is usually the case in single-gallery experiments. Table VI shows the performance improvement due to fusion of all 16 face experts in single-gallery experiment of FRGC v2.0. This time, enough training data are available for normalization: min–max score normalization parameters are calculated from the whole v1.0 database. It is seen that with the correct estimation of normalized scores, the fixed combination rules, such as the sum and product rules, offer the best improvements: they obtain 93.56% and 93.08% identification rates, which are 5.25% and 4.77% better, respectively, than the best individual face expert (PC-ICA).

VIII. WHICH EXPERTS TO INVITE FOR CONSULTATION

A. *Classifier Selection by Sequential Floating Backward Search (SFBS)*

In Section VII-C, we have fused the decisions of all the 16 face experts. However, it is not immediately obvious whether including all experts in a fusion scheme is the best scheme to follow simply because these individual experts may be

TABLE VI
RANK-1 IDENTIFICATION PERFORMANCES (IN PERCENT) OF THE FUSION METHODS ON THE FRGC V2.0 DATABASE

Fusion Method	Fused Experts	Accuracy	Improvement
Best Expert		88.31	
SUM	All	93.56	5.25
PRODUCT	All	93.08	4.77
PLUR	All	92.12	3.81
BORDA COUNT	All	92.24	3.93
MOD. PLUR	All	92.63	4.32
HIGH. CONF.	All	90.01	1.70
SUM (SFBS selection - v2.0 db)	{PC-XYZ, SN, CURV-PD, DI-DCT, PC-ICA, TEX-GABOR, TEX-PIXEL}	95.45	7.14
SUM (SFBS selection - v1.0 db)	{PC-XYZ, CURV-SI, CURV-PD, CURV-K, TEX-PIXEL, TEX-GABOR, DI-ICA, PC-NMF}	94.18	5.87

correlated; hence, they may not be proffering useful information. One idea is to use a classifier-selection method to design a better ensemble of experts [53]. The brute-force solution would be to construct all possible ensembles and select the best one. However, this is not practical given the number of combinations. Therefore, several heuristics such as incremental addition [54], pruning [55]–[57], and evolutionary algorithms [58], [59] are proposed in the classifier-selection literature.

We use a similar approach and formulate the classifier-selection problem as a feature-selection problem. In analogy to the feature-selection methods, we consider each classifier as a feature and apply the SFBS [60] to find the near-optimal subset for each fusion technique. The SFBS-based classifier-selection algorithm can be stated as follows.

1) Initialization step: Start with the total ensemble set (Ω_{in}) of all the face experts: $\Omega_{\text{in}} = \{e_1, \ldots, e_n\}$. Set the discarded face-expert subset to an empty set: $\Omega_{\text{out}} = \emptyset$.
2) Exclusion step: For each face expert $e_i \in \Omega_{\text{in}}$, remove this expert from Ω_{in} and obtain the candidate subset $\Omega_{\text{cand}} = \{\Omega_{\text{in}} - e_i\}$. Calculate the classification rate of the candidate ensembles. Select the candidate subset, which produces the best classification rate (Ω^*_{cand}). If the accuracy of the selected candidate is greater than or equal to the accuracy of the set Ω_{in}, then perform the following updates: $\Omega_{\text{in}} = \Omega^*_{\text{cand}}$, and $\Omega_{\text{out}} = \{\Omega_{\text{out}} \bigcup e_i\}$. Otherwise, stop.
3) Inclusion step (If the cardinality of the $\Omega_{\text{out}} > 2$): Form a candidate subset Ω_{can} by including a single face expert e_i from the previously discarded face-expert subset Ω_{can}: $\{\Omega_{\text{in}} \bigcup e_i\}$. If the classification rate of the candidate set Ω_{can} is better than the accuracy of the set Ω_{in}, then include expert e_i to the subset $\Omega_{\text{in}} = \{\Omega_{\text{in}} \bigcup e_i\}$ and remove e_i from the $\Omega_{\text{out}} = \{\Omega_{\text{out}} - e_i\}$. Try all of the remaining experts in the subset Ω_{out} to include to Ω_{in} in this fashion.
4) Try the exclusion and inclusion steps successively until there is no performance improvement. Output the subset Ω_{in}.

We have applied the SFBS algorithm to SUM, PRODUCT, PLUR, HC, and MOD-PLUR fusion schemes and found near-optimal subsets for E_1. The results are shown in Table VII. The second column of Table VII shows the selected face experts in the found subsets. It is clear from the classification accuracies of experiment E_1 that it is possible to get better ensembles in terms of identification performance. For example, the MOD-PLUR fusion rule attains 95.22% identification rate by combining only eight face experts, which is significantly better than using all of the 16 face experts in the original MOD-PLUR method (93.63%). It should be noted that these subsets were found by applying SFBS for experiment E_1, and the recognition accuracies of the other experiments were reported for these specific subsets. This explains the performance degradation of the PRO fusion rule in E_4. We have chosen to report the accuracies for experiments E_2, E_3, and E_4 in order to test the generalization ability of the SFBS-based classifier-selection algorithm. It is more appropriate to apply the SFBS algorithm to a separate validation set and then to report the final classification rates on an independent test set. However, our main concern in this paper is not to design a classifier-selection algorithm but to give a proof of the concept that it is possible to construct better ensembles without using all of the available face experts. However, for the FRGC v2.0 experiments, we can measure the generalization ability of SBFS-based ensemble construction by selecting the best classifier ensemble using the v1.0 database and then by reporting its identification accuracy on the v2.0 set.

Floating search-based ensemble construction also offers improved accuracy for the FRGC v2.0 experiments. The ninth row of Table VI shows the selected experts found by the SFBS algorithm on the v2.0 database for the SUM rule method in experiment E'_1. These seven classifiers attain 95.45% identification rate which is 7.14% better than the best single face expert. It is worthwhile to note that TEX-PIXEL approach, which can be considered as a weak classifier, is included in the best ensemble subset. The reason for this behavior will be clearer in the next section when we perform correlation analysis of the base face experts. The last row of Table VI shows the generalization ability of the SFBS-based ensemble formation. Here, the classifier subset is selected as the optimum one (SUM rule) from the independent v1.0 set (see Table VII, SUM rule). This ensemble, which is trained on the v1.0 database, achieves 94.18% accuracy on the v2.0 database. Hence, it is still superior to fusing all individual experts within v2.0 by 0.42 percentage points and also improves the best single individual expert's performance by 5.87%. This finding clearly supports our claim that judicious selection of classifiers for score fusion can be quite beneficial.

TABLE VII
SELECTED CLASSIFIER SUBSETS AND THEIR IDENTIFICATION PERFORMANCES (IN PERCENT) FOR DIFFERENT FUSION METHODS ON THE FRGC V1.0 DATABASE

Fusion	Expert Subset	E_1	E_2	E_3	E_4
SUM	PC-XYZ, CURV-SI, CURV-PD, CURV-K, TEX-PIXEL, TEX-GABOR, DI-ICA, PC-NMF	94.23 (2.35)	99.14 (0.43)	99.75 (0.08)	**100.00 (0.00)**
PRO	PC-XYZ, CURV-SI, CURV-PD, CURV-H, CURV-K, TEX-GABOR, DI-DFT, DI-ICA, DI-NMF	88.39 (-3.49)	99.14 (0.43)	99.75 (0.08)	99.89 (-0.11)
PLUR	CURV-SI, CURV-PD, CURV-K, TEX-PIXEL, TEX-GABOR, DI-DCT, PC-ICA	94.84 (2.96)	99.28 (0.57)	99.75 (0.08)	**100.00 (0.00)**
HC	CURV-SI, CURV-PD, CURV-H, TEX-GABOR, PC-ICA	94.69 (2.81)	98.99 (0.28)	99.42 (-0.25)	**100.00 (0.00)**
MOD-PLUR	PC-XYZ, CURV-SI, CURV-PD, TEX-PIXEL, TEX-GABOR, DI-DFT, DI-NMF, PC-ICA	**95.22 (3.34)**	**99.50 (0.79)**	**99.92 (0.25)**	**100.00 (0.00)**

TABLE VIII
2×2 PROBABILITY COMPUTATION FOR TWO CLASSIFIERS C_i AND C_j, WHERE $N_{11} + N_{01} + N_{10} + N_{00} = 1$

	C_j correct (1)	C_j wrong (0)
C_i correct (1)	N_{11}	N_{10}
C_i wrong (0)	N_{01}	N_{00}

B. Correlation Analysis of Face Experts

The SFBS-based construction of the face ensembles has shown that some of the base classifiers are redundant and that their inclusion may lead to suboptimal identification rates. To substantiate this finding, we consider the correlation of binary decision outputs of the face experts [61], which is computed as follows:

$$\rho_{i,j} = \frac{N_{11}N_{00} - N_{10}N_{01}}{\sqrt{(N_{11}+N_{10})(N_{01}+N_{00})(N_{11}+N_{01})(N_{10}+N_{00})}} \quad (3)$$

where, for classifiers C_i and C_j, N values denote the probabilities for the respective pair of correct/incorrect outputs and can be calculated as in Table VIII.

Given the 16 experts in Table III, we have computed 120 pairwise correlation values, with significant correlations between certain pairs of face experts. In order to visualize the multidimensional relationships between face experts, we have first obtained the dissimilarities between pairs of classifiers by using $d_{i,j} = 1 - \rho_{i,j}$. Here, $d_{i,j}$ can be considered as a distance measure between classifier pairs. Then, we applied MDS algorithm to construct a 2-D space $\Re^2$ where the coordinates denote the individual classifiers. The space of the two largest eigenvectors suffices to reasonably reproduce the space of face experts. Fig. 9(a) shows the reproduced face experts for the FRGC v1.0 experiments in 2-D coordinate system as black dots. In Fig. 9(a), the visually delineated face expert clusters (dashed ellipses) are also depicted. There are five salient clusters, and with few exceptions, each cluster matches one of the face representation methods. For instance, curvature-, depth-image-, point-cloud-, and texture-based face experts form their own clusters. Fig. 9(b) shows the same analysis for the face experts in the FRGC v2.0 experiments. Note that, now, nearly all of the face experts employ subspace techniques in their feature extraction process. Visual inspection of the clusters reveals the same conclusion for the base experts in the FRGC v2.0 experiments: Groups are formed according to the representation techniques used and not according to the feature extraction methods employed.

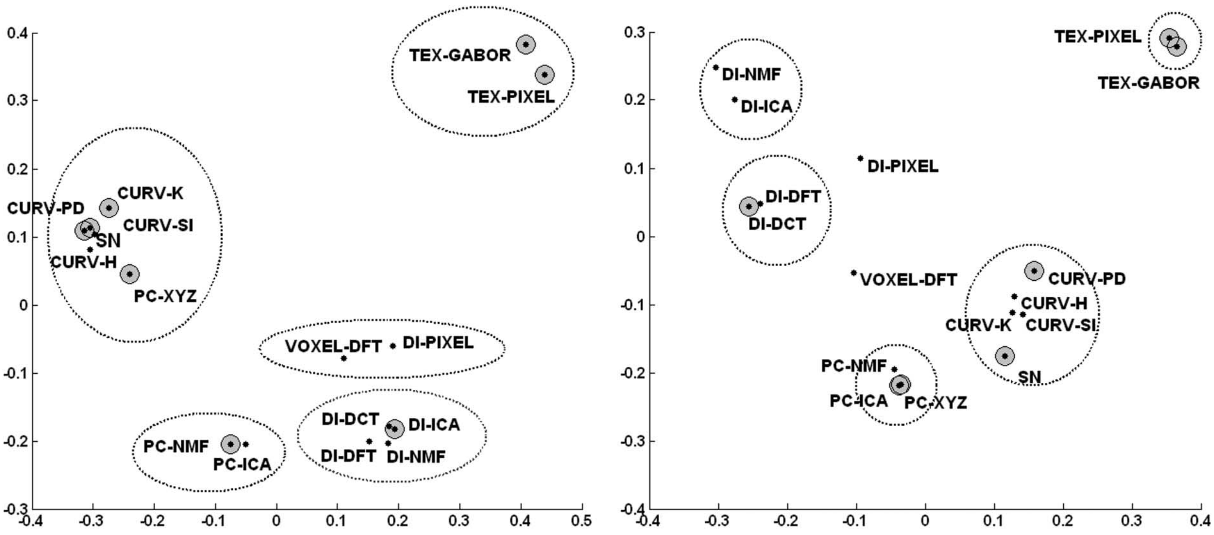
In Fig. 9(a), gray circles denote the selected face classifiers for the SUM fusion rule (see Table VII, second row). Examination of the selected face experts reveals that experts must come from different clusters. Similarly, for the FRGC v2.0 experiments, the best ensemble subset includes experts from curvature-, depth-image-, point-cloud-, and texture-based representations [Fig. 9(b)].

We can profit from this correlation map of experts to construct an alternative expert subset, i.e., a fusion ensemble. This method consists in handpicking one classifier from each cluster, as in Fig. 9, in order to enforce diversity of opinions. The heuristic is the greedy approach of choosing the best performing face expert in clusters. The upper part of Table IX shows the performance of the five fusion rules applied to the handpicked ensembles, namely, CURV-PD, TEX-GABOR, DI-DCT, and PC-ICA for the FRGC v1.0 experiment—the E_1. We exclude the VOXEL-DFT method because its solo classification performance is very low. Comparison of these results reveals the advantage of getting guidance from clustering of experts. The only drop in accuracy occurs for the PLUR method since voting within an ensemble of small cardinality is known not to perform well. The lower part of Table IX shows the identification performances of selecting the best expert from each cluster for the FRGC v2.0 experiment. The selected experts are the same as the ones used in the FRGC v1.0 experiments. It is reconfirmed that selecting the best experts from different clusters is not only less costly but is also better in terms of performance.

We conclude then that the set of experts yielding the highest score does not necessarily contain experts only with high individual scores, but that experts in a consultation team should be from different categories even if their individual scores are lower.

C. Classifier Selection by Best-N Method

A second alternative method to construct the ensemble would be to combine the best N face experts. However, since this method does not exploit the diversity of decision takers, we conjecture that it may not perform as well. In order to validate



(a) (b)

Fig. 9. Correlation analysis of the face experts. In both images, black dots denote the 2-D positions of the face experts calculated from the MDS analysis of pairwise correlations. Gray circles denote the expert subset found from (a) the SUM rule for FRGC v1.0 experiment E_1 and (b) the SUM rule for FRGC v2.0 experiments. In (a), large dashed ellipses denote visually salient clusters.

TABLE IX
IDENTIFICATION PERFORMANCES (IN PERCENT) WITH SELECTION OF CLASSIFIER ENSEMBLES BY CLUSTERING

	SUM	PRO	PLUR	HC	MOD-PLUR
			FRGC v1.0		
All 16 experts	92.03	72.23	93.40	93.32	93.63
Best of Clusters	92.26	85.82	91.35	94.23	94.01
			FRGC v2.0		
All 16 experts	93.56	93.08	92.12	90.01	92.63
Best of Clusters	94.55	94.35	91.19	91.64	92.97

this hypothesis, we have combined the best N face experts, where $N \in \{2, 3, \ldots, 16\}$, with MOD-PLUR fusion rule in FRGC v1.0 experiments. The fusion performance of the best ensembles is shown in Fig. 10(a) (black dotted curve), whereas the two lines correspond to the accuracies of the total ensemble (the $N = 16$ case) and the SFBS ensemble, respectively. As expected, the indiscriminate ensemble of the best ones performs worse than the judiciously chosen SFBS subset case. Fig. 10(b) shows the performance behavior of the best N approach for the SUM rule in the FRGC v2.0 experiment. Although the performance gradually improves by adding new base experts, it is always suboptimal when compared to the SFBS-ensemble performance.

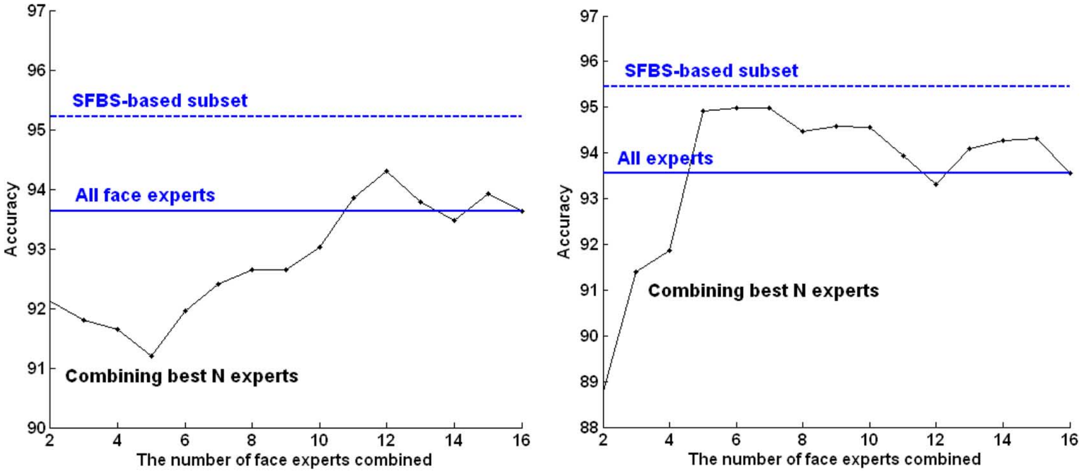
In the 3-D face recognition literature, the fusion of only two experts, one for the shape modality and one for the texture modality [13], [26], [27], is a common method. For completeness, we also present the results of this restricted fusion scheme. In Table X, we provide the results of combining best texture- and shape-based experts for the FRGC v1.0 and v2.0 experiments. From these performance figures, one can see that the combination of texture and shape experts has comparable accuracy to that of fusing all 16 face experts but does not perform as good as the best ensemble subset that employs diverse representations.

D. Overall Comparison of Fusion Schemes and Classifier Selection Methods

The bar chart in Fig. 11 shows the fusion results for the v1.0 experiments, where four fusion schemes (SUM, PLUR, HC, and MOD-PLUR) and four ensemble-construction algorithms (Ens-All, Ens-SFBS, Ens-Cluster, and Ens-BestN) are presented. The optimal ensemble formation is given by the SFBS algorithm. The choices of SFBS indicate the importance of diversity and complementariness in ensemble formation. In fact, the cluster-guided method (Ens-Cluster) satisfies the diversity conditions and hence performs better than fusing the best N individual experts (see Ens-BestN in Fig. 11). These findings are also proved on a bigger database (FRGC v2.0) where SFBS-based ensemble construction algorithm selects diverse experts and obtains the best identification accuracy.

In terms of fusion algorithms, the following conclusions can be drawn according to the availability of sufficient training data. If you have insufficient training data that lead to suboptimal estimation of score ranges: 1) in good ensembles, the SUM rule does not perform as well as the others; 2) if one has several face experts, plurality voting can be a better alternative to the SUM rule; 3) it is possible to improve plurality voting with the aid of confidence weights; 4) if there are few experts, then selecting the class having the HC (not the smallest score or distance) can lead to better identification rate than plurality voting. Otherwise, if you have enough training data to estimate correct score ranges, then the use of fixed rules such as the SUM or PRODUCT could be a better alternative.

Table XI illustrates the performances of different algorithms in the literature, which use FRGC v2.0 for identification simulations. In all of these systems, the performance of the proposed approach is benchmarked via single-gallery experiments where the earliest scans of each subject are placed into the gallery set. However, the experimental setups are different with different sizes of gallery and probe sets. In this respect, the experimental protocol used by Chang *et al.* [23] is more challenging since



(a) (b)

Fig. 10. Identification performances of fusing best N classifiers: (a) MOD-PLUR rule for the FRGC v1.0 experiments and (b) SUM rule for the FRGC v2.0 experiment. The x-axis denotes the number of classifiers fused, whereas the y-axis denotes the identification accuracy. Black dotted curve denotes the rank-1 accuracy of the best N fusion method. The horizontal dashed line and the horizontal solid line denote the performances of the fused ensemble for 1) SFBS-based face expert subset and 2) using all 16 face experts in the ensemble, respectively.

TABLE X
IDENTIFICATION PERFORMANCES (IN PERCENT) WITH FUSION OF SHAPE- AND TEXTURE-BASED EXPERTS USING THE SUM RULE

FRGC v1.0		FRGC v2.0	
TEX-GABOR + CURV-PD	93.63	TEX-GABOR + PC-ICA	93.42

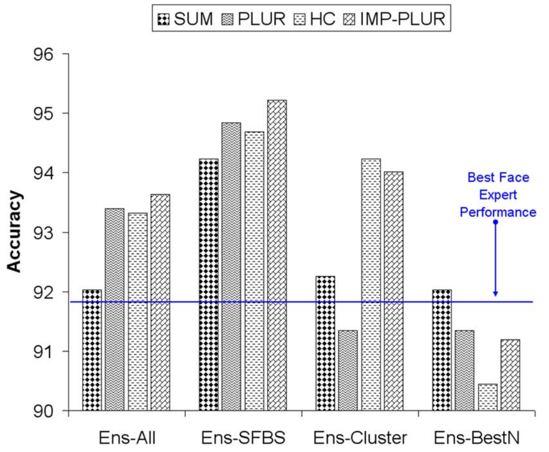


Fig. 11. Overall comparison of 1) the fusion techniques and 2) ensemble-construction methods. Ens-All: All 16 experts in the ensemble, Ens-SFBS: Subset of face experts selected by the SFBS method, Ens-Cluster: Selection of best experts from each cluster (see Fig. 9), and Ens-BestN: Selection of the most accurate five face experts. Horizontal line denotes the best face expert's accuracy in the experiment E_1.

they conducted recognition experiments on a larger database spanning both FRGC v1.0 and v2.0 image sets. Furthermore, their results are obtained via a fully automatic face recognition system, whereas our system employs manual landmarking for registration. Thus, the performance figures should be compared with respect to the relative difficulty of each experimental setup.

IX. CONCLUSION

In this paper, we have analyzed the impact of face data representation, feature selection, and fusion for 2-D/3-D face recognition. We have designed a diverse set of 2-D/3-D face recognizers that differ in the face representation and/or in the discriminative features they extract from these representations. Recall that one combination of face representation and feature type is denoted as a "face expert." We have demonstrated that it is possible to improve the recognition performance with a consultation session between these experts, where expert decisions are fused.

We have conducted our experiments on the FRGC v1.0 and v2.0 data sets. We have used the experimental configurations used by the most recent studies. In the experiments on the FRGC v1.0 data set, we have used all experimental configurations E_i. However, in FRGC v2.0, we restricted our attention to E_1 experiments, where the gallery contains a single training image per subject. In the experiments with FRGC v2.0, we have used the FRGC v1.0 data set to learn feature subspaces and selections for expert consultations. We have conducted extensive experiments on the effectiveness of different features, different representations, and different fusion rules. By experimenting with different training-set sizes, we were able to draw conclusions on the effect of training sets. The expert-selection experiments led to an understanding on the effect of combining diverse experts. Our insights and conclusions can be summarized as follows.

1) Representation is more important when training set is small. The acquired face data in 3-D can assume one of the forms of point clouds, surface normals, depth images, curvatures, or 3-D voxels. In 2-D, it assumes the form of gray-level texture images. The depth image derived from the original 3-D face is also treated as a 2-D image. In experiments where the training-set size is very small, the effect of representation type dominates: Curvature-based experts always scored the best, and 2-D textures always remained inferior.
2) The effect of matching feature dominates when training-set size gets larger. The second tier of the analysis is the feature extraction stage. For 3-D face data, we have used two varieties of features, namely, the subspace features

TABLE XI
COMPARISON OF RANK-1 IDENTIFICATION RATES IN THE LITERATURE FOR SINGLE-GALLERY EXPERIMENTS ON THE FRGC V2.0 DATABASE (NOTE THAT THE EXPERIMENTAL SETUP IN EACH STUDY IS DIFFERENT)

Reference	Number of gallery faces	Number of probe faces	Landmarking scheme	Rank-1 Accuracy
Passalis et al. [19]	466	3541	Automatic	89.5
Chang et al. [23]	449	3939	Automatic	91.9
Chang et al. [23]	449	3939	Manual	92.9
Faltemier et al. [62]	410	3541	Automatic	94.9
Our approach (SFBS-based ensemble)	410	3542*	Manual	95.45

* After the original distribution of FGRC database, several incorrect subject IDs were fixed. After correction, there are 465 subjects where 55 of them have only one scan and the total number of probe images (using subjects having at least two scans) is 3542.

(DFT, DCT, NMF, and ICA) and the spatial geometric features (point cloud, shape index, surface normals, and principal curvatures). For 2-D data, we have limited ourselves to Gabor texture features. Subspace-based methods such as application of ICA and NMF on point clouds gave superior results when a large training set is available. One important conclusion is that all 3-D face representation types (point clouds, surface normals, depth images, curvature images, and 3-D voxels) have similar identification performances provided that its matching feature is selected and that the gallery contains at least two data items per subject. Instances of a matching feature are the following: DCT or DFT features for depth images, shape index for curvature representation, and NMF for point cloud. In our experiments, the only exception to this rule was with 2-D textures: the solo performance of 2-D textures was inferior for any matching feature. However, they were useful as the assisting features in a fusion setting.

3) Fusion of experts is beneficial, but normalization is critical. The availability of multiple features and representation allows performance improvement via fusion. However, expert scores can be very diverse, and score normalization may not work when training-set size is small. We solved this disparity problem by using range normalization and a differential measure. Another conclusion was that the simple yet effective plurality voting can be improved by taking into consideration the expert confidences.
4) Fuse only the most diverse. We came to the conclusion that the face experts taking role in the decision session should have origins from different face representations and not from different features of the same representation. Interestingly enough, the experts cluster in a 2-D map after MDS according to their underlying representation data. The fusion of intelligently selected experts helps most in the challenging single training case, where additional 1.75 and 7.14 points of accuracy are gained for FRGC v1.0 and v2.0, respectively. It was shown that inviting everybody is not necessarily a good idea and that an expert-selection algorithm, much in the same way a feature selection, works better.

As future work, the optimal partitioning of the face in patches, with possibly smaller patches near the eyes and larger patches on the front or cheeks, must be explored. Each such patch will be represented by its DFT/DCT features locally and then concatenated into a face feature vector. NMF [63] and ICA can similarly benefit from patch-based analysis. In addition, we have observed that the depth field suffers from the unnatural acceleration at the border when cheeks meet background abyss. The effect of this artifact can be mitigated with some windowing function. We have considered in this paper only score-, rank-, and decision-level fusion. The merit, if any, of the feature-level fusion should be analyzed. In fact, it would be conceivable to design 3-D face recognition schemes that exploit feature- and decision-level fusion in some optimum combination.

REFERENCES

[1] A. Colombo, C. Cusano, and R. Schettini, “3D face detection using curvature analysis,” *Pattern Recognit.*, vol. 39, no. 3, pp. 444–455, Mar. 2006.

[2] A. Mian, M. Bennamoun, and R. Owens, “Automatic 3D face detection, normalization and recognition,” in *Proc. 3rd Int. Symp. 3DPVT*, 2006, pp. 735–742.

[3] S. Malassiotis and M. G. Strintzis, “Robust real-time 3D head pose estimation from range data,” *Pattern Recognit.*, vol. 38, no. 8, pp. 1153–1165, Aug. 2005.

[4] A. Mian, M. Bennamoun, and R. Owens, “Matching tensors for pose invariant automatic 3D face recognition,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 2005, vol. 3, p. 120.

[5] A. Mian, M. Bennamoun, and R. Owens, “Face recognition using 2D and 3D multimodal local features,” in *Proc. ISVC*, 2006, pp. 860–870.

[6] C. Samir, A. Srivastava, and M. Daoudi, “Three-dimensional face recognition using shapes of facial curves,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1858–1864, Nov. 2006.

[7] B. Gökberk, M. O. İrfanoğlu, and L. Akarun, “3D shape-based face representation and facial feature extraction for face recognition,” *Image Vis. Comput.*, vol. 24, no. 8, pp. 857–869, 2006.

[8] H. Dutağacı, B. Sankur, and Y. Yemez, “3D face recognition by projection-based features,” in *Proc. SPIE—Conf. Electronic Imaging: Security, Steganography, Watermarking Multimedia Contents*, 2006, pp. 194–204.

[9] K. W. Bowyer, K. Chang, and P. Flynn, “A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition,” *Comput. Vis. Image Underst.*, vol. 101, no. 1, pp. 1–15, Jan. 2006.

[10] G. Pan, S. Han, Z. Wu, and Y. Wang, “3D face recognition using mapped depth images,” in *Proc. IEEE Workshop Face Recog. Grand Challenge Experiments*, 2005, p. 175.

[11] T. Russ, M. Koch, and C. Little, “A 2D range Hausdorff approach for 3D face recognition,” in *Proc. IEEE Workshop Face Recog. Grand Challenge Experiments*, 2005, p. 169.

[12] A. Abate, M. Nappi, S. Ricciardi, and G. Sabatino, “Fast 3D face recognition based on normal map,” in *Proc. IEEE Int. Conf. Image Process.*, 2005, vol. 2, pp. 946–949.

[13] X. Lu, A. Jain, and D. Colbry, "Matching 2.5D face scans to 3D models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 31–43, Jan. 2006.

[14] G. Pan and Z. Wu, "3D face recognition from range data," *Int. J. Image Graph.*, vol. 5, no. 3, pp. 573–593, 2005.

[15] T. Papatheodorou and D. Reuckert, "Evaluation of automatic 4D face recognition using surface and texture registration," in *Proc. 6th Int. Conf. Autom. Face Gesture Recog.*, 2004, pp. 321–326.

[16] B. Gökberk, A. A. Salah, and L. Akarun, "Rank-based decision fusion for 3D shape-based face recognition," in *Proc. Audio-Video-Based Biometric Person Authentication* T. Kanade, A. Jain, and N. K. Ratha, Eds., 2005, vol. 3456, pp. 1019–1029.

[17] M. O. İrfanoğlu, B. Gökberk, and L. Akarun, "3D shape-based face recognition using automatically registered facial surfaces," in *Proc. Int. Conf. Pattern Recog.*, 2004, pp. 183–186.

[18] M. Koudelka, M. Koch, and T. Russ, "A prescreener for 3D face recognition using radial symmetry and the Hausdorff fraction," in *Proc. IEEE Workshop Face Recog. Grand Challenge Experiments*, 2005, p. 168.

[19] G. Passalis, I. Kakadiaris, T. Theoharis, G. Toderici, and N. Murtuza, "Evaluation of 3D face recognition in the presence of facial expressions: An annotated deformable model approach," in *Proc. IEEE Workshop Face Recog. Grand Challenge Experiments*, 2005, p. 171.

[20] X. Lu and A. K. Jain, "Deformation modeling for robust 3D face matching," in *Proc. IEEE Comput. Soc. Conf. CVPR*, 2006, pp. 1377–1383.

[21] A. Bronstein, M. Bronstein, and R. Kimmel, "Three-dimensional face recognition," *Int. J. Comput. Vis.*, vol. 64, no. 1, pp. 5–30, Aug. 2005.

[22] [Online]. Available: http://www.sic.rma.ac.be/~beumier/DB/3d_rma

[23] K. Chang, K. Bowyer, and P. Flynn, "Adaptive rigid multi-region selection for handling expression variation in 3D face recognition," in *Proc. IEEE Workshop Face Recog. Grand Challenge Experiments*, 2005, p. 157.

[24] C. Xu, Y. Wang, T. Tan, and L. Quan, "Automatic 3D face recognition combining global geometric features with local shape variation information," in *Proc. Int. Conf. Autom. Face Gesture Recog.*, 2004, pp. 308–313.

[25] K. I. Chang, K. W. Bowyer, and P. J. Flynn, "An evaluation of multimodal 2D + 3D face biometrics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 4, pp. 619–624, Apr. 2005.

[26] C. Ben Abdelkader and P. A. Griffin, "Comparing and combining depth and texture cues for face recognition," *Image Vis. Comput.*, vol. 23, no. 3, pp. 339–352, Mar. 2005.

[27] Y. Wang and C.-S. Chua, "Robust face recognition from 2D and 3D images using structural Hausdorff distance," *Image Vis. Comput.*, vol. 24, no. 2, pp. 176–185, Feb. 2006.

[28] Y. Wang and C.-S. Chua, "Face recognition from 2D and 3D images using 3D Gabor filters," *Image Vis. Comput.*, vol. 23, no. 11, pp. 1018–1028, Oct. 2005.

[29] M. Husken, M. Brauckmann, S. Gehlen, and C. von der Malsburg, "Strategies and benefits of fusion of 2D and 3D face recognition," in *Proc. IEEE Workshop Face Recog. Grand Challenge Experiments*, 2005, p. 174.

[30] A. A. Salah, H. Çınar, L. Akarun, and B. Sankur, "Robust facial landmarking for registration," *Ann. Telecommun.*, vol. 62, no. 1/2, pp. 1608–1633, 2007.

[31] H. Çınar, A. A. Salah, L. Akarun, and B. Sankur, "2D/3D facial feature extraction," in *Proc. SPIE—Conf. Electronic Imaging*, 2006, pp. 441–452.

[32] A. Srivastava, X. Liu, and C. Hesher, "Face recognition using optimal linear components of range images," *Image Vis. Comput.*, vol. 24, no. 3, pp. 291–299, Mar. 2006.

[33] F. Stein and G. Medioni, "Structural indexing: Efficient 3-D object recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 125–145, Feb. 1992.

[34] C. Dorai and A. K. Jain, "Cosmos—A representation scheme for 3D free-form objects," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 10, pp. 1115–1130, Oct. 1997.

[35] F. Mokhtarian, N. Khalili, and P. Yuen, "Multi-scale free-form 3D object recognition using 3D models," *Image Vis. Comput.*, vol. 19, no. 5, pp. 271–281, Apr. 2001.

[36] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin, "Matching 3D models with shape distributions," in *Proc. IEEE Int. Conf. SMI*, 2001, pp. 154–166.

[37] T. Zaharia and F. Preteux, "Shape-based retrieval of 3D mesh models," in *Proc. IEEE ICME*, 2002, pp. 437–440.

[38] C. B. Akgül, B. Sankur, Y. Yemez, and F. Schmitt, "Density-based 3D shape descriptors," *EURASIP J. Advances Signal Process.*, vol. 2007, p. 16, 2007. Article ID 32 503, DOI:10.1155/2007/32503.

[39] J. J. Koenderink and A. J. van Doorn, "Surface shape and curvature scales," *Image Vis. Comput.*, vol. 10, no. 8, pp. 557–565, Oct. 1992.

[40] A. Hyvarinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Netw.*, vol. 13, no. 4/5, pp. 411–430, May/Jun. 2000.

[41] D. Lee and H. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.

[42] D. Lee and H. Seung, "Algorithms for nonnegative matrix factorization," in *Proc. Advances Neural Inf. Process. Syst.*, 2001, vol. 13, pp. 556–562.

[43] L. Wiskott, J.-M. Fellous, N. Kuiger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 775–779, Jul. 1997.

[44] C. Liu, "Gabor-based kernel PCA with fractional power polynomial models for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 5, pp. 572–581, May 2004.

[45] B. Gökberk, L. Akarun, and E. Alpaydın, "Selection of kernel location, frequency, and orientation parameters of 2-D Gabor wavelets for face recognition," *Perception (Absract)*, vol. 32, p. 1, 2003.

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

[47] D. H. Wolpert, "Stacked generalization," *Neural Netw.*, vol. 5, no. 2, pp. 241–259, 1992.

[48] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, "Adaptive mixtures of local experts," *Neural Comput.*, vol. 3, no. 1, pp. 79–87, 1991.

[49] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, Aug. 1996.

[50] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Proc. ICML*, 1996, pp. 148–156.

[51] K. Chang, K. W. Bowyer, and P. J. Flynn, "Face recognition using 2D and 3D facial data," in *Proc. ACM Workshop Multimodal User Authentication*, 2003, pp. 25–32.

[52] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *Proc. Int. Conf. Comput. Vis. Pattern Recog.*, 2005, pp. 947–954.

[53] C. Demir and E. Alpaydın, "Cost-conscious classifier ensembles," *Pattern Recognit. Lett.*, vol. 26, no. 14, pp. 2206–2214, Oct. 2005.

[54] R. Caruana, A. Niculescu-Mizil, G. Crew, and A. Ksikes, "Ensemble selection from libraries of models," in *Proc. ICML*, 2004, p. 18.

[55] G. Giacinto and F. Roli, "Design of effective neural network ensembles for image classification," *Image Vis. Comput. J.*, vol. 19, no. 9/10, pp. 697–705, 2001.

[56] D. D. Margineantu and T. G. Dietterich, "Pruning adaptive boosting," in *Proc. ICML*, 1997, pp. 211–218.

[57] A. L. Prodromidis and S. J. Stolfo, "Cost complexity-based pruning of ensemble classifiers," *Knowl. Inf. Syst.*, vol. 3, no. 4, pp. 449–469, Nov. 2001.

[58] D. Ruta and B. Gabrys, "Classifier selection for majority voting," *Inf. Fusion*, vol. 6, no. 1, pp. 63–81, Mar. 2005.

[59] Z.-H. Zhou, J. Wu, and W. Tang, "Ensembling neural networks: Many could be better than all," *Artif. Intell.*, vol. 137, no. 1, pp. 239–263, May 2002.

[60] A. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-22, no. 1, pp. 4–37, Jan. 2000.

[61] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. Hoboken, NJ: Wiley, 2004.

[62] T. Faltemier, K. Bowyer, and P. Flynn, "3D face recognition with region committee voting," in *Proc. 3rd Int. Symp. 3D Data Process., Vis. Transmiss.*, 2006, pp. 318–325.

[63] S. Z. Li, X. W. Hou, H. J. Zhang, and Q. S. Cheng, "Learning spatially localized, parts-based representation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2001, pp. 207–212.

Berk Gökberk received the B.S., M.Sc., and Ph.D. degrees in computer engineering from Boğaziçi University, Istanbul, Turkey, in 1999, 2001, and 2006, respectively.

He worked as a Research Assistant with the Department of Computer Engineering, Boğaziçi University between 1999 and 2006. He is currently with the Philips Research Laboratories, Eindhoven, The Netherlands. His research interests are in the areas of biometrics, computer vision, computer graphics, and pattern recognition.

Helin Dutağacı was born in Tunceli, Turkey, in 1976. She received the B.Sc. and M.Sc. degrees from the Boğaziçi University, Istanbul, Turkey, in 1999 and 2002, respectively, where she is currently working toward the Ph.D. degree.

She is currently a Research Assistant with the Signal and Image Processing Laboratory, Department of Electrical and Electronics Engineering, Boğaziçi University. Her major field of study is telecommunications and signal processing. Her research interests include computer vision, pattern recognition, 3-D object recognition, and biometrics.

Aydın Ulaş received the B.S. and M.Sc. degrees in computer engineering from Boğaziçi University, Istanbul, Turkey, in 1999 and 2001, respectively, where he is currently working toward the Ph.D. degree.

His research interests include machine learning and pattern recognition.

Lale Akarun (S'87–M'92–SM'02) received the B.S. and M.S. degrees in electrical engineering from Boğaziçi University, Istanbul, Turkey, in 1984 and 1986, respectively, and the Ph.D. degree from the Polytechnic University, New York, in 1992.

Since 1993, she has been a Faculty Member with the Department of Computer Engineering, Boğaziçi University. She became a Professor of computer engineering in 2001. Her research areas are face recognition, modeling and animation, and human activity and gesture analysis.

Prof. Akarun has worked on the organization committees of the IEEE NSIP99, EUSIPCO 2005, and eNTERFACE2007.

Bülent Sankur received the B.S. degree in electrical engineering from the Robert College, Istanbul, Turkey, and the M.Sc. and Ph.D. degrees from the Rensselaer Polytechnic Institute, Troy, NY.

He is currently with the Department of Electrical and Electronics Engineering Boğaziçi University, Istanbul. He has held visiting positions at the University of Ottawa, Ottawa, ON, Canada, the Technical University of Delft, Delft, The Netherlands, and the Ecole Nationale Supérieure des Télécommunications, Paris, France. His research interests are in the areas of digital signal processing, image and video compression, biometry, cognition, and multimedia systems.

Dr. Sankur was the Chairman of the International Conference on Telecommunications (ICT'96) and the European Conference on Signal Processing (EUSIPCO'05), as well as the Technical Chairman of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP'00).