Regional Registration for Expression Resistant 3-D Face Recognition

Neşe Alyüz, Berk Gökberk, and Lale Akarun, *Senior Member, IEEE*

***Abstract*—Biometric identification from three-dimensional (3-D) facial surface characteristics has become popular, especially in high security applications. In this paper, we propose a fully automatic expression insensitive 3-D face recognition system. Surface deformations due to facial expressions are a major problem in 3-D face recognition. The proposed approach deals with such challenging conditions in several aspects. First, we employ a fast and accurate region-based registration scheme that uses common region models. These common models make it possible to establish correspondence to all the gallery samples in a single registration pass. Second, we utilize curvature-based 3-D shape descriptors. Last, we apply statistical feature extraction methods. Since all the 3-D facial features are regionally registered to the same generic facial component, subspace construction techniques may be employed. We show that linear discriminant analysis significantly boosts the identification accuracy. We demonstrate the recognition ability of our system using the multiexpression Bosphorus and the most commonly used 3-D face database, Face Recognition Grand Challenge (FRGCv2). Our experimental results show that in both databases we obtain comparable performance to the best rank-1 correct classification rates reported in the literature so far: 98.19% for the Bosphorus and 97.51% for the FRGCv2 database. We have also carried out the standard receiver operating characteristics (ROC III) experiment for the FRGCv2 database. At an FAR of 0.1%, the verification performance was 86.09%. This shows that model-based registration is beneficial in identification scenarios where speed-up is important, whereas for verification one-to-one registration can be more beneficial.**

***Index Terms*—Biometrics, curvature descriptors (CDs), three-dimensional (3-D) face recognition, three-dimensional (3-D) registration.**

I. Introduction

BIOMETRIC systems try to identify human beings from their distinctive physiological and behavioral characteristics. Among popular biometric modalities such as iris, fingerprint, voice, hand geometry, and gait, human faces have several advantages which make them attractive for particular applications. Most importantly, the ease of acquiring facial images without the need of subject cooperation allows the use of face recognition systems in a diverse range of applications such as surveillance systems. However, although recent studies show that the performance of face recognition systems can reach the level of high security biometric modalities such as fingerprint and iris [1], it is still a very challenging task to recognize people from their faces under adverse scenarios. Particularly, the presence of illumination differences, in-depth pose variations, and facial expressions are important factors that affect the accuracy of a face recognition system.

With the use of three-dimensional (3-D) facial structure information, it is possible to cope with some of these challenges more efficiently compared to two-dimensional (2-D). In this work, we are interested in face recognition that exploits 3-D facial data, because examination of the 3-D surface geometry can lead to more accurate registration and recognition. By using 3-D information, the effect of illumination differences can be avoided and small pose changes can be rectified. However, regardless of whether 2-D or 3-D information is used, facial expression variations still complicate the task of identification by creating higher intraclass variance than interclass variance.

In this work, we aim to tackle these problems with the use of 1) an efficient facial surface registration approach and 2) by incorporating discriminative 3-D features. The first phase of any 3-D face recognition system, namely alignment/registration of facial surfaces, is the most crucial part and the final accuracy of the system heavily depends on the quality of the alignment module. In this paper, we propose a simple, fast, and effective region-based rigid registration approach. The novelty of the approach is that it requires a single registration for a given test face. The probe is registered in a two-pass algorithm: First, rigid registration to an average model, followed by registration to individual AvRMs. The algorithm is preceded by a novel automatic landmark localization module, which provides the initialization. The registration of facial parts to a generic model significantly speeds up the identification time because it is sufficient to perform only a single alignment to a generic model per facial region. Since all the gallery/training samples are previously registered offline to the same generic model, single alignment provides the dense correspondence information to every gallery image by default. Lastly, and most importantly, since dense correspondence is established and 3-D features are represented as an ordered feature vector, it is possible to utilize advanced pattern recognition tools either at the level of feature extraction or at the level of pattern classification. Traditional approaches like pairwise matching of two 3-D point sets are limited in that sense, since the 3-D point sets are unordered and the similarity can only be computed by means of geometrical measures.

At the second phase, after regional registration is performed, we study the benefits of using several 3-D features according

Manuscript received September 24, 2009; revised June 14, 2010; accepted June 16, 2010. Date of publication June 28, 2010; date of current version August 13, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Davide Maltoni.

N. Alyüz and L. Akarun are with the Department of Computer Engineering, Boğaziçi University, Istanbul 34342, Turkey (e-mail: nese.alyuz@boun.edu.tr; akarun@boun.edu.tr).

B. Gökberk is with the Department of Electrical Engineering Mathematics and Computer Science, University of Twente, Enschede 7500 AE, The Netherlands (e-mail: b.gokberk@utwente.nl).

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/TIFS.2010.2054081

1556-6013/$26.00 © 2010 IEEE

TABLE I
RANK-1 CORRECT CLASSIFICATION RATES REPORTED ON THE FRGCv2 FACE DATABASE. N/A ENTRIES REPRESENT CASES WHERE THE CORRESPONDING VALUES ARE EITHER NOT AVAILABLE OR NOT SPECIFIED EXPLICITLY, N AND NON-N ABBREVIATIONS STAND FOR NEUTRAL AND NON-NEUTRAL SETS, RESPECTIVELY. THE SCANS ARE LABELED WITH fs, ns, OR fns, DENOTING FIRST SCANS, NEUTRAL SCANS, AND FIRST NEUTRAL SCANS, RESPECTIVELY. VERIFICATION RESULTS FOR THE FRGCv2 EXPERIMENTAL SETUPS, NAMELY ROC I, ROC II, AND ROC III, ARE ALSO REPORTED FOR COMPARATIVE PURPOSES

	Identification Results					Verification Results		
Author,Year	Gallery Size	Probe Size	N vs. All	N vs. N	N vs. Non-N	ROC I	ROC II	ROC III
Queirolo *et al., 2010 [8]*	466 (fs)	3541	98.4%	N/A	N/A	N/A	N/A	96.6
Mahoor *et al.*, 2009 [9]	370 (ns)	370 (ns)	N/A	93.7%	N/A	N/A	N/A	N/A
Faltemier *et al.*, 2008 [10]	410 (fns)	N/A	98.1%	N/A	N/A	N/A	N/A	94.8
Faltemier *et al.*, 2008 [10]	466 (fs)	3541	97.2%	N/A	N/A	N/A	N/A	94.8
Mian *et al.*, 2008 [11]	466 (fs)	3541	93.5%	99.0%	86.7%	N/A	N/A	N/A
Mian *et al.*, 2007 [12]	466 (fs)	3541	N/A	98.82%	92.36%	N/A	N/A	N/A
Osaimi *et al.*, 2007 [13]	466 (fs)	1944 (ns)	N/A	93.78%	N/A	N/A	N/A	N/A
Kakadiaris *et al.*, 2007 [14]	466 (fs)	3541	97.3%	99.0%	95.6%	97.3	97.2	97.0
Passalis *et al.*, 2007 [15]	466 (fs)	3541	96.5%	N/A	N/A	95.8	95.3	94.7
Faltemier *et al.*, 2006 [16]	410 (fns)	3951 [1]	94.9%	N/A	N/A	N/A	N/A	88.8
Cook *et al.*, 2006 [17]	410 (fns)	N/A	94.63%	98.25%	N/A	93.7	92.9	92.0
Passalis *et al.*, 2005 [18]	466 (fs)	3541	89.5%	N/A	N/A	89.0	88.0	87.0

[1] In [16], the probe set size was reported as 3951. This should be a typo since 410 and 3951 do not add up to 4007, the total number scans in FRGCv2.

to their discriminative power. The most common approach is to use 3-D point coordinates to represent facial surfaces. In our study, we show that, although it is marginally better to utilize more advanced 3-D shape descriptors such as principle curvature directions, the use of statistical feature extraction and the application of linear discriminant analysis (LDA) to 3-D point cloud (PC) features outweighs this advantage.

In this work, we have used two 3-D face databases containing expressions: Face Recognition Grand Challenge Version.2 (FRGCv2) and Bosphorus. FRGCv2 is the most commonly used database for 3-D face recognition and we have obtained comparable performance to the best accuracy reported in the literature: 97.51%. This study is also a first in benchmarking the recently acquired and publicly available Bosphorus database, which contains an extensive range of expressions. A recognition rate of 98.19% was obtained on this database.

II. RELATED WORK

Over the past few years, there has been a growing interest in 3-D face recognition systems. A thorough coverage of previously proposed 3-D face recognition systems can be found in surveys [2]–[4] and a more detailed treatment of some fundamental concepts can be read from some book chapters[5]–[7]. Comparative performance analysis of 3-D systems with other biometric modalities such as high-resolution 2-D still face images and iris can be found from the results of the large-scale independent evaluation effort of Face Recognition Vendor Test (FRVT) 2006 [1]. According to the FRVT 2006 results, identification performance of the 3-D modality has been found to be comparable to high-resolution 2-D still face images. FRVT 2006 performance benchmarks show that the best *3-D shape only* system from the University of Houston attains a median false reject rate (FRR) of 5.2% at a false accept rate (FAR) of 0.1% on a sequestered face database of 3589 scans from 330 subjects. Another important observation is that when 3-D information is coupled with 2-D intensity information, the performance improvement is significant: the best 3-D+2-D system, from Viisage, achieves a median FRR of 1.0% at an FAR of 0.1%. In this section, we focus only on the most recent approaches that try to overcome the expression variance problem with a special emphasis on the regional approaches. Table I gives a list of these approaches together with rank-1 identification accuracies and verification rates, where present, that are obtained on the FRGCv2 database.

Expression insensitive 3-D face recognition systems naturally focus on rigid parts of faces. The use of the nasal region is a prominent example of such approaches. In [19], three overlapping nose regions are extracted and matching scores from these different classifiers are combined at the score level. For automatic landmark localization, some fiducial points (namely the nose tip, eye pits, and nose bridge) are located using curvatures. These landmark points are utilized to segment the face into circular regions. For classifier fusion, product and sum rules yield the best performance: On the FRGCv2 *SuperSet* database with a gallery of 449 subjects with one neutral scan per person, the reported results are 97.1% and 87.1%, respectively, for neutral probe and non-neutral probe sets. The results of this study are not included in Table I, since the database is different. In a similar study, Faltemier *et al.* [16] used seven overlapping regions around the nose. The nose tip is located automatically by combining three different algorithms. For regional alignment, they have utilized the iterative closest point (ICP) algorithm. For each

region, classification votes are obtained using registration distances and threshold values. The regional classifiers are fused via committee voting. For the experiments, they have used the FRGCv2 database, and constructed the gallery with 410 subjects, each with a neutral scan. They provide a rank-1 recognition result of 94.9%. In their later work, Faltemier *et al.* [10] divided the face into a total of 38 regions, distributed over the whole facial area. The nose tip is automatically located and the regions are constructed using x and y offset values from the nose tip location and radius values to define the size of regions. The regional classification results are fused using a modified version of the Borda Count method. They have reported two different recognition results on the FRGCv2 with different gallery and probe sets. For the first set, the gallery contains 410 neutral scans from different subjects, and for the second set, the gallery consists of the first scan of each subject (which can either be neutral or non-neutral) making a total of 466 scans. The recognition results are 98.1% and 97.2%, respectively, for the first and second experimental sets. In all of these studies, the ICP-based core matcher should perform alignment for every gallery face, a time-consuming task when the gallery set is large.

Passalis *et al.* [18] utilize an annotated deformable face model, that is divided into different facial regions. The facial scans are rigidly registered to the model using ICP to obtain pose-invariance, where the model is elastically deformed to fit the registered scan afterwards. From the deformed model, the deformation image is obtained via UV parametrization and Haar wavelet filtering is applied for compression. For the experimental setup, the FRGCv2 database is divided into gallery and probe sets with 466 (first scan of each subject) and 3541 (remaining scans), respectively. Recognition scores for eye and nose areas, which are relatively resistant to expression variations, are reported as 85.8% and 81.5%, respectively. When the regional scores are fused, the recognition rate increases to 89.5%. In [15], this work is improved by using a simulated annealing method after the ICP for rigid registration of the scan to the annotated model. They report the recognition result as 96.5% on the FRGCv2 database. In [14], they further improve their work by adding the construction of a surface normal map. Both the geometry image and the normal map are analyzed using the Haar and Pyramid wavelet transforms, to obtain two sets of coefficients as distance metrics. The classifiers are then fused using the weighted sum approach. On the FRGCv2 database, they report results for neutral, non-neutral, and full probe sets as 99.0%, 95.6%, and 97.3%, respectively.

In [13], automatic nose tip localization is utilized and a region cropped around the nose tip is obtained. The cropped region is then triangulated and multiple local and global rank-0 tensors are computed. Two-dimensional histograms of these tensors are obtained and dimensionality reduction is done with the principal component analysis (PCA) method to form a single feature vector for each scan. The FRGCv2 database is used in experiments, where a gallery of 466 scans (first scan of each subject), and a neutral probe set of 1944 scans are constructed. A recognition rate of 93.78% is obtained for this neutral probe set. Mian *et al.* [12] developed a multimodal algorithm which combines 2-D and 3-D and the matching is handled in the hybrid mode where feature-based and holistic approaches are fused. Automatic extraction of inflection points around the nose tip are used to segment the face into eyes-forehead and nose regions, which are less affected by facial expressions. Separate matching of regions is handled with ICP and similarity measures are fused at the metric level. The FRGCv2 database is used for the experiments, with 466 and 3541 scans for gallery and probe sets, respectively. The use of 3-D information alone gives 98.82% and 92.36% recognition rates for neutral and non-neutral probe sets, respectively. In [11], they improve their multimodal method. In the 3-D space, they automatically detect key-points at locations with high shape variations. At each key-point, pose-invariant 3-D feature extraction is handled via surface fitting and regular resampling. PCA is applied on the extracted features and matching is obtained by fusing results at score and feature levels. On the FRGCv2 database, the gallery and probe sizes are 466 (first scan of each subject) and 3541 (remaining scans), respectively. When only the 3-D information is used, results on neutral, non-neutral, and full probe sets are 99.0%, 86.7%, and 93.5%.

Mahoor *et al.* [9] propose a method for 3-D face recognition from frontal range images. Their approach utilizes ridge images, consisting of points with maximum principal curvatures (points from eyes, nose, and mouth areas). For registration, two different methods are applied: Hausdorff distance and the ICP method. They obtain recognition results both on the FRGCv2 and the GavabDB databases. For the FRGCv2, they constructed the gallery and probe each with 370 neutral scans. The recognition accuracies are 58.92% and 91.8%, respectively, when Hausdorff distance and ICP methods are utilized for registration. Furthermore, if the whole surface is used for registration via ICP, a recognition rate of 93.7% is obtained.

In [8], the facial surface is divided into four regions: a circular and an elliptical region around the nose, an upper head region containing nose, eyes, and forehead areas, and a region consisting of the entire face. The regional registration is handled via simulated annealing using the most deformation-resistant areas to obtain expression invariance. The regional classifiers are fused via the sum rule. The FRGCv2 database is used for experiments, where the gallery contains the first scan of each subject and the remaining images constitute the probe set. A recognition result of 98.4% is obtained.

Cook *et al.* [17] used Log-Gabor templates (LGTs) on range images to deal with expression variations. A range image is divided into multiple regions both in spatial and frequency domains. Each individual region is classified separately and the results are fused at the score level. The facial image is divided into 147 regions and the size of the LGT response features are reduced by the PCA method. For classification, the Mahalanobis Cosine distance metric is used and the classifiers are fused by the sum rule. The experiments on the FRGCv2 database, with a gallery of neutral scans, yield a recognition performance of 94.63%.

In [20], Lu *et al.* combine surface matching with appearance-based matching. They apply a hybrid ICP algorithm in registering and matching phases of 3-D facial surfaces. In the hybrid ICP, two classical ICP algorithms, using point-to-point

and point-to-plane distances are the similarity metrics, where the first algorithm is used for alignment and the second for refinement. Coarse alignment prior to ICP is handled by extracting three corresponding feature points. For appearance-based matching, LDA is applied to 2-D textures. The weighted sum rule is used to combine the two classifiers. On a database of 200 subjects in the gallery and 598 probe scans with lighting, pose, and expression variations, recognition results of 86%, 77%, and 90% are obtained, respectively, for ICP, LDA, and ICP-LDA combination. In [21], Lu and Jain propose a method to model expression deformations to deal with expression variations. A control group consisting of a small number of subjects, is used to calculate different deformations caused by expressions. When matching a test scan to gallery faces, all deformation models obtained from the control group are applied to the gallery and the ICP algorithm is used to find the best fit. Experimental results are reported on a subset of FRGCv2, with a total of 150 scans from 50 subjects (each with one neutral, one smiling, and one surprise expression). Recognition rates of 97% and 87.6% are achieved, respectively, with and without the deformable models.

In [22], Li and Zhang use multiple intrinsic geometric descriptors such as angles, geodesic distances, and curvatures as features for an expression-invariant 3-D face recognition. For each individual feature, a set of weights are trained. To combine the attributes, a different set of weights are also trained. They have experimented with the GavabDB and a subset of the FRGCv2 containing a total of 180 scans from 30 subjects. For the GavabDB, recognition rates of 97.00% and 94.17% are obtained, respectively, for the leave-one-out (LOO) approach and for the normal reference (NR) approach. In the LOO approach, one scan for each subject is used as a probe face, and all the other faces constitute the reference system. In the NR method, all the neutral scans constitute the reference set, and the scans with expression variations form the probe set. On the subset of FRGCv2, they have obtained 96.67% and 98.89% recognition performances for the NR and LOO approaches, respectively. As a cross-database validation experiment, training was performed on the GavabDB to determine weights, and FRGCv2 subset was used as the probe set. Recognition rates of 85.34% and 95.56% were obtained with the NR and LOO methods, respectively.

III. Proposed System

The proposed system consists of four parts: 1) a novel automatic facial landmark detection algorithm, 2) a robust component-based registration that can deal with the large surface deformations caused by expressions, 3) discriminative 3-D feature extraction, and 4) a classifier fusion module. The automatic landmark detection algorithm locates five points around the nose region and these points are then used at the first phase of the coarse alignment step. Our region-based registration method is inspired by the average face model (AvFM)-based registration approach [23]–[25] and is extended to incorporate independent local regions as in [14], which will be referred to as the average region model (AvRM). The AvRM-based alignment offers several advantages such as using the generic facial parts as an index file, reducing the computational cost of registration due to the elimination of pairwise ICP registrations for every gallery image, and finally providing one-to-one correspondence of all surface points. After registration, we systematically study the importance of using several 3-D features ranging from *point coordinates*, invariant *curvature-based* features to *statistical features* obtained from point coordinates. At the last phase, we utilize several classifier fusion techniques, at *abstract* and *score level*, to deduce the identity of the given probe image. An illustration of the general outline of the proposed approach, in the best performing combination, is shown in Fig. 1.

A. Automatic Landmark Localization

The quality of the facial surface alignment methods, especially iterative approaches like the ICP method, relies on initial conditions, such as the starting positions of the facial surface pairs. In order to improve the convergence of the iterative registration methods, prealignment is often necessary. Most of the 3-D face recognition systems use facial landmarks during the prealignment, or *coarse alignment*, phase. Generally, the most distinctive facial features such as the nose tip, eye corners, and mouth corners are located for coarse registration. In our work, we use five fiducial points around the nose region that are mostly stable even under facial expression variations. These are left/right inner eye pits, nose tip and leftmost/rightmost points of the lower nose border region. Except for the situations where large in-plane rotations are present, all of these points can be localized efficiently and are sufficient for prealignment of facial surfaces.

Our landmark localization algorithm uses 3-D shape data only. We first detect the central profile contour, *facial symmetry axis*, and then search for the nose tip on the profile contour. In order to extract the vertical symmetry axis, we employ a symmetry operator that uses shape index values computed from surface curvatures. The use of curvature-based symmetry axis detection is advantageous since it is invariant to rotations and translations. The facial profile curve detection algorithm works as follows: First, for every point on the 3-D facial surface, we compute the shape index values that are based on maximum (κ_1) and minimum (κ_2) curvature values. Given κ_1 and κ_2, shape index $S(r) \in [0,1]$ of a surface point r is defined as

$$S(r) = \frac{1}{2} - \frac{1}{\pi}\tan^{-1}\frac{\kappa_1(r)+\kappa_2(r)}{\kappa_1(r)-\kappa_2(r)}. \tag{1}$$

By sampling $S(r)$, we obtain the shape index image I_S. Then, we use a local sliding window-based symmetry operator which computes a *symmetry map*, I_M, using I_S. The symmetry value of a pixel at the (i,j)th location of I_M is computed by a local window W of size $2N \times 2M$ centered at pixel (i,j)

$$I_M(i,j) = \sum_{m=-M}^{M}\sum_{n=0}^{N}\left|I_S(i+m,j-n) - I_S(i+m,j+n)\right|. \tag{2}$$

In I_M, pixels having smaller values denote regions of high symmetry. In our system, we set N and M to 15 pixels. A frontal 3-D face image without rotation variations is expected to have high symmetry map values along the vertical facial profile. With this assumption, we locate the vertical position of the symmetry

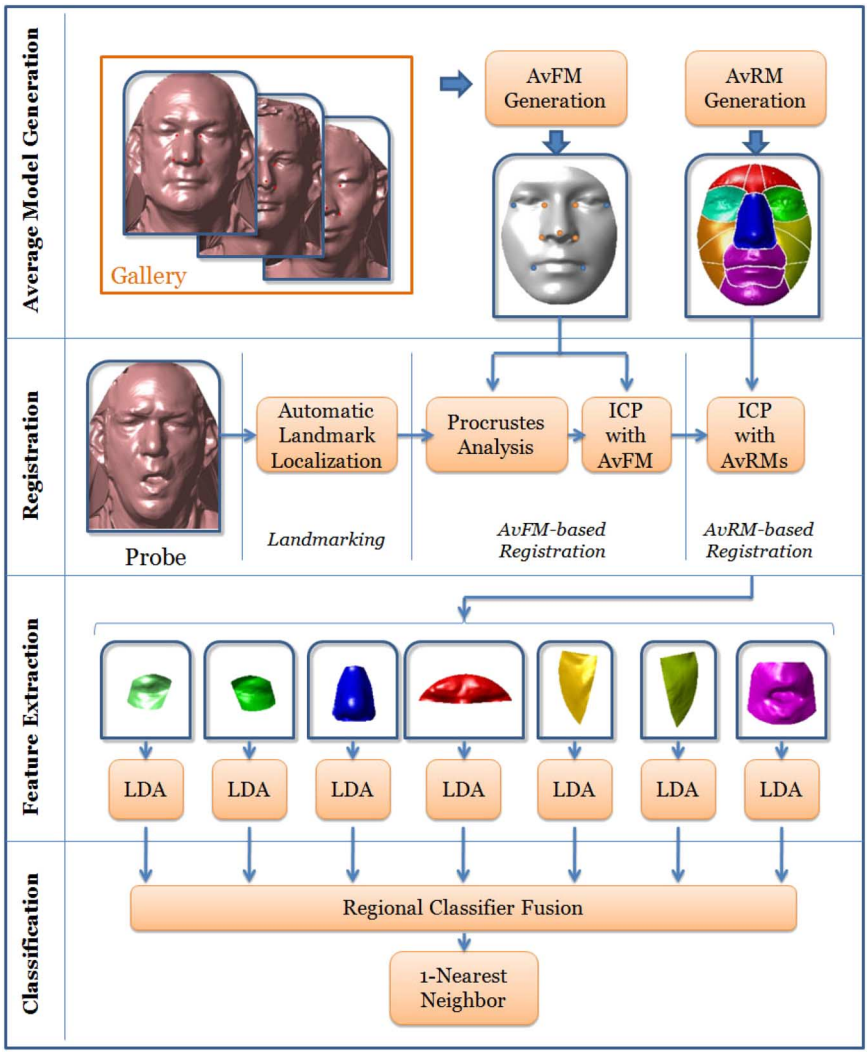


Fig. 1. Illustrative diagram of the proposed 3-D face recognition approach.

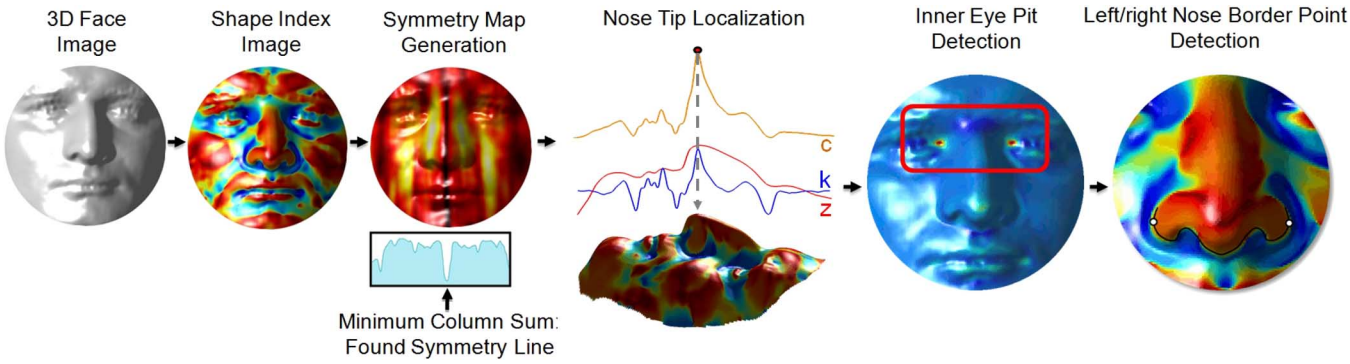


Fig. 2. Illustration of the automatic landmarking algorithm.

line in the symmetry map by selecting the vertical line which gives the minimum column-wise symmetry value sum.

In order to account for in-plane rotations of faces, we carry out the same procedure for different projection axes, i.e., by not only summing up symmetry values along the vertical lines but also using rotated lines. The projection axis producing the minimum symmetry sum gives the rotation angle of the face together with the position of the central profile line (see Fig. 2).

After finding the facial vertical profile contour, the nose tip location is found. For that purpose, we use both the depth measurements and the Gaussian curvature values along the profile axis. Gaussian curvature of a point r on a surface with principal curvatures κ_1 and κ_2 is defined as

$$K(r) = \kappa_1(r)\kappa_2(r). \tag{3}$$

Using a simple heuristic such as selecting the point having the biggest depth value as the nose tip position is not sufficient since in some cases, forehead or mouth may be closer to the camera due to expression or rotation variations. Therefore, we propose to combine Gaussian curvature values with depth measurements. Dome-like shape structures such as the nose tip region produce large Gaussian curvature values, thereby in combination with the depth information, the localization of the nose becomes more reliable. Let $z = (z_1, z_2, \ldots, z_n)$ and $k = (k_1, k_2, \ldots, k_n)$ be the normalized, depth, and Gaussian curvature value vectors along the profile line, respectively. We define a function of a combination of z and k as

$$c_i = z_i^2 k_i, \quad i = 1 \ldots n \tag{4}$$

and select $\arg\max_i c_i$.

The third step in automatic landmark localization is to find the inner eye pit locations. We observe that these points have cone-like shape structures around the upper nose area. Given the central facial profile and the nose tip position, it is easy to estimate a local search region for eye pits. In Fig. 2, the search window on a sample face can be seen. Since faces may have in-depth rotations, we use Gaussian curvature values to estimate the locations of eye pits, instead of using depth measurements. In the Gaussian curvature surface, cone-like structures produce values close to zero. Therefore, we search for the local minimum inside the search window and output these locations as the positions of eye pits.

Left and right outermost nose borders can be detected with the use of shape index descriptors efficiently as well. Saddle rut structures such as the nose border regions produce shape index values around 0.375. Therefore, we extract the nose border outline by a contour following approach where the pixels have saddle rut-like shapes. Using this approach, it is easy to extract the lower nose border contour, as shown in Fig. 2. Given the nose border contour, we select the rightmost and leftmost pixels along this curve as the rightmost and leftmost nose border points, respectively. More formally, let $C = \{(x_1, y_1), (x_2, y_2), \ldots, (x_l, y_l)\}$ denote the contour points where $|I_S(x_i, y_i) - 0.375| \leq \delta$, where δ is a small constant. It follows that the (x, y) locations of the left-most and right-most nose border points can be found by $x_{\text{left}} = \arg\min_x C$ and $x_{\text{right}} = \arg\max_x C$, respectively. y_{left} and y_{right} coordinates are the corresponding indices.

B. Three-Dimensional Face Registration

Three-dimensional registration establishes a one-to-one correspondence between the surface points of two given 3-D faces. The human face is a nonrigid surface which deforms in the presence of expressions initiated by muscle movements. The accuracy of rigid registration methods decreases when test scans with expressions are introduced. Region-based approaches try to overcome this difficulty by using smaller regions of faces [12], [14], [16], [19]. In region-based face recognition, a face is represented by a single robust region or it is considered as a composition of facial components. Rigid methods use ICP and rely on positions of landmarks on the face for initialization [12], [16], [19], while nonrigid methods elastically deform the surface to overcome the effect of expressions [14]. Our region-based registration approach provides a simple, fast, and robust two-pass alternative: We first employ ICP to register the facial surfaces to a common model, called the AvFM. This approach was previously used in [23]–[25] and the term AFM was used. In this paper, we abbreviate the average face model as AvFM to avoid confusion with the annotated face model (AFM) of [14]. The use of AvFM ensures that all gallery faces are in one-to-one correspondence. A second registration phase uses ICP registration to register individual regions to their respective AvRMs, starting from the initialization provided by the first phase.

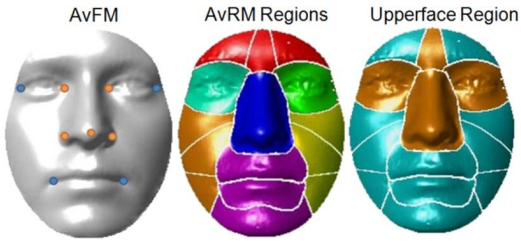


Fig. 3. AvFM and its landmark points computed from the FRGC database (leftmost image). Center and rightmost images show seven facial regions and upperface region for the AvRM, respectively.

In AvFM-based registration, the aim is to find the rigid transformation that will align a probe face, $P = \{\mathbf{p}_1, \ldots, \mathbf{p}_r\}$, with the AvFM, $M = \{\mathbf{m}_1, \ldots, \mathbf{m}_t\}$. The transformation $\mathbf{T}$ includes rotations around x, y, z axes, $\mathbf{R}_x, \mathbf{R}_y, \mathbf{R}_z$, respectively, and a translation $\mathbf{t}$

$$\mathbf{T}(\mathbf{p}_i) = \mathbf{R}_x \mathbf{R}_y \mathbf{R}_z \mathbf{p}_i + \mathbf{t}, \quad i = 1, \ldots, r. \tag{5}$$

The alignment error after the transformation is applied can be defined as

$$\sum_{j=1}^{t} \|\mathbf{T}(\mathbf{p}_i) - \mathbf{m}_j\| \tag{6}$$

where $\mathbf{m}_j$ is the point on the model which is in correspondence with the point $\mathbf{p}_i$. The optimal registration can be estimated as

$$\hat{\mathbf{T}} = \arg\min_{\mathbf{T}} \sum_{j=1}^{t} \|\mathbf{T}(\mathbf{p}_i) - \mathbf{m}_j\|. \tag{7}$$

As a holistic registration method, we use the iterative closest point (ICP) [26] to solve the linear numerical system defined over the whole surface points. Initial alignment prior to ICP iterations is carried out by the Procrustes analysis [27] using automatically located facial landmarks. The AvFM is constructed using coarsely registered gallery images by computing the mean shape. Details of AvFM construction algorithm can be read from [24] and [28]. An example AvFMs generated for the FRGC database is shown in Fig. 3 together with the landmarks.

For the AvRM-based registration, the construction of regional models is necessary. For this purpose, first of all an AvFM is generated from the gallery images using thin plate spline (TPS) warping, as described in [28]. Then, regional masks are created to divide the facial surface into patches that constitute the basic

building blocks. The facial patches are constructed by manually labeling corresponding areas on the AvFM. Patch construction on the AvFM is performed only once. The patches are collected into higher level components, namely the AvRMs. These regional models act as index files for the regional registration approach. In this work, we divided the face into a total of 15 patches, and from these patches we constructed seven meaningful regions: nose, left/right eye, forehead, left/right cheek, and mouth-chin. We also constructed a regional model for the area which is considered to be the region least affected by facial expression variations. This region is referred to as the upperface region and covers patches belonging to eye, nose, and forehead areas. The division of the facial surface into patches and the construction of regions from these patches are illustrated in Fig. 3.

The dense correspondence obtained between the face and the whole facial model acts as a coarse alignment for the regional approach. The aligned probe face, $P_{\text{reg}} = \{\mathbf{p}_1, \ldots, \mathbf{p}_t\}$, has the same number of points as the AvFM in exactly the same order. In AvRM-based registration, a second ICP is performed between the AvRM, $M^{(k)} = \{\mathbf{m}_1^{(k)}, \ldots, \mathbf{m}_{t_k}^{(k)}\}$ and the face previously registered to the AvFM P_{reg} to construct a local one-to-one correspondence of individual regions. The regions to be registered are considered independently of each other and for each component, different transformation parameters are calculated. For a region k, the transformation can be expressed as

$$\mathbf{T}^{(k)}(\mathbf{p}_i) = \mathbf{R}_x^{(k)}\mathbf{R}_y^{(k)}\mathbf{R}_z^{(k)}\mathbf{p}_i + \mathbf{t}^{(k)}, \quad i = 1, \ldots, t. \tag{8}$$

The alignment error after the transformation is applied can be defined as

$$\sum_{j=1}^{t} \left\| \mathbf{T}^{(k)}(\mathbf{p}_i) - \mathbf{m}_j^{(k)} \right\| \tag{9}$$

where $\mathbf{m}_j$ is the point on the AvRM model which is in correspondence with the point $\mathbf{p}_i$. The optimal registration for the considered region can be estimated as

$$\hat{\mathbf{T}}^{(k)} = \arg\min_{\mathbf{T}^{(k)}} \sum_{j=1}^{t} \left\| \mathbf{T}^{(k)}(\mathbf{p}_i) - \mathbf{m}_j^{(k)} \right\|. \tag{10}$$

The steps of our registration technique are summarized in the upper part of Fig. 1.

C. Three-Dimensional Features

1) PC Features: After the alignment phase, 3-D facial surfaces can be compared since they lie on the same coordinate system. A simple method is to use the coordinate differences between two corresponding surfaces. If the registered facial surfaces are resampled at the same (x, y) coordinates, then it suffices to use only the depth (z-coordinates) measurements in the computation of the PC value. More formally, let Φ be the whole facial surface composed of N local regions, ϕ_i, then $\Phi = \cup_{i=1\ldots N}\phi_i$. With the assumption of regular resampling in the PC method, each region ϕ_i is represented by a vector of z-depth measurements: $\phi_i = [z_1, z_2, \ldots, z_{M_i}]$, where each region ϕ_i contains M_i z-depth values. The dissimilarity between any two corresponding facial region then can be computed for person A and B as

$$D\left(\phi_i^A, \phi_i^B\right) = \frac{\left|\phi_i^A, \phi_i^B\right|}{M_i} \tag{11}$$

where $|.|$ denotes L_1-norm.

2) Statistical PC Features: A useful property of the generic AvRM-based registration is that 3-D facial features, particularly ϕ_i, are ordered vectors. In order to have a more compact and discriminative feature space, we propose to utilize LDA for the point coordinate features. Basically, we form a separate LDA space for every facial region. Construction of the LDA space, i.e., the computation of the LDA transformation matrix, is carried out by using an independent training set. Let Λ_i be the LDA transformation matrix found by region i. Then the regional LDA features γ_i can be found by the projection of ϕ_i: $\gamma_i = \Lambda_i\phi_i$. The dissimilarity between any two facial regions can be computed by the angular cosine distance measure in LDA subspace as

$$D\left(\gamma_i^A, \gamma_i^B\right) = 1 - \frac{\gamma_i^A \cdot \gamma_i^B}{\left|\gamma_i^A\right| \left|\gamma_i^B\right|}. \tag{12}$$

3) Curvature-Based 3-D Shape Descriptors: Given an erroneous alignment, PC representation-based similarity values are not discriminative. It is, therefore, necessary to consult better shape descriptors. In our system, we propose to use normal curvature descriptors (CDs) since they measure intrinsic characteristics of a surface, and are invariant to translations and rotations.

Normal curvatures measure the bending degree of a surface. For a 2-D surface, specifically, a Monge patch, s, characterized by a height function $f(u, v)$ defined over a support plane parameterized by (u, v), the intersection of s with planes defined by two orthogonal vectors in the tangent space produces plane curves. The direction at which the curvature of the plane curve is maximal or minimal determines the principal directions $\alpha_{\max}, \alpha_{\min}$. We use an analytical method outlined in [29] to estimate the principal directions. It is based on fitting a quadratic order surface of the form

$$z = f(x, y) = \frac{A}{2}x^2 + Bxy + \frac{C}{2}y^2 \tag{13}$$

in a neighborhood of the point of interest. The eigenvectors $e_{\max}, e_{\min}$ of the Weingarten matrix $W = \begin{pmatrix} A & B \\ B & C \end{pmatrix}$ can then be transformed by $\alpha_{\max} = e_{\max} \times [X_u X_v]$ and $\alpha_{\min} = e_{\min} \times [X_u X_v]$ to obtain principal directions $\alpha_{\max}, \alpha_{\min}$ in R^3. The support plane parameterized by (u, v) is defined by the orthogonal vectors X_u and X_v in R^3. Eigenvalues (λ_1, λ_2) of the Weingarten matrix correspond to principal curvature values (κ_1, κ_2). Coefficients A, B, and C are estimated by the least-squares technique. Using this method, we represent each point by a $(\alpha_{\max}, \alpha_{\min})$ three-vector pair (see Fig. 4 for a sample principal direction pair). Distance between corresponding points can then be calculated by the sum of the angle differences of maximal and minimal principal direction pairs. More formally, if only maximum principal curvature directions are considered, facial region i with M_i 3-D points is represented by $\Upsilon_i = [\alpha_{\max}^1; \alpha_{\max}^2; \ldots; \alpha_{\max}^{M_i}]$. The

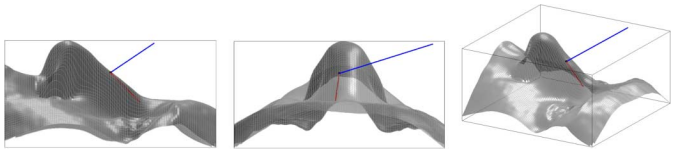


Fig. 4. Principle curvature direction pairs computed around the nose region, viewed from different angles. Red and blue lines show the minimum and the maximum principal curvature directions, respectively.

dissimilarity between two facial regions can then be computed by

$$D_{\max}\left(\Upsilon_i^A, \Upsilon_i^B\right) = \frac{\sum_{j=1}^{M_i} \operatorname{ang}\left(\alpha_A^j, \alpha_B^j\right)}{M_i} \tag{14}$$

where α_A^j denotes the jth maximum principal direction for person A. The function $\operatorname{ang}(.)$ denotes the angular difference between corresponding three-vector principal directions. Similarly, dissimilarity between facial regions is also computed for minimum principal directions, to obtain $D_{\min}$. The final curvature-based dissimilarity value is then computed by summing up these two values: $D = D_{\max} + D_{\min}$.

D. Classification: Fusion Techniques

In region-based techniques, each region acts as an independent classifier, and recognition results can be fused to obtain an improved overall performance. The fusion techniques can be grouped into three basic categories, namely score-level, rank-level, and abstract-level approaches [25]. In this work, we use score- and abstract-level fusion. In score-level fusion, the similarity measures obtained from different classifiers are combined using basic arithmetic rules. Two score-level methods are considered: sum rule (SUM) and product rule (PROD). Both of these rules operate on normalized distances. For distance normalization, we utilized the min–max normalization method.

In abstract-level fusion, each individual classifier produces a class label. The individual class labels are combined to provide a single label. In this category, plurality voting (PLUR) and modified plurality voting (MOD-PLUR) methods are considered. In PLUR, each expert provides the class label of the nearest gallery subject. Among the set of classifiers, the class label with the highest vote is assigned as the final label. When there are ties, the final label is randomly selected. In MOD-PLUR, the approach of plurality voting is improved, where for each classifier, a confidence value is estimated together with the class label. When there are ties, the decision is based on the confidence values. The confidence value is based on normalized scores. If $d = [d_1, d_2, \ldots, d_N]$ denotes the sorted dissimilarity values to N gallery samples in ascending order, a second score normalization is performed by

$$d_i' = \frac{(d_i - d_1)}{\operatorname{median}(d) - d_1}, \quad i = 2, \ldots, N. \tag{15}$$

After this score normalization, the classifier confidence can be defined as d_2'. The d_2' value gives the slope between the normalized scores of the first two top-ranked gallery classes. As the slope increases, the classifier gets more confident about its decision on the rank-1 class. For further details on confidence estimation, please refer to [25].

IV. Experimental Results

A. Databases

The main purpose of this work is to develop a 3-D face recognition system that is resistant to expression variations. For this purpose, we employed two 3-D face databases containing scans with facial expressions. The first one is the Bosphorus 3-D face database [30], which has a large variety of expressions. The second database considered is the FRGCv2 [31], which is a widely used database in the literature, exhibiting expression variations. In the following sections, details about the two databases are given.

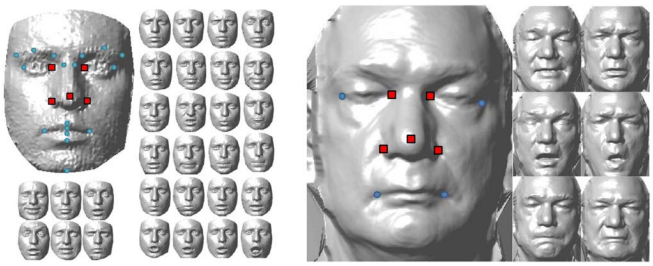
1) The Bosphorus 3-D Face Database: The Bosphorus database is a multiexpression and multipose 3-D face database. The database is constructed to enable testing of three main challenging scenarios for 3-D face recognition. For that purpose, both realistic and extreme *pose variations*, *expression variations*, and typical *occlusions* that may occur in real life are incorporated. A comprehensive set of scans that include both *emotional expressions* such as sadness, happiness, surprise, and *facial action units*, which are the building blocks of any facial mimic, were gathered, especially for automatic facial expression understanding studies. In order to incorporate realistic expressions, professional actors/actresses were also used during the acquisitions. For each subject, there are approximately 34 expressions, 13 poses, and 4 occlusions. The database includes a total of 4666 scans collected from 105 subjects, 61 men and 44 women. Professional actors/actresses constitute 29 of the subjects. For each scan, 22–26 manually labeled landmark points are present. The 3-D facial PC data is acquired using Inspeck Mega Capturor II 3-D, which has about 0.3-mm sensitivity in all dimensions and a scan consists of approximately 1.3-M points. The 3-D surfaces were processed to remove spikes and holes and to crop the facial area.

In this study, the focus is on expression variations: therefore, a subset of the database excluding pose variation and occlusion scans is considered. The subset has a total of 2919 scans, with roughly 34 different expression scans per subject. There are mainly two groups of facial expressions: The first group consists of action units (AUs) based on the Facial Action Coding System (FACS), which was developed for the taxonomy of plausible facial expressions of humans [32]. Among the 28 AUs, 20 lower face AUs, five upper face AUs, and three upper-lower combination AUs are taken into account. Expressions defined by AUs code the movement of several muscles; thus, some AUs are not present for some subjects who cannot control the related muscles. The second group of expressions is related to common emotions: happiness, surprise, fear, sadness, anger, and disgust. In Fig. 5(a), the manual landmark points present for each scan and the expression variability are illustrated.

As the experimental setup, we constructed a gallery set containing one neutral scan for each subject. The remaining scans constitute the probe set. Hence, gallery and probe set sizes are 105 and 2814, respectively.

TABLE II
AVERAGE EUCLIDEAN DISTANCES (mm) BETWEEN MANUAL AND AUTOMATICALLY FOUND LANDMARKS

	Inter Eye Distance	Left Inner Eye Corner	Right Inner Eye Corner	Nose Tip	Left Nose Corner	Right Nose Corner
Bosphorus	64	3.96	3.43	3.05	3.19	3.00
MLV (Bosphorus)		2.70	2.32	2.96	1.68	1.82
FRGCv2	70	4.90	5.05	3.26	4.68	4.51



(a) (b)

Fig. 5. Sample 3-D scans for the (a) Bosphorus and (b) FRGCv2 databases. For each database, manually located landmarks are shown on a neutral face. The orange-colored landmarks are used in the coarse registration phase. For the Bosphorus database, emotional expression variations are shown on the bottom left: happiness, sadness, surprise, fear, disgust, anger. On the right, scans for AU variations are shown. For the FRGC database, expression types are happiness, sadness, surprise, fear, cheek puffing, and disgust.

2) FRGCv2: The FRGCv2 database is widely used in the literature of 3-D face recognition due to its large number of subjects and large number of total scans. It consists of 4007 scans collected from 466 subjects using a Minolta Vivid 900/910 series sensor. The scans are frontal containing a number of facial expression variations such as happiness, sadness, surprise, anger, disgust, and cheek puffing. For this database, we manually labeled a set of nine landmark points on each facial surface. In Fig. 5(b), an example subject is shown with manual landmarks available and the expression variability is illustrated.

For the identification scenario, we designed an experimental setup with one image per subject in the gallery set and all the other images in the probe set. The gallery set constitutes a total of 466 scans, one scan per each subject. The images contained in the gallery are not restricted to be neutral, they are the first appearing scan of each subject. This experimental protocol is also used in [10], [11], and [14] and we have chosen the same setup to allow a direct comparison with the techniques proposed in those studies.

B. Automatic Landmark Localization Performance

Good landmarks are needed for convergence of the registration algorithm. The performance of the automatic landmark localization is thus, crucial. The average Euclidean distances of the automatically labeled landmarks to the corresponding manual landmarks are given in Table II for the Bosphorus and FRGCv2 databases. The average Euclidean distance between the eyes is 64 mm for the Bosphorus database. Therefore, it is seen that automatic landmark localization algorithm has an average error rate of 4%–6% of the inner eye distance. This accuracy is sufficient for a coarse registration, as will be shown by the identification accuracies in later sections. To evaluate the performance of the proposed automatic landmarking method better, we designed an experiment to observe the variability of the manual landmarking subject to the precision of the annotators. The five-point landmark set (the inner eye corners, nose tip, and the nose corners) is labeled by ten different annotators on a subset of the Bosphorus database, consisting of 20 scans. The average Euclidean distances given in the second row of Table II correspond to the *manual labeling variability* (MLV), which is the average distance of the manually labeled landmarks by the ten different subjects to the original manual landmarks. The results show that the variability of the automatic landmark locations and the variability that can be caused by the annotators are not significantly different. It is also evident that the outer nose corners are located more precisely both automatically and manually. Table III provides previously reported automatic landmarking accuracies on the FRGC database. As can be seen from Table III, some methods present average distance error in millimeters and others present localization rates in percentage given a distance threshold value. In addition to our FRGC results in millimeters (see Table II), we also provide localization rates (%) for commonly used threshold values, namely, 10, 12, and 20 mm, at the last row of Table III. When compared to the other results on the FRGC set, we see that our landmark localization method is sufficiently accurate.

The automatic landmark localization results are illustrated in Fig. 6(a) and (b) on a sample set of scans with facial expression variations for the Bosphorus and FRGCv2 databases, respectively. The original manually labeled landmarks are also shown, to permit visual interpretation of the results. It can clearly be seen that the eye and nose corner points can be located efficiently in the presence of expressions, enabling adequate results for the coarse registration phase.

C. Identification Results

Expression variations give rise to deformations on the facial surface. These deformations cause performance degradations of the registration approaches that treat the faces as rigid and *global* surfaces. To substantiate our assertion, we examined the AvFM-based rigid registration method on both the Bosphorus and the FRGCv2 databases. For coarse alignment of faces, Procrustes analysis utilizing the five-point landmark set is performed. The coarse alignment is followed by a fine registration step via the ICP algorithm. Subsequent to registration of the faces, the surfaces are considered as PCs and the Euclidean distances between a probe face and each of the gallery faces are computed. As a classification approach, the nearest neighbor algorithm is utilized to obtain identification results. In Table IV, the rank-1

TABLE III
OVERVIEW OF LANDMARK LOCALIZATION ALGORITHM PERFORMANCES ON THE FRGC DATABASE

Author,Year	Database	# Scans	Detected landmarks	Performance				Performance criteria
Zhao et al, 2009 [33]	FRGC v1	462	15 landmark points	Nose tip: 4.56 ± 2.51mm Right inner eye corner: 3.27 ± 1.14mm Left inner eye corner: 3.27 ± 1.51mm Left nose corner: 4.36 ± 3.73mm Right nose corner: 4.72 ± 2.12mm				Average distance(mm)
Romero and Pears, 2009 [34]	FRGC v2	4003	Inner eye corners, nose tip	Nose tip: 99.77% Inner eye corners: 96.82%				Less than 12mm
Dibeklioglu et al, 2008[35]	FRGC v1	314	Nose tip, Inner/outer eye corners, mouth	Nose tip: 99.10% Inner eye corners: 99.55%				Less than 24 pixels (14.4mm)
Mian et al, 2007 [12]	FRGC v2	4007	Nose tip	Nose tip: 98.3%				Not stated
Segundo et al, 2007 [36]	FRGC v2	4007	Nose tip, nose corners, nose base, inner eye corners	Nose tip: 99.95% Inner eye corners: 99.83% Nose corners: 99.76%				Not stated
Hose et al, 2007 [37]	FRGC v1	100	Nose tip, upper/lower nose corners, Outer eye corners	Nose tip: 3.17 ± 1.86mm (99.89%=20mm, 99.37%=10mm)				Less than 10 or 20mm
Salah and Akarun, 2006 [38]	FRGC v2	1893	Nose tip, inner/outer eye corners, mouth corners	Nose tip: 98.00% Inner eye corners: 97.10%				Less than 24 pixels (14.4mm)
Akakin et al, 2006 [39]	FRGC v1	235	Inner/outer eye corners, nose tip, mouth corners	7.81 pixels = 4.69mm (average distance of all landmarks)				Average distance(mm)
Chang et al, 2006 [19]	FRGC v2	4485	Inner eye corners, nose tip, nose bridge	99.4% (average distance of all landmarks)				Not stated
Lu and Jain, 2006 [40]	FRGC v1	953	Outer/Inner eye corners, nose tip, mouth corners	Nose tip: 8.3 ± 19.4mm Right inner eye corner: 8.3 ± 17.2mm Left inner eye corner: 8.2 ± 17.2mm				Average distance(mm)
Our method	FRGC v2	4007	Inner eye corners, nose tip, nose borders		$<$10mm	$<$12mm	$<$20mm	
				Nose tip	99.62	99.80	99.87	
				Inner Left Eye Corner	98.02	98.82	99.55	
				Inner Right Eye Corner	95.15	98.25	99.52	
				Left Nose Corner	98.62	99.30	99.87	
				Right Nose Corner	98.57	99.27	99.87	

recognition rates obtained via AvFM-based registration are reported on both databases, using manual and automatic landmarks in the coarse alignment phase. The first, second, and third rows are the identification performances for neutral, non-neutral, and for the full probe set, respectively. These results support our claim that the rigid registration accuracy decreases in the presence of facial expression variations. For the FRGCv2 database, which contains a large probe set of neutral and non-neutral scans, the performance degradation due to expression is about 40%. The performance decrease is also quite significant (30%) for the Bosphorus face database. Regarding the effect of manual and automatic landmarking on the identification performance, we see that performance decrease is quite small if landmarks are found automatically. By looking at the whole probe set (neutral + non-neutral), it is observed that rank-1 accuracies decrease by 0.14% and 0.36% for the FRGCv2 and Bosphorus databases, respectively, when automatic landmarks are used.

After showing that global ICP-based registration is not sufficient for non-neutral faces, we can now proceed to analyze local AvRM-based registration performances. As explained before, in the AvRM-based alignment, first a global ICP alignment is performed and then AvRMs are independently registered to a given probe facial surface. In Table V, the independent regional identification results using a single region are given. For comparative reasons, the identification rates obtained using the AvFM-based registration are given in the first row. As these results exhibit,

TABLE IV
IDENTIFICATION RESULTS OF THE AVFM-BASED APPROACH FOR MANUAL AND AUTOMATIC LANDMARKS

	FRGCv2 (Gallery Size: 466)			Bosphorus (Gallery Size:105)		
	Probe Size	Manual	Automatic	Probe Size	Manual	Automatic
Neutral Probes	1984	84.07	83.92	193	99.48	100.00
Non-Neutral Probes	1557	48.62	48.49	2621	69.71	69.29
All Probes	3541	68.48	68.34	2814	71.75	71.39

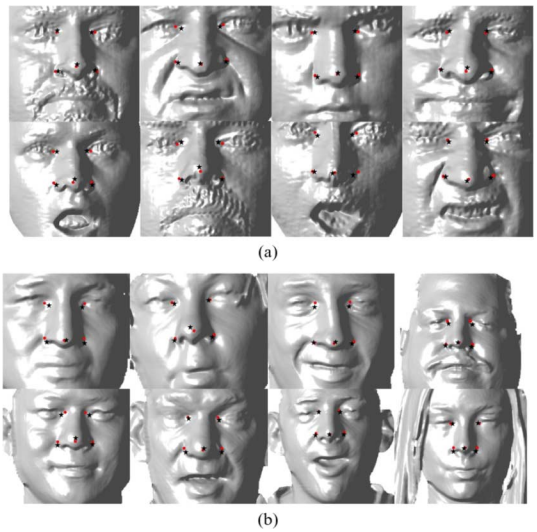


Fig. 6. Manually (red dots) and automatically (black stars) located landmarks shown on a sample set of scans for the Bosphorus and the FRGCv2 databases. (a) Bosphorus; (b) FRGCv2.

TABLE V
IDENTIFICATION RESULTS OF THE INDIVIDUAL REGIONS

	FRGCv2		Bosphorus	
	Manual	Automatic	Manual	Automatic
AvFM	68.48	68.34	71.75	71.40
Nose AvRM	85.12	84.98	86.96	86.70
Left Eye AvRM	65.15	65.04	60.23	60.27
Right Eye AvRM	64.90	64.78	62.30	62.26
Forehead AvRM	62.92	62.84	77.36	77.19
Left Cheek AvRM	34.11	33.95	32.76	32.66
Right Cheek AvRM	31.69	31.74	36.28	36.17
Chin-Mouth AvRM	41.51	41.46	36.64	36.57
Upperface AvRM	86.78	86.59	91.22	91.05

some regions are less affected by facial expressions, such as the nose, eye, and forehead regions. When the combination of these regions is used as a single AvRM, namely the upper-face AvRM, the best regional recognition rates are obtained. The cheek regions and the region containing the mouth and chin are the worst performing areas. This is basically due to the fact that facial expressions deform the mouth greatly and subsequently the cheek regions are affected by the mouth movement. Although their regional deformations are less than the mouth and chin area, cheek regions perform even worse, implicating their low discriminative ability.

An important observation from the results in Table V is that using only the nose region, it is possible to significantly improve the identification rates, compared to using the whole face with the standard ICP approach (the AvFM method). This finding is also compliant with the other studies that focus on the nasal region. However, we see that incorporating the forehead and eye regions with the nose, by forming a bigger upperface region, it is possible to improve the accuracy obtained by the nasal region alone. In terms of the landmarking method used in the coarse registration phase, we see that automatically located landmarks only slightly reduce the rank-1 identification accuracy for all the regions.

To improve the identification performances, we propose utilizing CDs as facial features instead of the PC of the surface. These features are invariant to rigid transformations, and are, therefore, less affected by erroneous registrations. Instead of using the whole PC for facial surface comparison, utilizing the intrinsic surface information residing in the principal curvature directions is beneficial. The global and local identification results using the CDs are reported in Table VI together with the previous results obtained using PC as the facial feature. As these results exhibit, the use of curvature directions is beneficial. The performance improvement is most distinct when the global registration approach is used, where the recognition rates increase by about 20%: For the FRGCv2 database, classical global ICP-based recognition rate improves from 68.34% to 87.15% and for the Bosphorus database, the improvement is from 71.40% to 94.92%. These results indicate that if a global registration is to be employed, it is better to utilize curvature features compared to using commonly used point coordinates. If a part-based registration is carried out, we also see an increase in the discriminative ability of individual regions by looking at the performances for facial regions in Table VI.

D. Fusion of Regional Classifiers

Although some regions are deformed less in the presence of facial expression variations, use of a single region is not sufficient for identification purposes. To improve the recognition results obtained by independent regional classifiers further, we propose to fuse the classification results. For each facial feature, namely PC or CD, we have eight regional classifiers: nose, left/right eye, forehead, left/right cheek, chin-mouth, and upperface classifiers. In Table VII, we present the fusion results for PC and CD methods using sum, product, plurality voting, and modified plurality voting fusion schemes. We also provide

TABLE VI
IDENTIFICATION RESULTS OF THE POINT CLOUD (PC) AND CURVATURE DESCRIPTORS (CDs)

	FRGCv2				Bosphorus			
	Manual		Automatic		Manual		Automatic	
	PC	CD	PC	CD	PC	CD	PC	CD
AvFM	68.48	87.40	68.34	87.15	71.75	95.38	71.40	94.92
Nose AvRM	85.12	92.01	84.98	91.81	86.96	97.94	86.70	97.69
Left Eye AvRM	65.15	86.53	65.04	86.53	60.23	90.94	60.27	90.76
Right Eye AvRM	64.90	85.60	64.78	85.20	62.30	91.68	62.26	91.54
Forehead AvRM	62.92	66.53	62.84	72.10	77.36	78.39	77.19	78.25
Left Cheek AvRM	34.11	56.76	33.95	56.68	32.76	40.58	32.66	40.47
Right Cheek AvRM	31.69	57.22	31.74	56.93	36.28	52.59	36.17	52.31
Chin-Mouth AvRM	41.51	54.14	41.46	58.94	36.64	43.89	36.57	43.71
Upperface AvRM	86.78	93.36	86.59	92.80	91.22	95.20	91.05	94.96

TABLE VII
FUSION RESULTS OF REGIONAL CLASSIFIERS USING PC, CD, AND COMBINED (PC+CD) FEATURES

		FRGCv2		Bosphorus	
		Manual	Automatic	Manual	Automatic
SUM	PC	81.11	61.25	86.11	76.15
	CD	92.57	90.93	98.47	98.19
	PC+CD	92.83	90.68	98.29	97.80
PROD	PC	88.39	88.11	95.91	95.56
	CD	93.36	91.81	98.47	98.15
	PC+CD	90.96	90.68	97.12	96.87
PLUR	PC	90.39	90.14	93.75	93.57
	CD	94.21	93.39	98.06	97.80
	PC+CD	95.72	94.99	98.39	98.05
MOD-PLUR	PC	91.39	91.16	94.92	94.63
	CD	94.27	93.56	98.26	98.01
	PC+CD	95.74	94.80	98.37	98.08

the fusion results of combining eight PC and eight CD classifiers, denoted as PC+CD in Table VII. In addition to the reported fusion schemes in Table VII, we have also tried several other fusion mechanisms such as highest confidence and the Borda count method. However, they have performed worse than the reported results in Table VII.

For the FRGCv2 database, the best identification accuracies are obtained by fusing 16 individual classifiers with the modified plurality voting scheme. If automatically found landmarks are used, MOD-PLUR achieves 94.80% rank-1 identification rate with the PC+CD classifiers. This is significantly better than the best individual classifier, namely, the nose classifier using the curvature features (91.81%, Table VI). It is also better than the best AvFM-based classifier (87.15%, AvFM-based CD, Table VI). If we compare the fusion methods, we see that if arithmetic rules such as sum/product are applied, increasing the number of base classifiers does not lead to an improvement. For instance, by inspecting the automatic FRGCv2 results, PC and CD fusions by the product rule achieve 88.11% and 91.81% accuracies. However, if PC+CD classifiers are fused with the product rule, the final accuracy is 90.68% which is not better than using CD fusion alone. In contrast, if voting-based fusion mechanisms are used, having more base classifiers always leads to a performance improvement. With this observation, we can deduce that it is better to fuse *many individual* classifiers with a *voting-based* mechanism. The same trend is also observed in the Bosphorus results. However, for the Bosphorus experiments, we see that sum/product rules are marginally better than the plurality schemes. On the other hand, as in the FRGCv2 results, having more classifiers, as using PC+CD, does not lead to superior results if arithmetic rules are employed. If automatic Bopshorus results are considered, it is seen that the best fusion accuracy, 98.19% with the sum-based fusion of CD classifiers is better than the best global AvFM-based classifier: 94.92% (see Table VI, CD-based AvFM classifier).

TABLE VIII
RANK-1 CLASSIFICATION RESULTS OF THE LDA-BASED AVRM APPROACH

	FRGCv2		Bosphorus (Evaluation Set)	
Gallery vs Probe	Manual	Automatic	Manual	Automatic
Neutral vs Neutral	98.59	98.39	99.47	100.00
Neutral vs Non neutral	97.11	96.40	99.60	99.25
Neutral vs All	97.94	97.51	99.59	99.31

E. Results of Statistical Features

In this section, we provide the classification results of using statistical point set-based features using the LDA technique. As explained before, ordered z-coordinates of the independently registered facial surfaces are used to construct LDA subspaces per region. In order to determine the LDA transformation matrix, we use separate training sets. For the FRGCv2 experiments, we use the FRGCv1 set which includes a total of 943 3-D scans. For the Bosphorus face database, we divide the whole database into two parts: 643 scans of 20 subjects are used

TABLE IX
OVERALL COMPARISON OF RECOGNITION RESULTS

		FRGCv2		Bosphorus (Evaluation set)	
Method	Description (registration method, features used)	Manual	Automatic	Manual	Automatic
AvFM-PC	Global reg., point set features	68.48	68.34	72.39	72.02
AvFM-CD	Global reg., curvature features	87.40	87.15	95.92	95.51
AvRM-PC	Regional reg., product fusion of point sets	88.39	88.11	96.47	96.24
AvRM-CD	Regional reg., product fusion of curvature	93.14	91.81	98.30	98.17
AvRM-PC-LDA	Regional reg., statistical LDA features	97.94	97.51	99.59	99.31

to construct LDA subspaces and the 2265 scans of 85 subjects are used to form an evaluation set (gallery and probe sets) for identification tests. The 20 subjects that are used for the LDA training are different from the ones in the evaluation set. In the Bosphorus evaluation set, there are 85 gallery images (single neutral image per person) and 2180 probe images. The rank-1 identification rates obtained by the product fusion of individual LDA-based regional classifiers are given in Table VIII. The results are provided in terms of Neutral versus Neutral and Neutral versus Non-neutral comparisons in order to analyze the behavior of the proposed scheme under expression variations. If we look at the FRGCv2 results with automatic landmarking, we see that a 97.51% rank-1 rate is achieved. Compared to the best performance of fusing raw features, i.e., using point sets or curvature directions directly without any statistical feature extraction, in the AvRM framework, that is fusing PC+CD with the PLUR scheme, we improve the accuracy from 94.99% (see Table VII) to 97.51%. This rank-1 classification rate obtained on the FRGCv2 database is the best reported accuracy in the literature and it is statistically significantly better than the best results obtained with the same experimental protocol (97.2% reported in Faltemier *et al.*, 2008 [10], 97.3% reported in Kakadiaris *et al.*, 2007 [14]) using a 0.05 level of significance. We have used the test provided in [10] to measure statistical significance. For the Bosphorus face database, fusion of LDA-based classifiers also provides very high identification rates: on the independent evaluation set 99.31% of the probe set is correctly classified. It should be noted that this performance value cannot be directly compared to the results provided in Table VII since the evaluation set is a subset of the whole database used in Table VII. A very important observation about our LDA-based regional approach is that non neutral probes are identified quite accurately compared to neutral probes. This proves that our proposed scheme, with the help of 1) regional registration and 2) the statistical subspace analysis is very beneficial and insensitive to expression variations. A very practical advantage of the LDA-based regional approach is the compactness of the feature vectors. The results shown in Table VIII are obtained by an LDA feature dimensionality of 90 per region. In real-world biometric applications, where the template size and matching speed are important, the use of such compact features is very crucial.

Lastly, in order to further analyze the generalization ability of the LDA approach, we perform *cross database* training for the FRGCv2 set. Basically, we train the LDA subspace with the Bosphorus training set and form the feature vectors for the FRGCv2 set by using the LDA space trained with the Bosphorus database. With cross database training, the rank-1 identification rate is 94.55% for the FRGCv2 database. This result implicates that even with such a challenging scenario of training with a completely different database with different sensor and different composition, it is possible to achieve quite acceptable recognition accuracy.

F. Overall Performance Comparison

In Table IX, we provide a summary of the classification performances of the methods studied in this paper. For each method, we emphasize the registration method (global versus local) and the types of features used (point sets, curvature based, and statistical). If the results are examined with respect to the landmarking method, we observe that automatic localization of facial fiducial points leads to a small performance decrease in general. For instance, the best performing approach, namely the AvRM-PC-LDA method, achieves 97.94% rank-1 identification rate on the FRGCv2 database with manually located landmarks. If automatically localized landmarks are used during registration, performance drops by 0.43% to 97.51%. For the Bosphorus database, performance degradation is even smaller. This finding shows that our proposed landmarking method is sufficient for identification.

With respect to the registration method utilized, we observe that for both PC and CD features, AvRM-based regional registration significantly improves the classification rates when compared to AvFM-based global registration. The performance improvement is more apparent for the PC features, from 68.34% to 88.11%, than for the CD features, from 87.15% to 91.81% (see automatic FRGCv2 results). This observation shows that point set features are more sensitive to registration errors thus leading to more improvement if better regional registration is carried out.

In terms of the 3-D features, we see that the use of raw point sets is always worse than using curvature-based features irrespective of the registration method employed. For both global and regional registration schemes, curvature features always attain better classification rates. However, we note that the performance gain is clearly more visible if the registration method is prone to errors as seen by the global registration results. By inspecting the automatic FRGCv2 results, we see that the performance improvement is 18.81% (from 68.34% to 87.15%) while

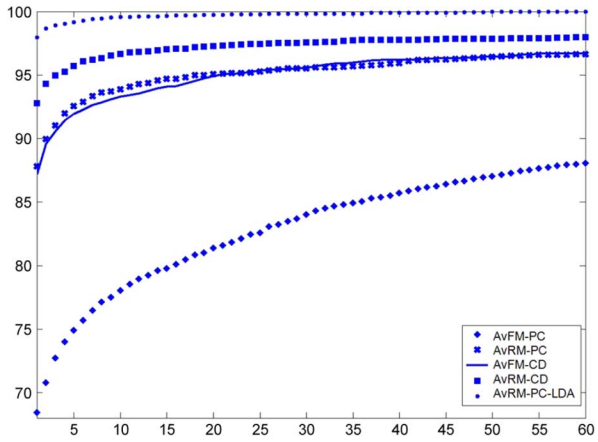


Fig. 7. CMC curves of the methods. First rank-60 results are shown for the manually landmarked FRGCv2 database.

for the AvRM-based regional registration the improvement is 3.70% (from 88.11% to 91.81%).

Lastly, we observe that by merging the power of regional registration through generic facial region models with the statistical feature extraction methods, the discriminative ability of 3-D features can be highly improved. The application of LDA, as a statistical feature extractor, improves the classification rates of the point set features from 88.11% to 97.51% for the FRGCv2 database, and from 96.24% to 99.31% for the Bosphorus database (see automatic landmarking results). Fig. 7 shows the cumulative match characteristics (CMC) curves of the methods for the manually landmarked FRGCv2 database. As verified by the CMC behaviors, the AvRM-PC-LDA method is superior to the others. We have also performed verification simulations using the AvRM-PC-LDA method on the FRGCv2 database using all possible pairwise genuine and impostor comparisons. The equal error rate (EER) obtained by the AvRM-PC-LDA method with automatically detected landmarks is 1.91% on the FRGCv2 database. In order to compare our verification performance with the ones reported in the literature, we carried out the standard verification tests of the FRGC protocol for the AvRM-PC-LDA method. Table X shows the verification rates for three different FRGC masks, namely ROC I, II, and III, at an FAR of 0.1%. Receiver operating characteristics (ROC) curves for the ROC I, ROC II, and ROC III masks are also shown in Fig. 8. When compared to the state-of-the-art verification rates provided in Table I, we see that our proposed scheme does not improve the state-of-the-art performances when used in the verification mode. We believe that fast registration of all the gallery scans with the probe image through the use of an AvFM is very efficient for fast matching computation in the identification mode. However, a further step of regional pairwise registrations between gallery and probe facial parts could be very beneficial for one-to-one matching in the verification mode where the speed of matching is not as crucial as in the identification mode.

One important advantage of our approach is its speed: Since the AvFM-AvRM approach aligns all gallery faces in the training phase, a single registration is sufficient for a test face. The computational complexity of ICP-based rigid registration is known to be $O(PQ)$, where P and Q are the

TABLE X
VERIFICATION PERFORMANCE (%) OF THE AVRM-PC-LDA METHOD ON THE FRGCV2 DATABASE. VERIFICATION RATES ARE COMPUTED AT FAR $= 0.1\%$

	Manual	Automatic
ROC I	85.81	85.39
ROC II	86.03	85.63
ROC III	86.09	85.64

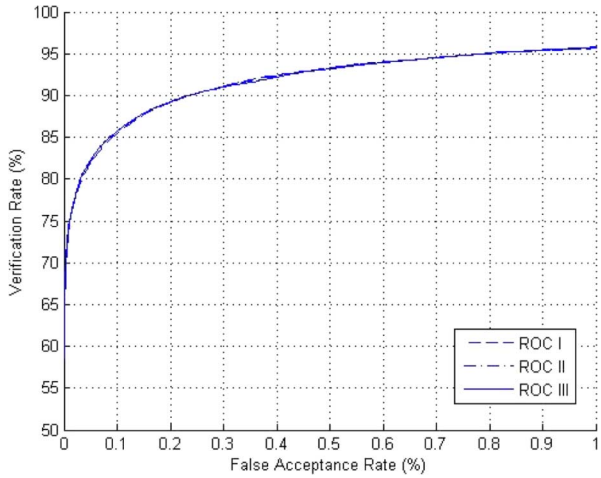


Fig. 8. Verification performance of the AvRM-PC-LDA method on automatically landmarked FRGCv2 database for the ROC I, ROC II, and ROC III masks. (Since the numerical values are very close, the three curves cannot be distinguished.)

number of points in the probe and gallery faces, respectively [9].The use of faster ICP implementations can reduce this complexity to $O(Q \log P)$. However, the use of a single ICP is still a significant advantage. In our implementation, we have used unoptimized MATLAB code running on a 64-bit Core i7 2.67-GHz PC with 12-GB RAM. The detailed timings for processing a single test face are as follows: Detecting five landmarks takes approximately 11 s: 8 s for curvature calculations and the rest for other operations. Registration to the AvFM takes 10 s. The second phase of registration involving eight individual ICP registrations, each ranging from approximately 8–23 s (on the average, 15 s), is actually parallellizable. It is remarkable that the rest of the distance calculations is negligible due to reduced dimensions by LDA: under 3 ms per region for the whole FRGCv2 gallery (466 faces).

V. CONCLUSION

In this paper, we have presented a fully automatic 3-D face recognition system which exploits facial surface characteristics to infer the identity of a person. Our focus was to design an expression insensitive face recognition system. Among other factors such as occlusions and large head pose rotations, facial expression variations present considerable challenges for a typical 3-D face recognition system. In order to achieve an accurate identification system under severe expression variations, it is essential to employ an efficient facial surface registration scheme. The main contribution of our work is the utilization of component-based regional registration methodology with the help of a

generic face model and generic region models which has advantages for 1) better registration under local facial surface deformations, 2) fast search in identification mode, and 3) the applicability of statistical feature extraction methods for unordered 3-D point data.

Our proposed locally rigid regional registration scheme has shown to be resistant to circumstances where faces exhibit expression variations. Our regional registration scheme automatically locates five facial landmarks around the nose region and performs two-phase rigid alignment. In the first phase, a coarse alignment is performed with a global AvFM and at the second step, generic regional templates are independently fitted to a given probe surface. Thereby, we can automatically divide any given facial surface into its meaningful parts, registered to a generic part model. The use of average facial region models during the registration considerably reduces the time complexity of the dense correspondence establishment with the gallery set. We show that with the use of component-based methodology, the performance of standard global ICP-based approach can be significantly improved. On the FRGCv2 set, this improvement is from 68.34% to 88.11%.

We also study the use of better 3-D features in terms of classification performance. As opposed to using popular point coordinate features, we propose to use curvature-based 3-D shape descriptors, particularly principal curvature directions. We show that principle curvature directions offer higher discriminative power than the point sets. Especially, if a global registration has been used, the accuracy gain is substantial. On the FRGCv2 set, the classification rate of a globally registered 3-D face recognizer improves from 68.34% to 87.15%. Together with the use of component-based registration for curvature features, this performance rate increases to 91.81%. This finding leads to a conclusion that it is possible to utilize more effective 3-D features than point coordinates, and principle curvature directions are good alternatives.

The use of generic template-based alignment also provides a natural ordering of 3-D facial features so that they can later be considered as *feature vectors*. As opposed to unordered classical point set matching methods commonly employed in 3-D face recognition systems, ordered feature representation allows the use of statistical pattern recognition techniques. We demonstrate the strength of this methodology by utilizing the LDA-based feature extraction method. For each separate facial region, we form an LDA subspace and combine the decisions of local classifiers via several score/abstract-level fusion schemes. A particular advantage of using an LDA-based system is the low space complexity of the biometric templates which also leads to a very fast matching. With the use of LDA features computed from depth measurements, we can achieve comparable performance to the best reported rank-1 identification rate, 97.51%, on the FRGCv2 database using a standard evaluation protocol. On this specific FRGCv2 experiment, 98.39% of the neutral probes (1984 scans) and 96.40% of the non-neutral probes (1557 scans) are identified correctly. Similar performance figures are also obtained on a more challenging multiexpression Bosphorus face database. On the evaluation set of the Bosphorus database, the proposed system attains 99.31% rank-1 classification rate. These identification results show that the proposed AvRM-based 3-D face recognition system with the use of discriminative features can be able to deal with expression variations and perform quite accurately. While the regional registration can cope with facial expression variations effectively, registering to an average model brings the ability to use dimensionality reductions techniques such as LDA. By registering each facial region to a common regional model (AvRM), we perform LDA in a smaller space where the main mode of variation is based on identity. Hence, the LDA in the regional spaces is able to capture identity variations better.

An interesting aspect of this work has been the use of two very different 3-D databases for experimental validation. While the FRGCv2 database is a commonly used benchmark, it uses a high sensitivity sensor that has a high capture time. The Bosphorus database has been collected with a less expensive device with a different sensor technology. The captured data is noisier, and the expression variation is much larger in the captured faces. We have shown that our techniques are valid, and roughly comparable, in both databases. Furthermore, we have shown that cross-database training is possible and beneficial. Future studies must investigate aggregation of data from different sensors.

REFERENCES

[1] P. Phillips, W. Scruggs, A. O'Toole, P. Flynn, K. Bowyer, C. Schott, and M. Sharpe, FRVT 2006 and ICE 2006 Large-Scale Results National Institute of Standards and Technology, NISTIR, 2007, vol. 7408.

[2] 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 Understanding*, vol. 101, no. 1, pp. 1–15, 2006.

[3] A. Scheenstra, A. Ruifrok, and R. C. Veltkamp, "A survey of 3D face recognition methods," *Lecture Notes Comput. Sci.*, vol. 3546, pp. 891–899, 2005.

[4] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino, "2D and 3D face recognition: A survey," *Pattern Recognit. Lett.*, vol. 28, no. 14, pp. 1885–1906, 2007.

[5] B. Gokberk, A. A. Salah, L. Akarun, R. Etheve, D. Riccio, and J. L. Dugelay, "3D face recognition," in *Guide to Biometric Reference Systems and Performance Evaluation*, D. Petrovska-Delacretaz, G. Chollet, and B. Dorizzi, Eds. New York: Springer-Verlag, 2008, pp. 1–33.

[6] A. F. Abate, S. Ricciardi, and G. Sabatino, "3D face recognition in a ambient intelligence environment scenario," in *Face Recognition*, K. Delac and M. Grgic, Eds. Vienna, Austria: I-Tech, 2007.

[7] T. Papatheodorou and D. Rueckert, "3D face recognition," in *Face Recognition*, K. Delac and M. Grgic, Eds. Vienna, Austria: I-Tech, 2007.

[8] C. C. Queirolo, L. Silva, O. R. P. Bellon, and M. P. Segundo, "3D face recognition using simulated annealing and the surface interpenetration measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 2, pp. 206–219, Feb. 2010.

[9] M. H. Mahoor and M. Abdel-Mottaleb, "Face recognition based on 3D ridge images obtained from range data," *Pattern Recognit.*, vol. 42, no. 3, pp. 445–451, 2009.

[10] T. C. Faltemier, K. W. Bowyer, and P. J. Flynn, "A region ensemble for 3D face recognition," *IEEE Trans. Inf. Forensics Security*, vol. 3, no. 1, pp. 62–73, Mar. 2008.

[11] A. S. Mian, M. Bennamoun, and R. Owens, "Keypoint detection and local feature matching for textured 3D face recognition," *Int. J. Comput. Vis.*, vol. 79, no. 1, pp. 1–12, 2008.

[12] A. Mian, M. Bennamoun, and R. Owens, "An efficient multimodal 2D-3D hybrid approach to automatic face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 11, pp. 1927–1943, Nov. 2007.

[13] F. R. Al-Osaimi, M. Bennamoun, and A. Mian, "Integration of local and global geometrical cues for 3D face recognition," *Pattern Recognit.*, vol. 41, no. 3, pp. 1030–1040, 2007.

[14] I. A. Kakadiaris, G. Passalis, G. Toderici, M. N. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis, "Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 640–649, Apr. 2007.

[15] G. Passalis, I. A. Kakadiaris, and T. Theoharis, "Intraclass retrieval of nonrigid 3D objects: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 2, pp. 218–229, Feb. 2007.
[16] T. C. Faltemier, K. W. Bowyer, and P. J. Flynn, "3D face recognition with region committee voting," in *The Third Int. Symp. 3D Data Processing, Visualization, and Transmission (3DPVT)*, 2006, pp. 318–325, IEEE Computer Society.
[17] J. Cook, V. Chandran, and C. Fookes, "3D face recognition using log-gabor templates," in *Proc. British Machine Vision Conf.*, 2006, pp. 83–92.
[18] G. Passalis, I. A. 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 Computer Society Conf. Computer Vision and Pattern Recognition (CVPR)—Workshops*, 2005, p. 171.
[19] K. I. Chang, K. W. Bowyer, and P. J. Flynn, "Multiple nose region matching for 3D face recognition under varying facial expression," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1695–1700, Oct. 2006.
[20] X. Lu, A. K. 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.
[21] X. Lu and A. K. Jain, "Deformation modeling for robust 3D face matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 8, pp. 1346–1357, Aug. 2008.
[22] X. Li and H. Zhang, "Adapting geometric attributes for expression-invariant 3D face recognition," in *IEEE Int. Conf. Shape Modeling and Applications*, 2007, pp. 21–32.
[23] B. Gokberk, M. O. Irfanoglu, and L. Akarun, "3D shape-based face representation and feature extraction for face recognition," *Image Vis. Computing*, vol. 24, no. 8, pp. 857–869, Aug. 1, 2006.
[24] M. O. Irfanoglu, B. Gokberk, and L. Akarun, "3D shape-based face recognition using automatically registered facial surfaces," in *Proc. Int. Conf. Pattern Recognition*, 2004, vol. 4, pp. 183–186.
[25] B. Gokberk, H. Dutagaci, A. Ulas, L. Akarun, and B. Sankur, "Representation plurality and fusion for 3-D face recognition," *IEEE Trans. Syst. Man Cybern. B, Cybern.*, vol. 38, no. 1, pp. 155–173, Feb. 2008.
[26] P. J. Besl and H. D. McKay, "A method for registration of 3D shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, Feb. 1992.
[27] C. Goodall, "Procrustes methods in the statistical analysis of shape," *J. Royal Statistical Society, Series B (Methodological)*, pp. 285–339, 1991.
[28] A. A. Salah, N. Alyuz, and L. Akarun, "Registration of 3D face scans with average face models," *J. Electron. Imag.*, vol. 17, no. 1, 2008, Article 011006.
[29] J. Goldfeather and V. Interrante, "A novel cubic-order algorithm for approximating principal direction vectors," *ACM Trans. Graphics*, vol. 23, no. 1, pp. 45–63, 2004.
[30] A. Savran, N. Alyuz, H. Dibeklioglu, O. Celiktutan, B. Gokberk, L. Akarun, and B. Sankur, "Bosphorus database for 3D face analysis," in *Proc. First Eur. Workshop Biometrics and Identity Management Workshop (BioID)*, 2008, pp. 47–56, Springer.
[31] P. J. Phillips, P. 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. IEEE Computer Society Conf. Computer Vision and Pattern Recognition (CVPR)*, 2005, vol. 1, p. 954.
[32] P. Ekman and W. V. Friesen, *Facial Action Coding System (FACS): A Technique for the Measurement of Facial Action*. Palo Alto, CA: Consulting, 1978.
[33] X. Zhao, E. Dellandréa, and L. Chen, "A 3D statistical facial feature model and its application on locating facial landmarks," *Lecture Notes Comput. Sci.*, vol. 5807, pp. 686–697, 2009.
[34] M. Romero and N. Pears, "Landmark localisation in 3D face data," in *Proc. Sixth IEEE Int. Conf. Advanced Video and Signal Based Surveillance*, 2009, pp. 73–78.
[35] H. Dibeklioglu, A. A. Salah, and L. Akarun, "3D facial landmarking under expression, pose, and occlusion variations," in *Proc. 2nd IEEE Int. Conf. Biometrics: Theory, Applications and Systems*, 2008, pp. 1–6.
[36] M. P. Segundo, C. Queirolo, O. R. P. Bellon, and L. Silva, "Automatic 3D facial segmentation and landmark detection," in *Proc. 14th Int. Conf. Image Analysis and Processing*, 2007, pp. 431–436.
[37] J. D'Hose, J. Colineau, C. Bichon, and B. Dorizzi, "Precise localization of landmarks on 3D faces using gabor wavelets," in *Proc. First IEEE Int. Conf. Biometrics: Theory, Applications, and Systems*, 2007, pp. 1–6.
[38] A. A. Salah and L. Akarun, "3D facial feature localization for registration," *Lecture Notes Comput. Sci., Multimedia Content Representation, Classification and Security*, vol. 4105, pp. 338–345, 2006.
[39] H. C. Akakin, A. A. Salah, L. Akarun, and B. Sankur, "2D/3D facial feature extraction," *Proc. SPIE: Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning*, vol. 6064, no. 1, 2006, Article 60641D.
[40] X. Lu and A. Jain, "Automatic feature extraction for multiview 3D face recognition," in *7th Int. Conf. Automatic Face and Gesture Recognition*, 2006, pp. 585–590.

Neşe Alyüz received the B.S. degree in computer engineering from Istanbul Technical University, Istanbul, Turkey, in 2005, and the M.Sc. degree in computer engineering from Boğaziçi University, Istanbul, Turkey, in 2008. She is currently working toward the Ph.D. degree at the Perceptual Intelligence Laboratory of Boğaziçi University, working on "Preprocessing of 3-D Data for 3-D Face Recognition."

Her research areas are face biometrics, computer vision, image processing, and pattern recognition.

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 senior scientist at Philips Research, Eindhoven, between 2006 and 2008. He is currently with the Signals and Systems chair of Department of Electrical Engineering Mathematics and Computer Science, University of Twente, The Netherlands. His research interests are in the areas of biometrics, computer vision, computer graphics, and pattern recognition.

Lale Akarun (M'86–SM'02) received the B.S. and M.Sc. degrees in electrical engineering from Bogazici University, Istanbul, in 1984 and 1986, respectively, and the Ph.D. degree from Polytechnic University, New York, in 1992.

Since 1993, she has been working as a faculty member at Bogazici University. She became a Professor of Computer Engineering in 2001. Her research areas are face and hand biometrics, and human activity and gesture analysis. She has worked on the organization committees of IEEE NSIP99, EUSIPCO 2005, eNTERFACE2007, and ICPR2010.