# Registration of three-dimensional face scans with average face models

**Albert Ali Salah**
Boğaziçi University
34342 Bebek
İstanbul, Turkey
Centrum voor Wiskunde en Informatica
Kruislaan 413, 1098 SJ
94079, The Netherlands
E-mail: salah@boun.edu.tr

**Neşe Alyüz**
**Lale Akarun**
Boğaziçi University
34342 Bebek
İstanbul, Turkey

**Abstract.** *The accuracy of a three-dimensional (3-D) face recognition system depends on a correct registration that aligns the facial surfaces and makes a comparison possible. The best results obtained so far use a costly one-to-all registration approach, which requires the registration of each facial surface to all faces in the gallery. We explore the approach of registering the new facial surface to an average face model (AFM), which automatically establishes correspondence to the preregistered gallery faces. We propose a new algorithm for constructing an AFM and show that it works better than a recent approach. We inspect thin-plate spline and iterative closest-point-based registration schemes under manual or automatic landmark detection prior to registration. Extending the single-AFM approach, we consider employing category-specific alternative AFMs for registration and evaluate the effect on subsequent classification. We perform simulations with multiple AFMs that correspond to different clusters in the face shape space and compare these with gender- and morphology-based groupings. We show that the automatic clustering approach separates the faces into gender and morphology groups, consistent with the other race effect reported in the psychology literature. Last, we describe and analyze a regular resampling method, that significantly increases the accuracy of registration. © 2008 SPIE and IS&T.* [DOI: 10.1117/1.2896291]

## 1 Introduction

With the advances in acquisition hardware and three-dimensional (3-D) recognition algorithms, 3-D face recognition has become an important biometric modality.[1] The illumination effects that impair two-dimensional (2-D) face recognition are alleviated to a large degree when using the surface of the face. Additionally, 3-D face recognition offers new possibilities of representation when compared to 2-D face recognition, and the choice of representation for the surface has a bearing on the accuracy and the speed of the 3-D face recognition algorithm. For instance, selecting a depth map to represent a 3-D face makes it possible to use almost all the techniques developed for 2-D face recognition, whereas storing true 3-D information potentially increases the accuracy.

The speed and accuracy bottleneck in 3-D face recognition is in the registration step, which follows the preprocessing. Test faces need to be aligned to gallery faces for comparison. Most of the algorithms start by coarsely aligning the faces, either by their centers of mass,[2] the nose tip,[3] or the eyes,[4] or by fitting a plane to the face and aligning it with that of the camera.[5] Good registration of the images is important for all local similarity measures.

In 3-D face classification applications, the most frequently employed registration approach is the iterative closest point (ICP) algorithm, which establishes a dense correspondence between two point clouds in a rigid manner.[6] Typically, a test face is registered to each gallery face separately,[7–9] and a point set distance is adopted for classification.

In the work of İrfanoğlu *et al.*, an alternative and fast method was proposed to register faces,[10] where an average face model (AFM) was employed to determine a single point-to-point correspondence. The gallery faces, previously used in generating the AFM, are already in dense correspondence with it. Thus, a single ICP is enough to register a test face, which is much faster for a reasonably sized gallery. This overcomes what has been reported as the major drawback of ICP.[8,11]

In this paper, we explore and expand AFM-based registration. Our methodology is summarized in Sec. 2, where we propose an approach that uses one AFM for each facial category. Section 3 provides a brief summary of the rigid and nonrigid registration methods that we use, along with a description of the automatic landmarking algorithm. The

Paper 07107SSR received Jun. 15, 2007; revised manuscript received Oct. 25, 2007; accepted for publication Nov. 28, 2007; published online Mar. 31, 2008. This paper is a revision of a paper presented at the SPIE Conference on Vision Geometry XV, February 2007, San Jose, California. The paper presented there appears (unrefereed) in SPIE Proceedings Vol. 6499.
1017-9909/2008/17(1)/011006/14/$25.00 © 2008 SPIE and IS&T.

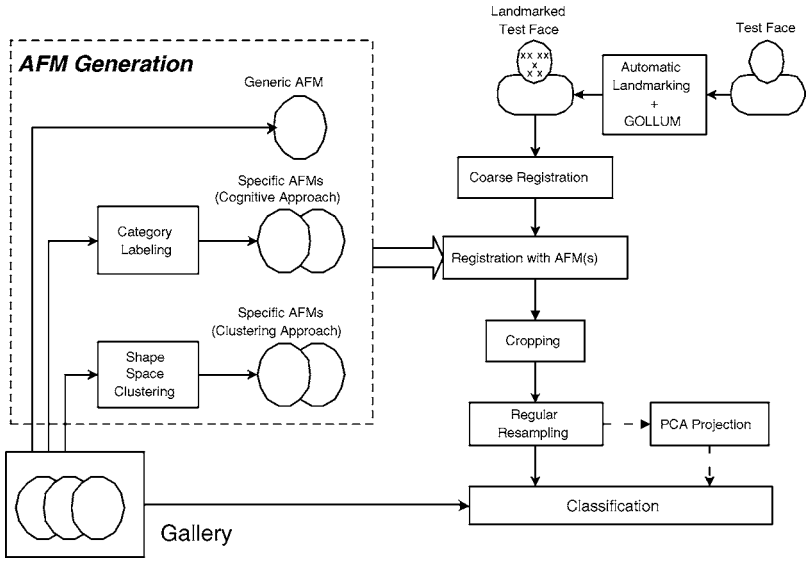


**Fig. 1** The fully automatic 3-D face recognition model with average face models. Three alternative ways of using AFMs are shown together.

novel AFM construction algorithm is presented in Sec. 4, where we also describe two different approaches for creating category-specific AFMs. We report our results on the FRGC version 1 face database in Sec. 5 and conclusions in Sec. 6.

## 2 Methodology

The goal of this paper is to investigate AFM-based registration for 3-D face recognition. Our setup therefore is a full 3-D face recognition system with preprocessing, landmarking, registration, and recognition. This setup is shown in Fig. 1, which shows the stages of processing for a new query.

The query image is first preprocessed to clear it from imaging artifacts, and a landmarking algorithm is run to detect seven landmarks (eye and mouth corners and the nose tip) on the face. The next step is a coarse registration, which is important for ICP, followed by rigid or nonrigid registration. The AFM (or multiple AFMs) are registered to the test face, and after the dense correspondence is established, the points that do not have correspondences in the AFM are removed. This effectively crops the face area and produces a set $S$ of 3-D points. After the cropping, depth values on the test face are resampled from a regular $x-y$ grid. We have used a simple triangle-based nearest-neighbor interpolation algorithm for this purpose.[12] After the dense correspondence, the point vectors representing the faces are of the same size, and it becomes possible to apply a principal components analysis (PCA) projection to store much smaller templates.

Figure 1 makes it clear that there are many design choices for the complete system, and each of these choices (manual versus automatic landmarking, type of coarse registration, single versus multiple AFMs, type of AFMs, rigid versus nonrigid registration, with or without regular resampling, classification with or without PCA projection) has an effect on the performance. We have designed experiments to show the contribution of each of these issues and inspected whether a trade-off is offered by a particular choice, or whether the setting indicates the superiority of one approach over the other.

The first set of experiments deal with the stage prior to the dense registration. The coarse registration in ICP and the nonrigid thin-plate spline (TPS)–based registration both require a couple of fiducial points for guidance. We evaluate the effect of errors in landmark detection by using 3-D ground-truth versus automatically located landmarks (Sec. 3.4). This permits us to analyze the algorithms under realistic assumptions, as automatic landmarking errors are not uniformly distributed. The details of the ICP and TPS methods are given in Secs. 3.2 and 3.3, respectively.

For coarse registration in ICP, we test four different methods. Our results indicate that the simple heuristics used in the literature leave room for improvement. Then we propose a novel AFM construction method and obtain very good registration results with a single, generic AFM.

Using a single AFM for registration and registering separately to each gallery face are the two extreme approaches. Between these extremes, using a few category-specific AFMs can be beneficial to accuracy and still be computationally feasible. To test this hypothesis, we propose two different methods for generating category-specific AFMs. What constitutes a facial category is an open issue; we contrast an approach based on cognitive justifications (detailed in Sec. 4.2) with one that is based on clustering on the shape space (Sec. 4.3). The cognitive approach allows recognition scenarios to use category information (e.g., the query for a male face is searched among the males only). But if the category information is not assumed to be available, it is still possible to register the test face to all of the available AFMs and take the best registration (the one with the smallest distance) as the outcome. We test both approaches in Sec. 5.

In the last part of this paper, we demonstrate that resampling of depth values from a regular grid improves accuracy significantly, and we justify this improvement by

a detailed analysis. We also show that a good registration ensures that a subspace projection can be used with little decrease in accuracy.

## 3 Registration Methods

We consider two different registration methods in this paper. In the first method, termed *TPS-based* in the experiments section, the test face is aligned to the average face with the TPS method, and the points not in correspondence with the AFM are cropped.[10] This method deforms the test face to fit the AFM, and the amount of deformation is proportional to the number and spread of the landmarks. At the limit of using all facial points as landmarks, the face deforms into the AFM, losing the discriminative information completely. However, with a few landmarks, corresponding facial structures are aligned.

In the second method, we use the iterative closest point method to align the test face with the AFM. ICP is a rigid registration method; hence, the test face is not deformed at all. TPS-based methods are completely guided by the landmarks, whereas ICP needs a coarse initialization. Intuitively, ICP will benefit from using category-specific AFMs, as the rigid registration is not able to cope with shape differences very well. A more similar average face will ensure that the dense correspondence will be established between points that have better structural correspondence. The TPS-based method will also benefit from category-specific AFMs, albeit for another reason: A more similar average face means that the test surface will be deformed less, and discriminatory information will not be lost.

We review the ICP and TPS algorithms here, along with the Procrustes analysis, the automatic landmark localization algorithm, and the regular resampling that are used by both.

### 3.1 *Procrustes Analysis*

Procrustes analysis is a statistical tool for the analysis of geometrical shapes.[13] A shape (or equivalently a *figure*) $P$ in $\mathbb{R}^p$ is represented by $l$ landmarks. Two figures $P:l\times p$ and $P':l\times p$ are said to have the same shape, if they are related by a special similarity transformation:

$$P' = \alpha P\Gamma + 1_l\gamma^T, \tag{1}$$

where the parameters of the similarity transformation are a rotation matrix $\Gamma:p\times p$, $|\Gamma|=1$, a translation vector $\gamma:p\times 1$, a positive scaling factor $\alpha$, and $\mathbf{1}_l$ is a vector of ones. By using the generalized Procrustes analysis, it is possible to derive a *consensus shape* for a collection of figures,[14] which is then used in registering new shapes into alignment with the collection by an affine transformation:

1. Center all shapes $P_i$:

$$M = \frac{1}{N}\sum_{i=1}^{N} P_i, \tag{2}$$

$$P_i = P_i - M. \tag{3}$$

2. Bring the shapes to a common scale.
3. Set the consensus shape $Y$ equal to the first shape $P_1$, as an initialization.
4. For $i=2,3,\ldots,N$, rotate $P_i$ to fit $Y$. $Y$ is reevaluated after each update of $P_i$ as

$$Y = \frac{1}{i}\sum_{j=1}^{i} P_j. \tag{4}$$

5. Repeat updating the $P_i$ and $Y$, while monitoring the residual sum-of-squares:

$$S_r = N[1 - tr(Y_t Y_t^T - Y_{t-1}Y_{t-1}^T)], \tag{5}$$

where $Y_t$ is the consensus at iteration $t$, and $Y_{t-1}$ is the consensus at iteration $t-1$. When $S_r$ is below a threshold (e.g., 0.0001, as suggested by Gower in Ref. 14), stop the iterations, and output the consensus shape.

### 3.2 *Iterative Closest Point*

In this section, we summarize the ICP procedure as defined in Ref. 6. Define the *unit quaternion* as a four vector $\vec{q_H}=[q_0 q_1 q_2 q_3]^T$, with $q_0\geq 0$, and $q_0^2+q_1^2+q_2^2+q_3^2=1$. The $3\times 3$ rotation matrix $H$ generated by this quaternion is:

$$H = \begin{bmatrix} q_0^2+q_1^2-q_2^2-q_3^2 & 2(q_1q_2-q_0q_3) & 2(q_1q_3+q_0q_2) \\ 2(q_1q_2+q_0q_3) & q_0^2+q_1^2-q_2^2-q_3^2 & 2(q_2q_3-q_0q_1) \\ 2(q_1q_3-q_0q_2) & 2(q_2q_3+q_0q_1) & q_0^2+q_1^2-q_2^2-q_3^2 \end{bmatrix}. \tag{6}$$

Let $\vec{q_T}=[q_4 q_5 q_6]^T$ be a translation vector. Together with $q_H$, they make up the complete registration state vector $\vec{q}=[\vec{q_H}|\vec{q_T}]^T$. Let $P=\{\vec{p_i}\}$ be a data point set to be aligned with the model point set $Y=\{\vec{y_i}\}$. The two models will have the same number of points, and ICP will put the points with the same indices into one-to-one correspondence. Denoting the number of points in each model with $N$, the objective function minimized by the ICP procedure is:

$$f(\vec{q}) = \frac{1}{N}\sum_{i=1}^{N} \|\vec{y_i} - H(\vec{q_h})\vec{p_i} - \vec{q_T}\|^2. \tag{7}$$

Denoting the center of mass of the point set $P$ with $\mu_p$, and that of the model set with $\mu_y$, the cross-covariance matrix $\Sigma_{py}$ is given by:

$$\Sigma_{py} = \frac{1}{N}\sum_{i=1}^{N} [(\vec{p_i} - \vec{\mu_p})(\vec{y_i} - \vec{\mu_y})^T]. \tag{8}$$

The cyclic components of the matrix $A_{ij}=(\Sigma_{py}-\Sigma_{py}^T)_{ij}$ are used to form a column vector $\Delta=[A_{23}\ A_{31}\ A_{12}]^T$, which in turn is used to form a symmetric $4\times 4$ matrix $Q(\Sigma_{py})$:

$$Q(\Sigma_{py}) = \begin{bmatrix} tr(\Sigma_{py}) & \Delta^T \\ \Delta & \Sigma_{py}+\Sigma_{py}^T - tr(\Sigma_{py})I \end{bmatrix}, \tag{9}$$

where $I$ is the $3\times 3$ identity matrix. The optimum rotation is given by the unit eigenvector $\vec{q_H}=[q_0 q_1 q_2 q_3]^T$ corresponding to the maximum eigenvalue of the matrix $Q(\Sigma_{py})$. The optimum translation vector is given by:

$$\vec{q_T} = \vec{\mu_y} - H(\vec{q_H})\vec{\mu_p}. \tag{10}$$

We use $\vec{q}(P)$ to denote the point set $P$ after the application of the transformation represented by $\vec{q}$. The ICP algorithm computes and applies these transformations iteratively. A step-by-stop description of the algorithm follows:

1. Initialize ICP by setting $P_0=P$, $\vec{q_0}=[1,0,0,0,0,0,0]^T$, and $k=0$. The registration is defined relative to $P_0$, which requires a coarse registration. Steps 2 to 5 are applied iteratively, until convergence is achieved within a tolerance $\tau$.
2. Compute the closest points: $Y_k=\mathcal{C}(P_k,Y)$. The computational cost of this step is $O(N_pN_y)$ in the worst case, where $N_p$ is the number of points on the registered point cloud, and $N_y$ is the number of points on the model shape.
3. Compute the registration: $\vec{q_k}=\mathcal{Q}(P_0,Y_k)$. The computational cost is $O(N_p)$.
4. Apply the registration: $P_{k+1}=\vec{q_k}(P_0)$. The computational cost is $O(N_p)$.
5. Terminate the iteration if the change in the mean square error is below the preset threshold $\tau$. A heuristic value for $\tau$ is a multiple of $[tr(\Sigma_y)]^{1/2}$, where $\Sigma_x$ is the covariance matrix of the model shape, and the square root of its trace is a rough indicator of model shape size.

In our simulations, the test face acts as the model shape, and the cropped gallery face is aligned to it. The points of the test scan that are put into one-to-one correspondence with the model are retained, and the rest are discarded. If the registration is correct, this procedure automatically gives a good cropping, possibly including hair and clutter removal. In rare cases where a streak of hair extends over the face center, the registration and the subsequent classification will be inaccurate.

Previous work on ICP shows that a good initialization is necessary for fast convergence and an accurate end result. We compare several approaches for the coarse registration. In our first approach, the point with the greatest depth value is assumed to be the tip of the nose, and a translation is found to align it to the nose tip of the AFM. This is the fastest and simplest heuristic used in the literature,[15] and we expect it to perform well with near-frontal faces. In the second approach, we use the manually determined nose tip (i.e., ground truth) in the coarse alignment. In the third approach, we use the Procrustes analysis to bring seven manually determined landmark points (inner and outer eye corners, nose tip, and the mouth corners) into alignment with the average face model. Last, we use Procrustes analysis to align automatically determined landmarks with the average face model. The automatic landmarking errors are not random and cannot be simulated by injecting noise to the manually determined landmarks, except by modeling the specific landmarking procedure.

### 3.3 *Thin-Plate Splines*

The TPS model expresses the bending energy of a thin metal plate fixed at certain points.[16] At the heart of the model is a special surface function:

$$z(x,y) = -U(r) = -r^2\log(r^2), \tag{11}$$

with $r=(x^2+y^2)^{1/2}$ equal to the Euclidean distance of point $(x,y)$ to the origin.

For a set of anchor points $P_i=(x_i,y_i)$, $i=1\ldots n$, the TPS interpolation is a vector-valued function $f(x,y)=[f_x(x,y),f_y(x,y)]$ that maps the anchor points to their specified homologs $P_i'=(x_i',y_i')$, $i=1\ldots n$ and specifies a surface that has the least possible bending, as measured by an *integral bending norm*. We will give a summarizing mathematical specification of the model here.

Define $r_{ij}=|P_i-P_j|$ to be the distance between the points $i$ and $j$. Also define the following matrices:

$$K=\begin{bmatrix} 0 & U(r_{12}) & \ldots & U(r_{1n}) \\ U(r_{21}) & 0 & \ldots & U(r_{2n}) \\ \ldots & \ldots & \ldots & \ldots \\ U(r_{n1}) & U(r_{n2}) & \ldots & 0 \end{bmatrix}, \tag{12}$$

$$P=\begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \ldots & \ldots & \ldots \\ 1 & x_n & y_n \end{bmatrix}, \tag{13}$$

and

$$L=\left[\begin{array}{c|c} K & P \\ \hline P^T & O \end{array}\right], \tag{14}$$

where $O$ is a $3\times 3$ matrix of zeros. Let $V$ be a matrix made up of the homologs of the anchor points:

$$V=\begin{bmatrix} x_1' & x_2' & \ldots & x_n' \\ y_1' & y_2' & \ldots & y_n' \end{bmatrix}. \tag{15}$$

Define $w_i$ and the coefficients $a_1$, $a_x$, and $a_y$ as:

$$L^{-1}(V|\mathbf{0},\mathbf{0},\mathbf{0}) = (w_1,w_2,\ldots,w_n,a_1,a_x,a_y)^T. \tag{16}$$

The function $f(x,y)$ is defined as:

$$f(x,y) = a_1 + a_x x + a_y y + \sum_{i=1}^{n} w_i U(|P_i-(x,y)|). \tag{17}$$

$f(x,y)$ minimizes the nonnegative integral bending norm $I_f$ over all such interpolants:

$$I_f = \int\int_{\mathbb{R}^2}\left[\left(\frac{\partial^2 f}{\partial x^2}\right)^2 + 2\left(\frac{\partial^2 f}{\partial x\,\partial y}\right)^2 + \left(\frac{\partial^2 f}{\partial y^2}\right)^2\right]\mathrm{d}x\mathrm{d}y. \tag{18}$$

The TPS function $f(x,y)$ is invariant under rotations and translations. It maps the landmarks $P_i$ to their homologs $P_i'$, and defines a smooth interpolation for the rest of the points on the surface. $P_i$ and $P_i'$ taken together exactly specify the function $f(x,y)$ and are therefore crucial to the accuracy of the deformation. In the TPS-based registration method we use, $P_i$ are the landmarks defined on the average face model, and $P_i'$ are the corresponding landmarks of the test scan.

### 3.4 *Automatic Landmark Localization*

Registration of facial images is usually guided by a few fiducial points of established correspondence (e.g., nose tip and eye and mouth corners). These anchor points have a great influence on the resulting registration. There are very few completely automatic systems for face recognition. Most research results are reported on cropped and aligned faces, or the existence of a few manually located landmarks is assumed. This is apparently necessary, because the quality of the landmarks can be of greater consequence to the final accuracy than the classification algorithm itself.

To understand the extent of the dependence on landmarks, we have contrasted manually obtained ground-truth with the results from a recent automatic landmark localization algorithm.[17–19] In this section, we briefly summarize this algorithm.

The method we employ is a coarse-to-fine approach, based on local feature information. The inner and outer eye corners, the nose tip, and the mouth corners (seven landmarks) are localized. These landmarks were selected because statistical information (rather than heuristics) can guide the search, as they correspond to distinguishable local features (whereas the tip of the chin, for instance, does not correspond to a very clear local feature). During the training phase, $7\times7$ patches are cropped from around each landmark in the downsampled depth image ($60\times80$). These patches are points in a 49-dimensional space, and their distribution is modeled with a mixture of factor analysers. We employ the Incremental Mixtures of Factor Analyzers (IMoFA) algorithm that automatically tunes the model complexity to the complexity of the data distribution.[20]

The main assumption in factor analysis is that the $d$–dimensional data $\boldsymbol{x}$ are generated in a $p$–dimensional manifold, represented with $\boldsymbol{z}$, with $d\gg p$. The vectors that span the lower dimensional manifold are called *factors*, and the relationship between the factor space and the data space is given by

$$\boldsymbol{x}-\boldsymbol{\mu}=\Lambda z+\boldsymbol{\epsilon}, \tag{19}$$

where $\boldsymbol{\mu}$ is the data mean, $\boldsymbol{\Lambda}$ is the factor loading matrix, and $\boldsymbol{\epsilon}$ models the independent variance at each data dimension. The IMoFA algorithm starts with a simple, one-factor representation of the data and gradually increases the complexity of the model while monitoring a separate validation set. The resulting model is a mixture with components that have different numbers of factors, thereby modeling clusters with different intrinsic dimensionality.

During the testing phase, this generative model is used to produce likelihoods for each point on the image. The highest-likelihood locations for each landmark are passed to GOLLUM (Gaussian Outlier Localization with Likelihood Margins), a structural correction algorithm that can recover from localization errors.[17] The remaining errors are locations on the face with characteristics that are statistically similar to the landmark in question, and they conform to the general face pattern within a margin, as they have passed through structural correction. For instance, the corners of the eyebrows may be selected as eye corners, or the cleft between the chin and the mouth may be detected as the mouth. The landmark detection was performed on the 3-D depth map, which is found to be less informative than 2-D images but much more robust to illumination changes.[18]

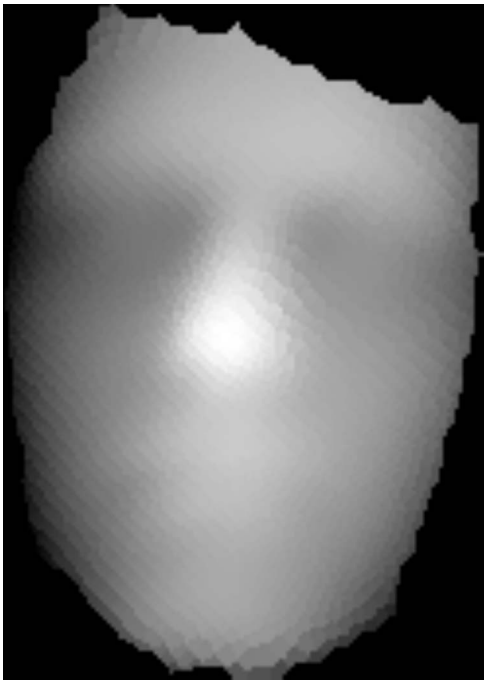

**Fig. 2** The AFM generated with the method of İrfanoglu *et al.*[10]

The coarse localization is complemented with a fine-level search on the original $480\times640$ range image. In the fine-level search, only a relatively small ($41\times41$) area around the coarsely located landmark is searched. The first and second depth gradients are estimated in vertical and horizontal directions and convolved with kernels learned during the training stage. Each kernel is a map with values ($\pm1$), and the cascade of kernels is constructed incrementally to produce the maximum output for the desired landmark shape. These kernel outputs are combined, and the highest response is taken as an indicator of the presence of the sought landmark. This approach is computationally simple, fast, and improves the quality of the coarse landmarks. See Ref. 19 for more details on the automatic landmarking procedure.

### 3.5 *Regular Resampling*

The sampling of 3-D information can be a cause of error in the comparison stage. We have used a simple triangle-based nearest-neighbor interpolation for regular resampling of depth information from the aligned models.[12] The method relies on a Delaunay tesellation of the original data, such that none of the original data points are contained in any circumspheres of the simplices of the tesellation. Then, for each $x$ and $y$ position of the regular grid, the nearest vertex in the tesellation is located, and its depth value is used. The grid boundaries are derived from the generic cropped AFM, at the resolution of the range sensor. Consequently, the effect on the AFM is that the points are slightly moved to conform with a regular $x-y$ grid. We will provide a graphical intuition on why regular resampling is beneficial in Sec. 5.

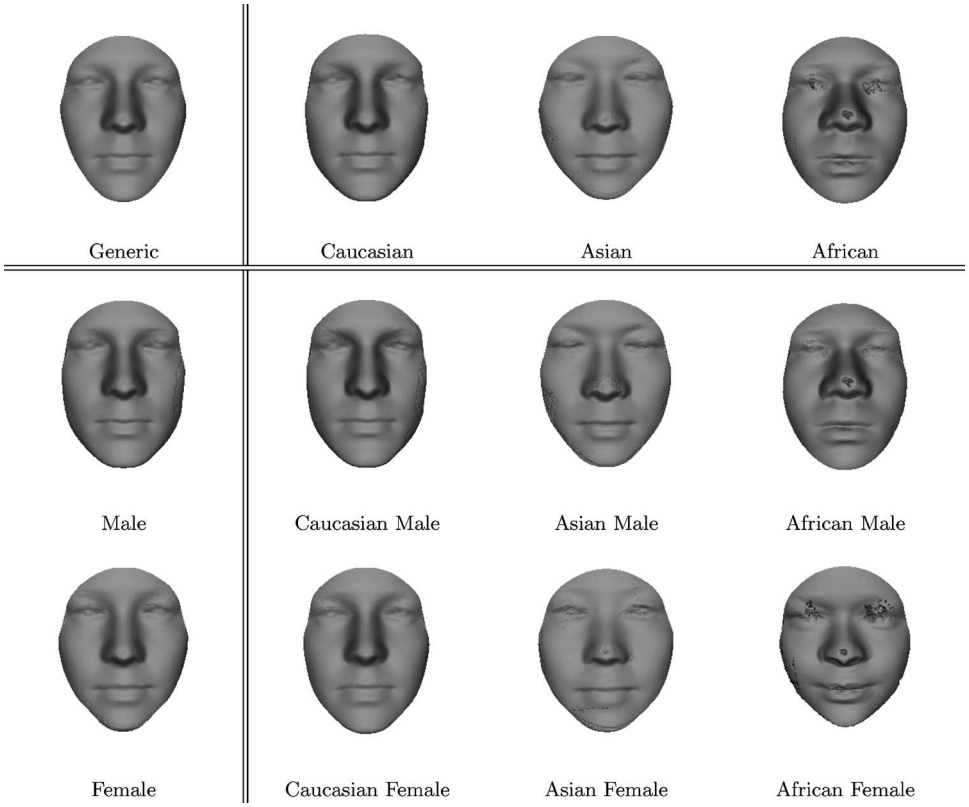


**Fig. 3** Average faces for different morphology and gender combinations.

## 4 Average Face Models

In Ref. 10, a method for generating the AFM was described. In this method, training faces are manually landmarked, and the TPS method is used to register all the training faces to a *consensus shape*. One of the faces is selected to be the AFM candidate, and from each of its vertices, distances to each other training face are computed. If a vertex has no corresponding point in every training face closer than a threshold value, that vertex is trimmed. The trimmed vertices are usually the ones at the boundaries of the face area. The remaining vertices, and their corresponding closest points for each training face are averaged to create the final AFM. This procedure creates a very smooth facial surface, as shown in Fig. 2.

In this section, we describe a novel method of generating the AFM, which is experimentally determined to produce better results. We also describe two methods of generating category-specific AFMs, a *cognitive method* that is based on the other-race effect, and a *clustering approach* that is based on clustering in the shape space, respectively.

### 4.1 *Construction of an AFM*

Using a set of landmarked training faces, we generate an AFM with the following procedure:

1. Using Procrustes analysis, a consensus landmark distribution is found on the training set.
2. The landmarks of the consensus shape are rectified to present a fully frontal face, centered at the origin of the 3-D coordinate system. This heuristic is used to facilitate the use of the transformed range image in later stages. Rectification is achieved by rotating the face so that the eye and mouth planes, as found by fitting surfaces to the appropriate landmarks, are parallel to the $x$ axis and the $z$ axis as much as possible.
3. TPS deformation is computed for the training faces, which warps the landmarks of each face to the consensus shape perfectly and interpolates the rest of the points.
4. The depth values of the interpolated face are resampled from a regular $x$-$y$ grid. This ensures that all added faces have points with overlapping $x$ and $y$ values, and the depth values are given for matching points. For the simple range image representation, this extra offline computation leads to much faster online model comparison.
5. A cropping mask is defined to encompass the facial area, and faces are cropped before they are added to the average face model. For creating this mask, we calculate the maximum distance from the nose tip to any landmark in the consensus shape. We add a 10 percent margin to this distance and retain all points closer than this value to the nose.
6. After all the training faces are added, depth values are simply averaged. Samples of AFMs generated with this method can be seen in Fig. 3.

The database we use was collected with a laser sensor that typically generates holes (especially at the eyes and the mouth) or other artifacts. The preprocessing sometimes falls short of repairing larger errors. On the other hand, the proposed AFM generation method is robust with respect to

the number of images: Generating AFMs with a training set three times bigger than the original set resulted in very similar AFMs, with an average of 0.4($\pm$0.3) mm point-to-point distance to the old AFMs. Since the landmark locations are retained, the facial characteristics are not smoothed out, even when a large number of samples are used for AFM generation.

### 4.2 *Cognitive Approach to Multiple AFM Generation*

When humans see faces, they perceive each face individually, and not merely as specimens of a generic object category.[21] The mechanism that operates for enhanced within-category recognition is what we can call *expertise* of that category. Tong *et al.* remark that expertise-related visual judgements involve enhanced within-category discrimination, as opposed to between-category discrimination, the former requiring a magnification of differences in similar objects, and the latter calling for a neglecting of differences to group similar items together.[22]

People have great difficulty recognizing faces of another race if they are not exposed to these faces for prolonged periods. This phenomenon is termed the *other-race effect*. In the light of the Tong *et al.* experiments, it seems reasonable that during the acquisition of face expertise, the transformations learned by the visual cortex serve to magnify the differences between individual faces, as indicated by the statistical distribution of the encountered facial features. By this reasoning, the other-race effect suggests that different face morphologies exhibit different statistical distributions of distinguishing characteristics. Valentine has previously employed principal component analysis to find different subspace projections and obtained results that mimic the other-race effect.[23]

We stress that our aim is not to detect the *race* of a person; therefore, we use the term *morphology* to denote similar facial surface morphology characteristics. Based on the cognitive cues, we predict better recognition rates if the faces are clustered into morphological or gender groups that exhibit greater intragroup similarity and the discriminative features are learned within each group separately. This is not trivially true for all pattern recognition applications, as the grouping reduces the number of training samples and consequently runs the risk of impairing learning conditions. In this approach, the gender and morphology were determined manually, by visual inspection of the faces.

### 4.3 *Clustering Approach to Multiple AFM Generation*

If our hypothesis of metaclassification is correct, we expect morphology and gender to be discriminating dimensions of the face space. However, we do not want to categorize faces into races explicitly, as this approach has ethical consequences. Can the gender and race determination during the training (and possibly, in the testing) stage be evaded? For our simulation purposes, we have roughly assigned facial images into African, Asian, and Caucasian morphological face classes. The other-race effect suggests that racial-morphology-based clusters *exist* in the face space, and an unsupervised clustering method can recover those clusters, among other meaningful structures. Thus, it is not necessary to recover the race and gender of a person; the clustering will hopefully provide us with a useful set of average faces to serve in metaclassification with increased discrimination within clusters.

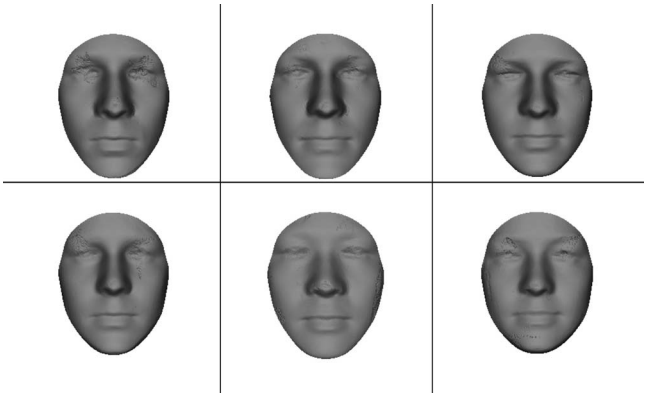

**Fig. 4** Shape space cluster means.

By performing a shape space–based clustering, we will also answer an interesting question: Are gender and racial morphology major factors in determining the face shape? The intuitive answer is affirmative, but the clustering provides us with an explicit way to test this hypothesis.

We propose to take a straightforward race- and gender-blind clustering approach with the $k$-means algorithm. The clustering is performed on the aligned coordinates of seven facial landmarks. We specify the number of clusters for the shapes and initialize the cluster consensus shapes by random selection from the training samples. At each iteration, we align the training samples to the consensus shapes of the clusters via Procrustes analysis and assign each sample to the cluster with least-average distance to the cluster consensus. We then reestimate the cluster consensus shapes from the samples assigned to the cluster and iterate until the total distance stabilizes. The number of clusters was set to six to allow comparison with the cognitive approach.

The clustering gives us a number of cluster consensus shapes and assigns each training face to one of these clusters, without ensuring that all morphological categories are represented. We apply our AFM generation algorithm to these reduced training sets separately, and obtain one AFM for each cluster. These models can be seen in Fig. 4.

## 5 Experiments

### 5.1 *The 3-D Face Database*

We use the FRGC 2D-3D version 1 face database in our experiments.[24] The database contains 943 near-frontal depth images from 275 subjects, stored in a point cloud representation. The laser scanner that was used to collect the database samples points from the surface of the face in regular intervals. Although the number of points per face can be different according to distance to the scanner at the time of acquisition, the coordinates of the points are given in a proper metric scale.

We use images from 195 subjects, with one training face in the gallery and 1 to 4 test faces (for a total of 659 test faces). This experimental setup is used in Ref. 33, and we chose to use the same setup to allow a direct comparison with a host of techniques. Subjects with only a single image in the dataset and images with incorrect texture-depth cor-

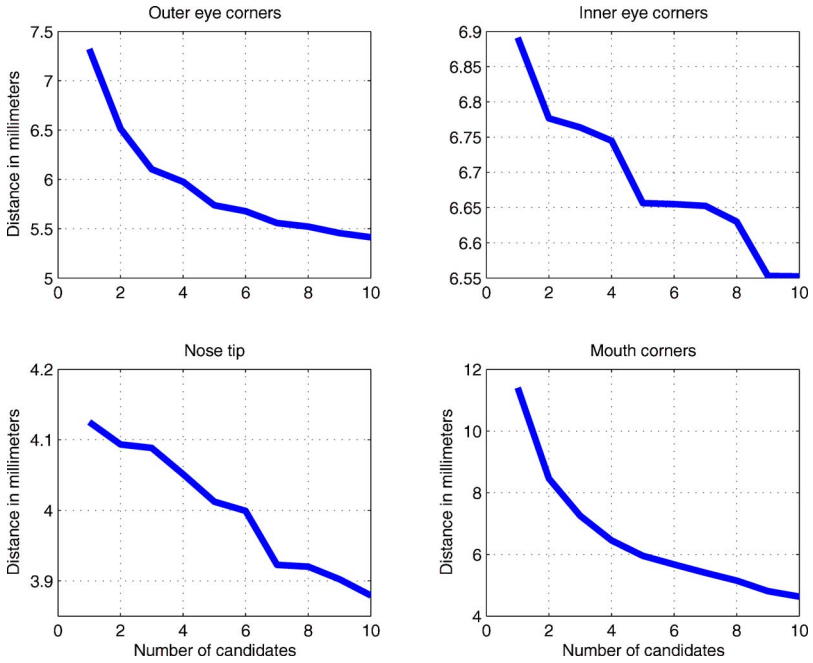


**Fig. 5** Accuracy of the automatic landmarking method, as indicated by average distance to true landmark in millimeters.

respondence are not used. We work only with 3-D information for landmarking and registration; 2-D is not used at all. We design a number of experiments to answer various questions. Each subsection deals with a particular question and reports relevant simulation results. The overall system has many dimensions, ruling out a factorial experimentation protocol.

For each method, we run a recognition experiment and a verification experiment. For the recognition experiments, the rank-1 recognition rate (R1) is reported. The point set difference (PSD) is used as a distance metric for recognition. In the verification experiments, each of the 659 test faces is used for one genuine and 194 false claims. Thus, the number of false claims is two orders of magnitude higher. The equal error rate (EER), for which the false acceptance rate is equal to the false rejection rate, is reported under these conditions.

### 5.2 *Preprocessing*

The preprocessing of 3-D information consists of several steps. Missing points and holes can be filled by local interpolation or by making use of facial symmetry.[25] However, it is a well-known fact that faces are not truly symmetrical. Gaussian smoothing and linear interpolation are used to eliminate irregularities in both texture and range images.[2,3,26] Background clutter and hair artifacts axe usually manually removed in the literature,[2,3,27] whereas in our approach, the cropping of the face area is fully automatic.

If there is excessive acquisition noise in a dataset, it will be difficult to assess the accuracy of algorithms, as the deterioration due to noise can be more pronounced than improvement due to a better algorithm. Datasets are frequently cleared of noisy samples.[26,28] Mean and median filters are also employed to reduce local noise.[5,26] The preprocessing used in this paper includes a $9\times9$ mean and a $9\times9$ median filter applied in succession, followed by linear interpolation for filling the gaps. The filter sizes are experimentally determined and are not optimized for actual shape sizes.

### 5.3 *Automatic Landmarking*

Figure 5 shows the accuracy of the automatic landmarking procedure in terms of millimeter-distance from the manually obtained ground-truth. The evaluation considers one to ten candidates for each landmark location, as indicated by the $x$ axis. The curves begin at the average distance for the best candidate and decrease as better locations are found among the candidates. The manually annotated ground-truth is accurate up to several millimeters, and a lower distance to the ground-truth indicates a better localization. The quality of landmarking also depends on the type of landmark; the nose tip is easier to find in 3-D, as expected.[29–32] The reader is referred to Ref. 19 for more details on the performance of the algorithm.

### 5.4 *Classification*

The most natural way of classification for point sets in dense correspondence is using the point set distance (PSD) measure. Since all points are in full correspondence, we have two 3-D vectors of the same length, and the PSD is simply the sum of all Euclidean distances of the points in the query and gallery faces:

$$\arg\min_{j} \sum_{i\in S} \|\mathbf{x}_i - \mathbf{a}_{i,j}\|, \tag{20}$$

where $\boldsymbol{x}_i$ is the $i$'th point of the test scan, and $\boldsymbol{a}_{i,j}$ is the $i$'th point of the $j$'th gallery sample. $S$ denotes the set of points put into dense correspondence.

**Table 1** Effect of coarse alignment on ICP.

| | Nose-tip heuristic | Automatic landmarks + Procrustes | Nose ground truth | Manual landmarks + Procrustes |
|---|---|---|---|---|
| R1 | 82.85 | 87.86 | 90.60 | 92.11 |
| EER | 14.25 | 8.12 | 6.60 | 6.20 |

**Table 2** Simulation of TPS registration with deformation EER.

| | Gender | Morphology | Gender and morphology |
|---|---|---|---|
| Generic AFM | 16.79 | 18.50 | 13.87 |
| Specific AFM | 14.64 | 16.97 | 11.47 |

In the cases where we employ regular resampling, the $x$ and $y$ coordinates of these points are the same across all samples, and therefore need not be taken into account. The 3-D PSD becomes equivalent to an L1 distance computed on the depth values, and for classification, the smallest L1 distance between gallery images and the test images is sought to indicate the correct class $j$:

$$\arg\min_j \sum_{i \in S} |\boldsymbol{x}_i^z - \boldsymbol{a}_{i,j}^z|, \tag{21}$$

where $\boldsymbol{x}_i^z$ is the depth value of the $i$'th point of the test scan, and $a_{i,j}^z$ is the depth value of the $i$'th point of the $j$'th gallery sample.

For authentication, a threshold $\tau$ is used to validate the claimed identity $j$:

$$\sum_{i \in S} |z_i - a_{i,j}| < \tau, \tag{22}$$

and $\tau$ can be used to tune a trade-off between false acceptance rate (FAR) and false rejection rate (FRR). The equal error rate (EER) is the reported error rate for the particular $\tau$ for which FAR and FRR are equal on the test set. While it makes sense to use one-to-one ICP for the authentication setting in practice, obtaining a lower EER by just changing the AFM set is still indicative of a better registration, and therefore it is reported for some of our simulations. Furthermore, by storing only the projection coefficients for gallery faces, the storage requirement can be greatly reduced. (For instance, instead of about 32,000 3-D points, a 50-dimensional vector can be stored per subject.) In this case, the subject-specific 3-D model is not available for registration, and an AFM will be used for authentication.

### 5.5 *Coarse Registration*

Table 1 shows the effect of coarse alignment methods on ICP-based registration. As we have suspected, the nose-tip heuristic performs the worst. Automatic localization of seven landmarks and using Procrustes alignment works better than the nose-tip heuristic. For ICP, using the nose ground-truth works well in this dataset, because the faces we deal with are mostly upright and frontal. Ideally, three landmarks should be used to accomodate greater pose differences in 3-D. Finally, the seven manual landmarks with the Procrustes analysis give us an upper bound on the performance of the ICP-PSD method.

We have also contrasted our AFM construction method with the method of Irfanoğlu *et al.*[10] on ICP. Manual landmarks were used, and initialization was by Procrustes alignment. With their smoother AFM, a rank-1 recognition rate of 86.34 was achieved, as opposed to our 92.11 percent. Similarly, the EER was higher with their AFM by more than 2 percent.

### 5.6 *Meta-Classification*

Do metaclassification and more specialized individual experts increase discrimination? We have tested this hypothesis by employing the average faces that are generated from groups of training faces. We have grouped the training samples for gender and morphology and generated average face models (AFM) for each group. Figure 3 shows range images obtained for the average face, for male and female averages, for three morphological group averages (roughly Caucasian, Asian, and African), and for all combinations, respectively. The morphology does not correspond to a clear-cut distinction. The morphological group of a given face will be determined by its proximity to the average face of the group.

In Table 2, the authentication experiment results with or without specific average faces are shown for TPS-based registration. We have supplied both the generic-AFM-based system and the specific-AFM-based system with the categorical information, which improves the accuracy by itself. However, any improvement in the specific system with regard to the corresponding generic system is strictly due to better registration. We have computed distances between the test face and the gallery faces with an $L_1$ distance metric and trimmed the worst 2 percent of the correspondences. The EER without any use of categorical information is 20.10 percent. The results reported under these conditions show that specific AFM usage is beneficial in this case. We stress that the main purpose of the PSD is to evaluate the effect of AFM usage; the EER is generally too high for these experiments because of deformations in the registration, but the gain obtained by introducing specific models is more pronounced in comparison to ICP-based experiments.

The point set distance after aligning the face to the female and male averages can be used in gender classification. This simple method works with 80 percent accuracy.

### 5.7 *Shape Space Clustering*

The shape space method divides the training samples into clusters of similar size. The number of clusters should be small to have a maximal computational gain and can be increased as long as the accuracy increase is considered worth the computational cost. However, the number of training samples per cluster should be sufficiently high (depending on the quality of the range data and preprocessing); otherwise, the constructed AFM will be poor. We have specified six clusters, to allow a comparison with the full

**Table 3** Comparison of specific AFMs; rank-1 recognition rates.

| | Manual landmark + ICP | Automatic landmark + ICP | Manual landmark + TPS | Automatic landmark + TPS |
|---|---|---|---|---|
| Generic | 92.11 | 87.86 | 52.20 | 42.64 |
| Gender | 90.14 | 86.65 | 54.63 | 45.52 |
| Morphology | 89.98 | 86.80 | 53.87 | 44.92 |
| Gender and morphology | 91.05 | 86.49 | 56.90 | 47.95 |
| Shape space derived | 93.78 | 91.20 | 47.65 | 41.58 |

morphology-gender combination case, and we ran our algorithm on the training part of the FRGC version 1 dataset. Figure 4 shows the cluster means.

Table 3 shows recognition rates for ICP- and TPS-based systems with manual or automatic landmarks. The first row shows the results obtained with a single generic AFM. The next three rows show results with gender-, morphology-, and gender+morphology–based specific AFMs. The results for the last row are obtained with six shape space–derived clusters. For this last case, the registration does not benefit from the injection of categorical information, and each test sample is compared with all the training samples. The best result is obtained with shape space–derived specific AFM and ICP (93.78 percent). As a baseline experiment, we have also tested one-to-all ICP, where each test face is registered to each gallery face for distance computation. With manual landmarks, we have obtained 89.07 percent rank-1 recognition rate, which further demonstrates the usefulness of AFM-based registration.

We have also tested adding the previously excluded single-image subjects to the gallery. This decreases the accuracy of all the recognition methods, as the 195-class problem becomes a 275-class problem. The ICP method (with manual landmarks and generic AFM) suffers 2.6 percent accuracy loss with the enriched gallery.

Our cognitively motivated subspace hypothesis is confirmed, if the shape space clustering automatically creates clusters with a dominant gender, or puts samples from one race into a single group. (Currently available 3-D face datasets do not have a balanced morphology-gender distribution, where the cognitive hypothesis can be tested more thoroughly.) This is what more or less happens, and we do have clusters dominant in a single gender or a single morphology. This effect is more pronounced when we increase the size of the training set by neutral images of 371 subjects from the FRGC version 2 dataset. Figure 6 shows the distribution of morphology and gender in each group as pie charts.

### 5.8 *Regular Resampling and Subspace Projection*

In this section, we apply the eigenface method to 3-D range images registered to different AFMs with ICP. A projection from the space of 3-D points can be computed only if the number of points is the same for each query and gallery face. Our method ensures that this is indeed the case. If the accuracy loss after the projection is small, this means that the registration was successful, and the selected subspace dimensionality is sufficient to capture the discriminative dimensions of variation.

The depth values of the range images are not sampled regularly from an $x$-$y$ grid, and so far we have computed 3-D point distances, instead of a simpler one-dimensional (1-D) comparison based on the depth. Since we deal with aligned shapes in dense correspondence, applying a regular resampling will make it possible to discard two dimensions from the point cloud, making the subsequent comparison and subspace projection easier. We will also show that the regular resampling helps classification by making the distance measurement more accurate.

The dimensionality of the subspace is determined heuristically. We set the number of eigenvectors so that at least 95 percent of the variance is accounted for. The eigenface method is contrasted with the PSD method, which, due to regular resampling, uses the sum of squared distances of depth values only.

The regular resampling lightens the computational burden of comparing the test sample with gallery images. Furthermore, the computed projection allows us to store much smaller gallery faces. For example, in experiments with the generic AFM, roughly 32,000-dimensional face vectors are represented with 50-dimensional vectors after the subspace projection. This means that the comparison with gallery faces will be much faster. Table 4 shows that the accuracy loss due to subspace projection does not exceed 1 percent, if there are sufficient training samples. For the gender and morphology combination, the training set is very limited, and consequently there are only 15 eigenvectors with nonzero eigenvalues in one of the groups. (Henceforth, we denote the number of eigenvectors with $p$.) This number is too small to represent the facial variation, and the accuracy decrease is about 3 percent. The morphology results were obtained with $p=33$, and the gender results with $p=49$. For the results in this section, we have grouped African faces and Caucasian faces into a single category, as we had too few samples from the African category (i.e., two clusters for morphology, and four clusters for gender and morphology runs).

Before analyzing the results, we will give some additional experimental results reported on this dataset. In Ref. 33, Gökberk *et al.* use the same experimental protocol as ours to compare a wealth of classification methods: point set difference; non-negative matrix factorization (NMF);

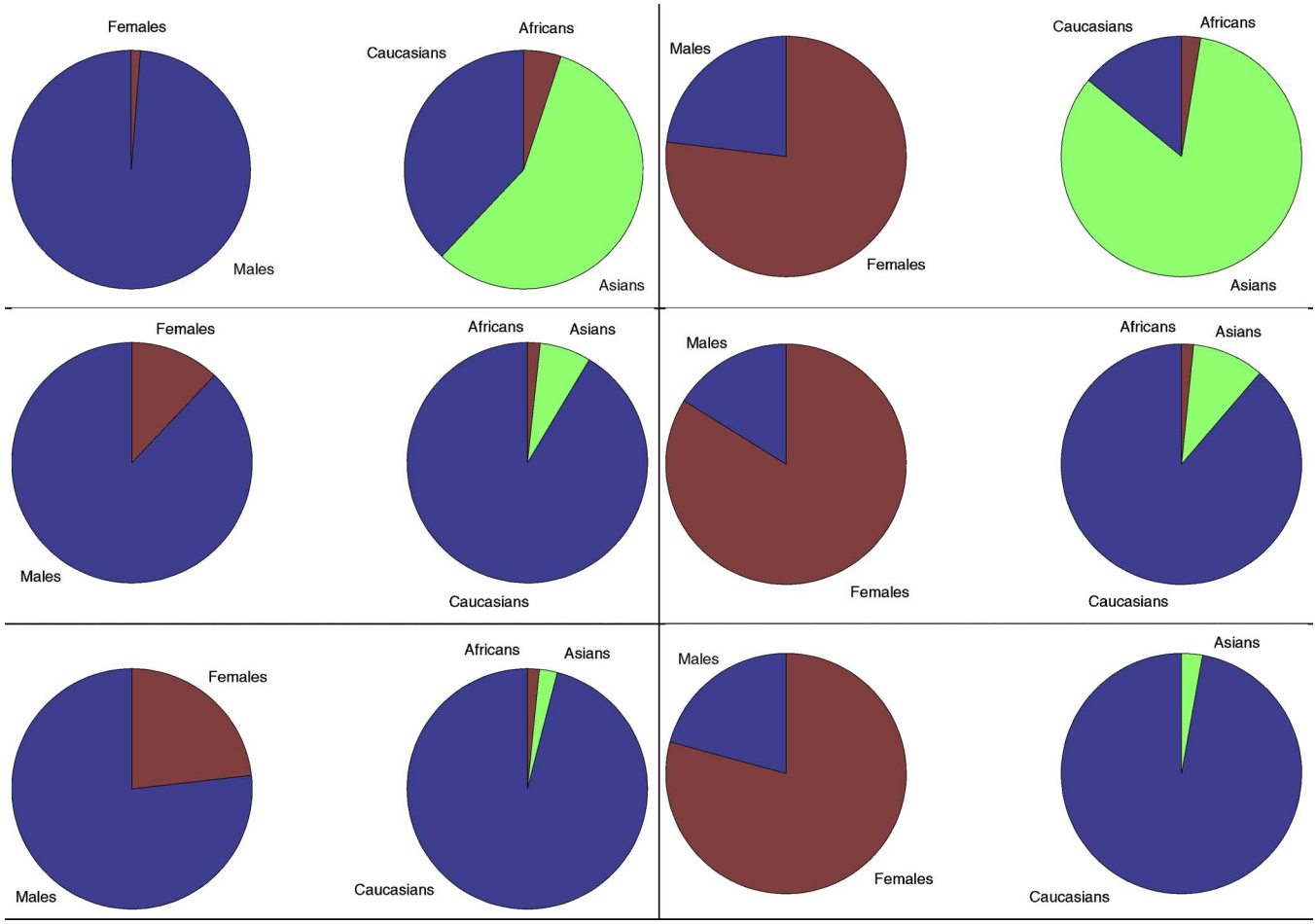


**Fig. 6** Shape space clustering distributions on the enriched training set. For each cluster, the gender and morphology distributions are shown in separate pie charts. For six clusters, we have three dominantly male clusters and three dominantly female clusters, and one dominantly Asian cluster per gender. The African class has too few samples to create a perceptible effect.

and independent component analysis (ICA) coefficients for point clouds; discrete cosine transform (DCT), discrete Fourier transform (DFT), PCA, linear discriminant analysis (LDA), and ICA projections on depth images; shape indices; mean and principal curvatures; 3-D voxel DFT coefficients; and 2-D Gabor wavelet coefficients. Manually annotated landmark positions were used for an AFM-based ICP registration (with an AFM generated with the method of Irfanoğlu *et al.*[10]). The best classification results are based on shape indices (90.06 percent), principal directions (91.88 percent), and surface normals (89.07 percent). The best accuracy after classifier fusion is 93.63 percent, obtained with modified plurality voting.[33]

For projection-based methods, the input dimensionality needs to be fixed. Subsequently, PCA is not directly applicable in 3-D prior to the cropping (which in turn uses the

**Table 4** Subspace projection after ICP+resampling; rank-1 recognition rates.

| | Manual landmark PSD | Manual landmark Eigenface | Automatic landmark PSD | Automatic landmark Eigenface |
|---|---|---|---|---|
| Generic | 98.18 | 98.03 | 98.03 | 97.88 |
| Gender | 96.81 | 96.21 | 95.30 | 95.90 |
| Morphology | 96.51 | 96.05 | 95.60 | 94.84 |
| Gender and morphology | 96.97 | 93.78 | 94.99 | 91.96 |
| Shape space derived | 98.18 | 97.72 | 98.03 | 96.81 |

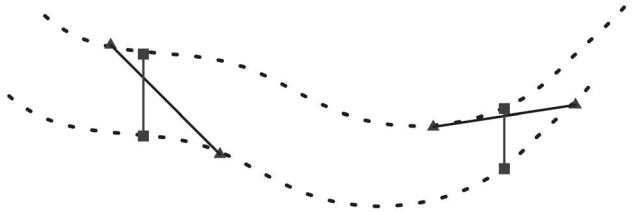

**Fig. 7** Regular resampling for ICP is beneficial. Registered surfaces are shown as dotted lines. Surface points are depicted with triangles before resampling and with squares after resampling.

alignment), as we have different number of 3-D points per subject. It is possible to apply PCA on the range image without 3-D registration, by using the landmarks to define a bounding box on the range image. In Ref. 34, this approach is used, where manual landmarking is followed by rescaling of the face area to a fixed bounding box. By using a cropping mask, the effect of boundaries were reduced, and a PCA transformation was used to compute Mahalanobis cosine distance for classification. With an experimental setup on the FRGC version 1 dataset very similar to ours (198 training and 670 probe images were used, subjects with a single image are excluded), the reported accuracy is 87.04 percent.

The ICP results reported in Table 4 are much better than the results reported in Table 3 and the results reported in the literature. With the generic AFM and the automatic landmarks (i.e., the fully automatic system) the PSD method without resampling has a rank-1 recognition rate of 87.86 percent, whereas after resampling, it has 98.03 percent accuracy. The reason is depicted graphically in Fig. 7. The facial surface (shown symbolically as a dotted line in the figure) is irregularly sampled by the laser scanner (two points per facial surface, shown as black triangles). The ICP registration brings these surfaces into alignment by global rigid matching. Hence, the corresponding points may not be in close alignment locally, although the sum of all displacement vectors is at a local minimum.

Regular resampling produces depth values at regular $x$ and $y$ intervals (shown as black squares). These points give a more realistic indication of the distance between the two surfaces, unless the absolute depth gradient is very high. In the latter case, small displacements in the $x$-$y$ plane will result in big changes in depth, making an irregular, point-to-point 3-D comparison the logical choice. However, the facial surface as represented by a range image has few points with sharp depth changes (i.e., the nose ridge, mouth and eye corners, and face boundary). Our cropping procedure eliminates the face boundary, and greatly reduces the number of these points. Consequently, the regular resampling is indispensable for AFM-based registration.

When we inspect the samples that are classified correctly after resampling, but not before, we see that the error due to point irregularities is large enough to disturb classification. Figure 8 depicts the mean distance differences between the correct class and the incorrect class for these samples. For samples with irregular point distributions, the point-to-point distance terms have a large variance, and this error is distributed all over the face, as shown with light color. However, for regular resampled point distributions, large areas on the face have very low error, as shown with dark color. As we predicted, the error in resampled faces peaks for locations with greater depth gradient, and especially for the nose ridge. Since the nose ridge is a relatively small area of the face, increased error here is compensated by decreased error on the larger facial surfaces. Further-

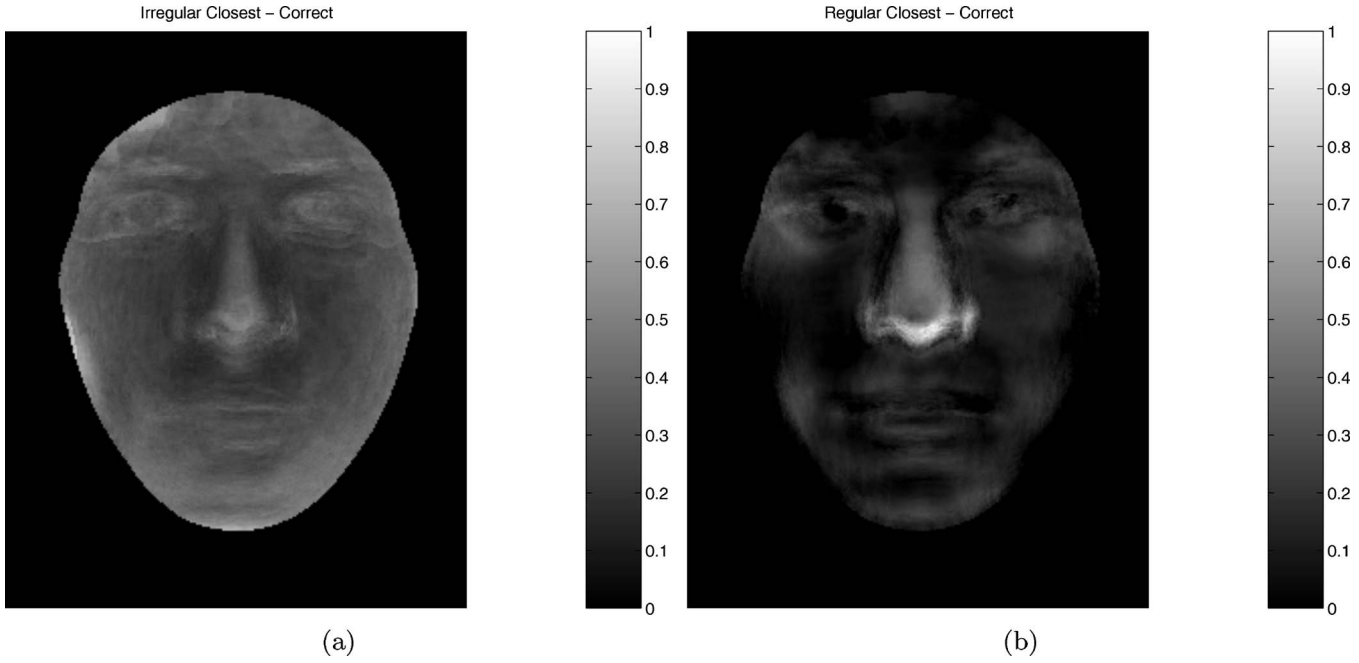


**Fig. 8** Mean point-to-point distance differences in classification. The distances to the correct sample are subtracted from the distances to the closest sample. (a) Irregularly sampled points create errors uniformly on the face surface. (b) Regular sampling reduces the errors on the inner face, close to the nose, where the registration is most accurate.

more, the nose ridge error is increased for competing classes as well. Consequently, the nose area becomes useful in discrimination, even though sometimes it is the highest-error area during registration with the correct gallery sample.

The resampling does not have a high computational cost, as the points are already ordered in the range image. For one-to-all ICP, it is possible to perform a similar resampling. However, if the gallery faces are not in alignment, the resampling has to be performed online for each gallery face separately. Another benefit of using the AFM is that the resampling is performed just once for each test face, and the computation is offline for gallery samples.

When we experiment with race and gender information, we make sure that for each comparison, the training and test samples are registered to the same AFM. For example, in the simulations where the gender is available, we register a male test face with the male AFM before comparing it to the male faces in the gallery, but we use the female AFM for comparisons with female gallery faces. Thus, we have two options when using categorical AFMs: We can either inject ground-truth information (for instance, by setting intragroup distances to infinity), or we can let the system decide on the face category by picking the match with the smallest distance, like we do for shape space clustering–based categories. The results reported in Table 4 are obtained with the latter method. Our simulations show that injecting the ground-truth increases the accuracy only by 0.5 to 1 percent. This means that for this dataset, cross-gender and cross-morphology errors are relatively rare, and we obtain categorical information with great accuracy by simply selecting the best gallery face.

Another issue we have considered was the number of points in different AFMs. The AFM for the females contains roughly 20 percent fewer points that the AFM for the males. In our first experiments, we have used a single mask to crop faces in all categories. This procedure gives faces with equal numbers of points. The results of Table 4 are obtained by allowing each category to have a different number of points after cropping, and by normalizing the Euclidean distances accordingly. For the gender case, this procedure increased the recognition accuracy by a half percent.

## 6 Conclusions

We have evaluated ICP- and TPS-based registration of 3-D faces under automatic and manual landmarking. For real-time 3-D face recognition, the computational requirements of the algorithms must be taken into consideration. The much slower ICP method is viable only if the registration is speeded up through the use of average face models. We have proposed an AFM-based method for this purpose and demonstrated its usefulness. By extending the paradigm to multiple AFMs, we have proposed two approaches for generating category-specific AFMs and contrasted these. Last, we have proposed a regular resampling step that increases the accuracy and speed of classification.

Our results show that ICP is superior to the faster TPS-based method in accuracy. Simulations with improved non-rigid approaches have shown that these methods depend heavily on the accuracy of landmarks and also require a greater number of those.[19] However, the beneficial effect from specific AFM use is more evident in TPS methods that use either automatic or manual landmarks.

We propose a cognitively based approach and a clustering approach to generating specific AFMs. Clustering on shape space produces good AFMs, increases the accuracy of registration, but also reveals natural groups depending on morphology and gender in the face space. This is interesting, as it provides an indication that the other-race effect has a physical basis even for the 3-D information contained in a face.

The reasoning behind multiple AFM usage is that categorical information can act as a filter to reduce the candidates for recognition, and an average face more similar to the test face can ensure a better registration. In the case that cross-gender and cross-morphology confusions are relatively rare, the injection of categorical information does not increase the accuracy.

Another observation regarding multiple AFMs is that specific AFM models have different numbers of points. A male face usually contains 20 percent more points than a female face, simply because female human faces are typically smaller. When we align a face to the female and the male AFMs, the distribution of distances is different in the center of the face and at the periphery. Using a smaller AFM (the one for the females, or Asians, for instance) will effectively remove the points close to the periphery from the distance calculation. This can be an issue for one-to-all ICP approaches as well.

Our experimental results have also confirmed that ICP is sensitive to initialization and that automatic landmarking as a preprocessing step is beneficial to ICP. Manual inspection showed that none of the test cases had gross registration deficiencies in the fully automatic method. The nose-tip heuristic may be useful in frontal faces, but the hair, clothing, and sometimes the chin can be erroneously detected as the nose tip. The error due to incorrect nose localization can be gauged by looking at the results of the simulations that use the ground-truth for the nose in initialization. We should also keep in mind that the database we use is made up of near-frontal faces. The nose-tip heuristic will perform worse in other pose settings. On the other hand, our results confirm that the nose tip is more important for the registration than any other landmark we have used.

The resampling procedure we propose is based on the relative flatness of the facial surface and the fact that cropping eliminates the facial boundary, where the cheeks can produce a sharp depth gradient in the frontal view. With this method, the error due to irregular sampling by the laser scanner is compensated, and there is a significant increase in the accuracy. As a result of AFM-based registration, the cropped models have the same number of points, and these points are matched on the depth map after resampling. This helps us in projecting the depth map to a much lower dimensional manifold via PCA, while retaining a high recognition accuracy. We obtain the best results with ICP and shape space clustered AFMs, and our reported results with resampling are significantly better than results reported in the literature without resampling for the same experimental protocol.

## Acknowledgments

This work is supported by FP6-NoE BIOSECURE, Boğaziçi University Project BAP-03S106, and TÜBİTAK Project 104E080. We thank our anonymous reviewers for their contributions.

## References

1. A. A. Salah, N. Alyüz, and L. Akarun, "Alternative face models for 3D face registration," *Proc. SPIE* **6499**, 64990E (2007).
2. A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Expression-invariant 3D face recognition," in *Audio- and Video-Based Person Authentication*, J. Kittler and M. S. Nixon, Eds., pp. 62–70, Guildford, UK (2003).
3. A. Srivastava, X. Liu, and C. Hesher, "Face recognition using optimal linear components of range images," *Image Vis. Comput.* **24**, 291–299 (2006).
4. S. Lao, Y. Sumi, M. Kawade, and F. Tomita, "3D template matching for pose invariant face recognition using 3D facial model built with iso-luminance line based stereo vision," in *Proc. Intl. Conf. Patt. Recog.*, Vol. 2, pp. 911–916 (2000).
5. B. Achermann and H. Bunke, "Classifying range images of human faces with Hausdorff distance," in *Proc. Intl. Conf. Patt. Recog.*, pp. 809–813 (2000).
6. P. Besl and N. McKay, "A method for registration of 3-D shapes," *IEEE Trans. Pattern Anal. Mach. Intell.* **14**(2), 239–256 (1992).
7. C. S. Chua, F. Han, and Y. K. Ho, "3D human face recognition using point signature," in *Proc. Intl. Conf. Automatic Face and Gesture Recognition*, Vol. 1, pp. 233–238 (2000).
8. M. Hüsken, M. Brauckmann, S. Gehlen, and C. von der Malsburg, "Strategies and benefits of fusion of 2D and 3D face recognition," *Proc. IEEE Intl. Conf. on Computer Vision and Pattern Recognition*, Workshops, p. 174 (2005).
9. X. Lu, A. K. Jain, and D. Colbry, "Matching 2.5D face scans to 3D models," *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(1) 31–43 (2006).
10. M. O. Irfanoğlu, B. Gökberk, and L. Akarun, "3D shape-based face recognition using automatically registered facial surfaces," *Proc. Intl. Conf. Patt. Recog.*, Vol. 4, 183–186 (2004).
11. T. Faltemier, K. Bowyer, and P. Flynn, "3D face recognition with region committee voting," in *Proc. 3rd Int. Symp. 3D Data Processing, Visualization, and Transmission*, pp. 318–325 (2006).
12. D. F. Watson, *Contouring: A Guide to the Analysis and Display of Spacial Data*, Pergamon Press (1994).
13. C. Goodall, "Procrustes methods in the statistical analysis of shape," *J. R. Stat. Soc. Ser. B (Methodol.)* **53**(2), 285–339 (1991).
14. J. C. Gower, "Generalized Procrustes analysis," *Psychometrika* **40**(1), 33–51 (1975).
15. D. Colbry, G. Stockman, and A. K. Jain, "Detection of anchor points for 3D face verification," *Proc. IEEE Intl. Conf. on Computer Vision and Pattern Recognition*, Workshops, p. 118 (2005).
16. F. L. Bookstein, *Morphometric Tools for Landmark Data: Geometry and Biology*, Cambridge University Press (1991).
17. A. A. Salah, H. Çinar Akakin, L. Akarun, and B. Sankur, "Robust facial landmarking for registration," *Ann. Telecommun.* **62**(1–2), 1608–1633 (2007).
18. A. A. Salah and L. Akarun, "3D facial feature localization for registration," in *Int. Workshop Multimedia Content Representation, Classification and Security*, B. Gunsel *et al.*, Eds., Vol. 4105/2006, pp. 338–345, New York (2006).
19. A. A. Salah, , "Biologically motivated 3D face recognition," PhD dissertation, Dept. of Computer Engineering, Boğaziçi University (2007).
20. A. A. Salah and E. Alpaydin, "Incremental mixtures of factor analyzers," *Proc. Intl. Conf. Patt. Recog.*, Vol. 1, pp. 276–279 (2004).
21. I. Gauthier and M. J. Tarr, "Becoming a 'Greeble expert': exploring the face recognition mechanism," *Vision Res.* **37**, 1673–1682 (1997).
22. M. H. Tong, C. A. Joyce, and G. W. Cottrell, "Are Greebles special? Or, why the fusiform fish area would be recruited for sword expertise (if we had one)," in *Proc. 27th Annual Cognitive Science Conference*, Lawrence Erlbaun, Mahwah, NJ (2005).
23. T. Valentine, "A unified account of the effects of distinctiveness, inversion and race in face recognition," *Q. J. Exp. Psychol.* **43A**, 161–204 (1991).
24. P. J. Phillips, P. J. Flynn, W. T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. J. Worck, "Overview of the face recognition grand challenge," *Proc. IEEE* **I**, 947–954 (2005).
25. F. Tsalakanidou, S. Malassiotis, and M. Strinzis, "Integration of 2D and 3D images for enhanced face authentication," *Proc. Intl. Conf. Automatic Face and Gesture Recognition*, pp. 266–271 (2004).
26. K. Chang, K. Bowyer, and P. Flynn, "Multi-modal 2D and 3D biometrics for face recognition," in *Proc. IEEE Int. Workshop on Analysis and Modeling of Faces and Gestures*, Vol. 1, pp. 187–194 (2003).
27. V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(9), 1063–1074 (2003).
28. Y. Wang, C. Chua, and Y. Ho, "Facial feature detection and face recognition from 2D and 3D images," *Pattern Recogn. Lett.* **23**, 1191–1202 (2002).
29. C. Boehnen and T. Russ, "A fast multi-modal approach to facial feature detection," in *Proc. 7th IEEE Workshop on Applications of Computer Vision*, pp. 135–142 (2005).
30. K. Chang, K. Bowyer, and P. Flynn, "Multiple nose region matching for 3D face recognition under varying facial expression," *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(10), 1695–1700 (2006).
31. C. Xu, T. Tan, Y. Wang, and L. Quan, "Combining local features for robust nose location in 3D facial data," *Pattern Recogn. Lett.* **27**(13), 1487–1494 (2006).
32. A. Lu and A. Jain, "Automatic feature extraction for multiview 3D face recognition," in *Proc. 7th IEEE Int. Conf. AFGR*, pp. 585–590 (2006).
33. B. Gökberk, H. Dutağaci, A. Ulaş, L. Akarun, and B. Sankur, "Representation plurality and fusion for 3–D face recognition," *IEEE Trans. Syst., Man, Cybern., Part B: Cybern.* in press.
34. K. I. Chang, K. W. Bowyer, and P. J. Flynn, "An evaluation of multimodal 2D+3D face biometrics," *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(4), 619–624 (2005).

**Albert Ali Salah** worked as a research assistant in the Perceptual Intelligence Laboratory of Boğaziçi University, where he was part of a team working on machine learning, face recognition, and human-computer interaction. After receiving his PhD in computer engineering under the supervision of Prof. Lale Akarun in 2007, he joined the Signals and Images research group at Centrum voor Wiskunde en Informatica (CWI) in Amsterdam. With his work on facial feature localization, he received the inaugural EBF European Biometrics Research Award in 2006.

**Neşe Alyüz** received her BS degree in computer engineering from Istanbul Technical University in 2005 and her MS degree from Boğaziçi University in 2008. She is currently a graduate student at the Perceptual Intelligence Laboratory of Boğaziçi University, working on expression invariant face recognition. Her research areas are three-dimensional face recognition, face modeling, and expression recognition.

**Lale Akarun** received BS and MS degrees in electrical engineering from Boğaziçi University, Istanbul, in 1984 and 1986, respectively. She obtained her PhD from Polytechnic University, New York, in 1992. Since 1993, she has been working as a faculty member at Boğaziçi University. She became a professor of computer engineering in 2001. Her research areas are face recognition, modeling, and animation of human activity, and gesture analysis. She has worked on the organization committees of IEEE NSIP99, EUSIPCO 2005, and eNTERFACE 2007. She is a senior member of the IEEE.