3-D Face Recognition Under Occlusion Using Masked Projection

Nese Alyuz, Berk Gokberk, and Lale Akarun, *Senior Member, IEEE*

***Abstract*—With advances in sensor technology, the three-dimensional (3-D) face has become an emerging biometric modality, preferred especially in high security applications. However, dealing with occlusions covering the facial surface is a great challenge, which should be handled to enable applicability to fully automatic security systems. In this paper, we propose a fully automatic 3-D face recognition system which is robust to occlusions. We basically consider two problems: 1) occlusion handling for surface registration, and 2) missing data handling for classification based on subspace analysis techniques. For the alignment problem, we employ an *adaptively-selected-model*-based registration scheme, where a face model is selected for an occluded face such that only the valid nonoccluded patches are utilized. After registering to the model, occlusions are detected and removed. In the classification stage, a masking strategy, which we call *masked projection*, is proposed to enable the use of subspace analysis techniques with incomplete data. Furthermore, a regional scheme suitable for occlusion handling is incorporated in classification to improve the overall results. Experimental results on two databases with realistic facial occlusions, namely, the Bosphorus and the UMB-DB, are reported. Experimental results confirm that registration based on the adaptively selected model together with the masked subspace analysis classification offer an occlusion robust face recognition system.**

***Index Terms*—3-D face recognition, 3-D registration, biometrics, curvature descriptors.**

I. Introduction

IN BIOMETRIC systems, human beings are identified by distinctive features, such as physiological and behavioral characteristics. As a biometric modality, the human face is widely preferred because of several advantages: Due to its contactless acquisition, it is well accepted among users. Furthermore, its applicability to noncooperative scenarios makes it suitable for a range of applications such as surveillance systems. However, in noncooperative and uncontrolled scenarios, recognizing individuals from their faces is a challenging task. The factors that degrade the performance of a face recognizer, include presence of illumination differences, in-depth pose variations, facial expression variations, and the presence of occlusions. In the three dimensional (3-D) domain, challenges caused by illumination, pose, and expression variations can be better handled. However, extreme occlusion variations still complicate the task of identification. In this work, we propose a 3-D face recognition system that is robust under realistic occlusions.

In our approach, occlusion handling is considered in the registration and the classification stages. For alignment of occluded surfaces, we employ a registration scheme which adaptively selects an alignment model, where for each probe face, a model including the probable nonoccluded facial parts is employed. By adaptively selecting a model, it is possible to discard the effect of occluding surfaces on registration. After alignment, occluded regions are discarded. Occlusion-removed facial surfaces contain missing data points. In this paper, we propose a *masked projection* technique that can cope with missing data. Furthermore, we utilize a regional approach to improve the classification performance, where different regions serve as separate classifiers.

A. Related Work

Developments in 3-D sensor technologies have increased interest in 3-D face recognition. In [31], it is shown that by using 3-D face, it is possible to obtain competitive results when compared with other modalities such as iris and high-resolution 2-D facial images. A thorough survey of previously proposed 3-D face recognition systems can be found in [1], [9], [35] and detailed fundamental analysis of concepts are given in [2], [21], [28]. In this section, we focus on the recent face recognition approaches dealing with realistic occlusion variations, both in 2-D and 3-D.

1) 2-D Techniques: Although variations caused by pose and expression have attracted increased research effort, the problem of handling occlusions has not been discussed frequently. In the 2-D face recognition studies, there has been a few approaches considering occlusion variations. In most of these studies, the aim is occlusion handling for recognition and the registration problem is not considered: Experimental results are usually reported on databases where the faces are assumed to be accurately registered prior to recognition.

Some studies are based on subspace analysis methods, where the aim is either occlusion-robust projection or missing data compensation. In [30], Park *et al.* consider occlusions caused only by eyeglasses and propose a method to compensate for the missing data. Initially, the glasses region is extracted using color and edge information. The offline-generated eigenfaces from a set of nonoccluded images are then used together with the extracted glasses region for missing data compensation. In

Manuscript received August 20, 2012; revised November 26, 2012 and November 26, 2012; accepted March 15, 2013. Date of publication April 01, 2013; date of current version April 16, 2013. This work was supported by TUBITAK Grant 108E161. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Zhenan Sun.

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

B. Gokberk is with the Department of Electrical Engineering Mathematics and Computer Science, University of Twente, 7500 AE Enschede, 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.2013.2256130

1556-6013/$31.00 © 2013 IEEE

[36], occlusion variations are handled by eliminating facial parts where occlusions frequently occur. Several subsets of images are created through masking facial regions both in training and test faces. Using masked training images, different face projection spaces are created through Principal Component Analysis (PCA) and majority voting is applied to fuse multiple classifiers. In [18], an approach for combining discriminative and reconstructive methods is proposed for better handling of images with outlier pixels. The general discriminative model is rewritten by incorporating the feature vectors corresponding to the reconstructive model. In addition, the truncated projection matrix is extended to retain the complete discrimination power.

Other holistic approaches can be considered as model-based methods. In [16], De Smet *et al.* proposed an iterative approach for the parameter estimation of 3-D morphable model fitting procedure. Concurrently, a visibility map defining the occlusions is modeled by Markov Random Fields (MRF), which accounts for spatial coherence of occlusions. The visibility map is used to exclude occluded regions from further computations. Similar to the morphable model formulation, Park *et al.* [29] proposed to encode all the geometric quantities and the structural information residing in a facial surface as an Attributed Relational Graph. Identification is achieved by partial matching of these graphs. In [24], Lin and Tang proposed a method which encapsulates the occlusion detection and recovery problems through a generative process. A Bayesian formulation is proposed, where the quality assessment model is constructed by learning *a priori* information from a set of images.

Another approach for occlusion handling considers the facial surface as a combination of partitions. When local patches are considered separately, the areas where occlusions occur can be compensated for, in the classifier fusion phase. In [27], the facial surface is divided into local regions. Each region is modeled individually by a mixture of Gaussian distributions, and fusion is achieved by probabilistic evaluation of regional matches. In [23], Kim *et al.* propose a part-based local representation approach based on Independent Component Analysis (ICA). ICA representations are constructed for local regions corresponding to salient parts such as eye, nose, and lip areas. Conservation of discriminative features is achieved by reordering of basis images. In [38], a face image is represented by applying multiscale and multiorientation Gabor filters and obtaining the Local Binary Pattern (LBP) map. Recognition is achieved by matching regional histograms.

Recently, there has been increasing interest in the area of sparse representation techniques. For robust face recognition against occlusions and corruptions, Wright *et al.* [37] proposed an identification technique, where the occlusion robustness is obtained by sparsely representing corrupted pixels. Additionally, identification performance for occluded facial images is improved by block partitioning. In [22], a sparse representation technique based on correntropy is proposed for occlusion handling. Nonnegativity constraint is introduced to obtain a more sparse and efficient solution. In [39], Zhou *et al.* proposed to improve sparse representation methods for handling of contiguous occlusions by including prior knowledge about the pixel-wise error distribution. The spatial continuity of both corrupted and uncorrupted pixels are modeled by Markov Random Fields. In these approaches, although sparse representation appears beneficial for occluded surfaces, best results are obtained when occlusions are manually removed or compensated for via block partitioning.

2) 3-D Techniques: In the studies using 3-D facial data, only a few consider facial occlusion detection, removal, restoration, and missing data handling [3], [11]–[14], [40]. In [11], Colombo *et al.* proposed to detect occlusions by analyzing the difference between an original face and its approximation by the eigenface approach. The regions detected as occlusions are removed and the locations of the missing parts are employed in the restoration process, which is handled by Gappy PCA [17]. Identification was performed by the Fisherfaces approach. In [12], they have refined the occlusion detection method by including the difference of the input image from the mean face. In the experiments, it was reported that restoration does not offer improvement for faces occluded by more than 30%. In [13], the authors have employed the occlusion detection and restoration idea to the face detection problem. Occlusions are detected roughly by thresholding the difference from a mean face, and Gappy PCA is utilized to discriminate between face and nonface images. In [14], they have integrated their previous occlusion detection approach [12] into the face recognition pipeline of [11] and reported experimental results both on synthetically occluded faces and on the Bosphorus database [34] which includes realistic occlusion variations.

In [3], a part-based method is proposed, where facial regions are aligned independently to average regional models. The regional division scheme is also employed in the classification stage, where the regional classification results are fused with different fusion techniques. The experimental results indicate performance improvement by the part-based system, both for expression and occlusion variations. In [6], the authors have proposed a nose-based registration scheme for better handling of occluded faces. Curvature information is utilized for automatic detection of the nose area, and an average nose model is used for fine alignment via Iterative Closest Point (ICP) algorithm. On the registered surfaces, occlusions are detected by analyzing the difference from the average face model, and the occlusion-removed surfaces are completed by a modified version of the Gappy PCA method. The restored faces are classified using different local masks, and multiple classifiers are fused for final identity estimation.

B. Proposed System and Contributions

This work introduces a new technique called *masked projection* for subspace analysis with incomplete data. We use the system outlined in Fig. 1: The preprocessing module includes the registration and occlusion removal steps. For alignment, the adaptive registration module[1] of [5] is utilized, which registers the occluded surfaces. By adaptively selecting the model, it is possible to discard the effect of occluding surfaces on registration. The occlusions are detected on the registered surfaces by thresholding point distances to an average face model. The training module works offline to learn the projection matrices

[1]Our adaptive registration approach was previously presented in a conference paper [5] with preliminary experiments on the UMB-DB database.

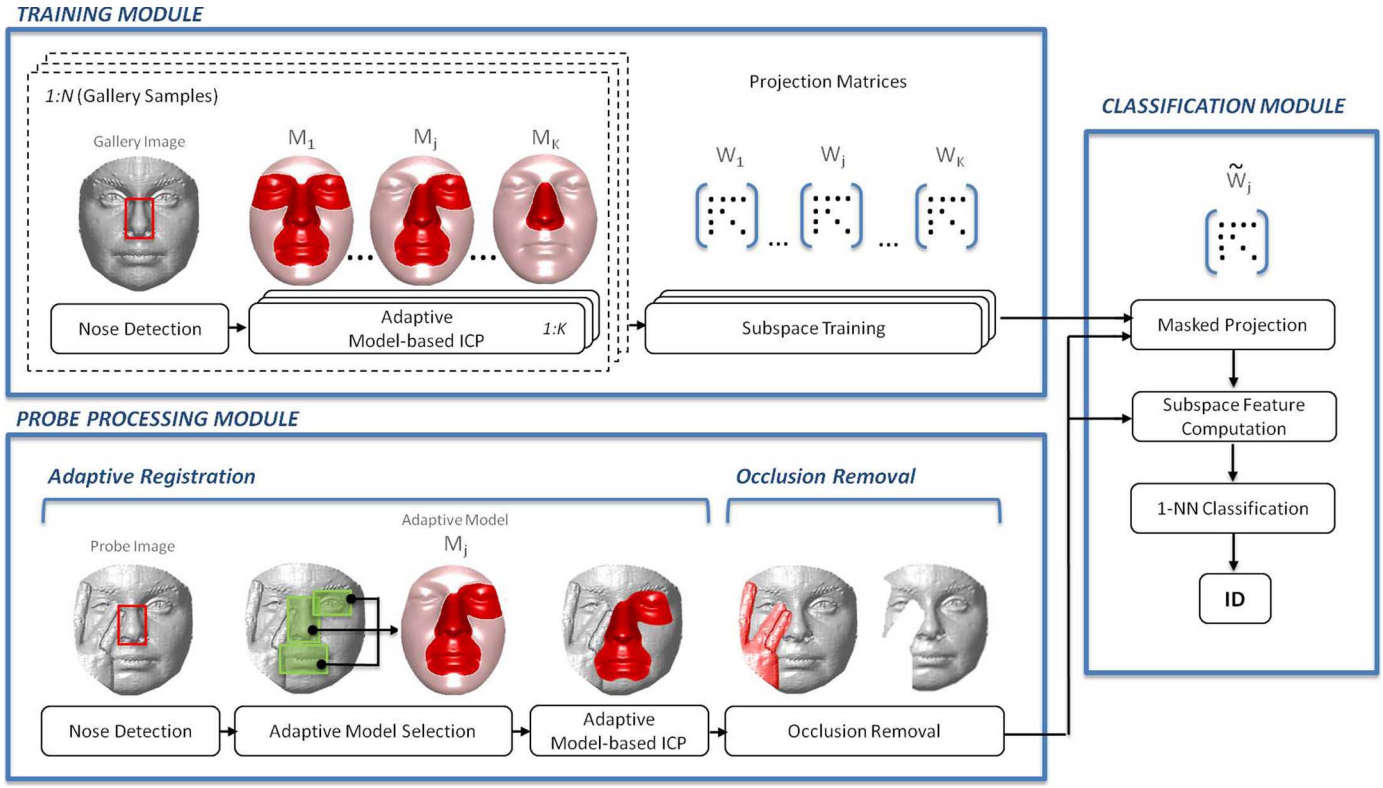


Fig. 1. Illustrative diagram of the proposed 3-D face recognition approach.

from the training set of nonoccluded faces for different regions. The classification module uses the occlusion mask of the probe image to compute the masked projection, and projects the probe image to the adaptive subspace. The identification is handled in the subspace by 1-nearest neighbor (1-NN) classifier. The proposed system is evaluated on two main 3-D face databases that contain realistic occlusions: (1) The Bosphorus, and (2) the UMB-DB databases.

The paper is organized as follows: In Section II, the details of the nose detection, registration and the subsequent occlusion detection phases are given. In Section III, the newly proposed adaptive subspace technique which can handle missing data is explained. The experimental results are reported in Section IV and the conclusions are given in Section V.

II. Registration and Occlusion Removal

In this work, the occlusion resistant registration approach initially proposed in [5] is employed. For the integrity of this work, we have summarized the details of the registration and occlusion detection processes.

A. Automatic Nose Detection

Iterative Closest Point (ICP) algorithm [8] is one of the most widely preferred methods for rigid registration of 3-D surfaces. However, iterative approaches like ICP are highly dependent on the initial conditions. In most of the current 3-D face recognition systems, the surface initialization is handled by accurately locating a set of facial landmark points [10], [26]. However, in the presence of occlusions, localization of the nose area rather than individual points appears to be a more effective detection strategy [6].

The nose detection algorithm [6] is based on the surface curvature information, which is advantageous due to its rotation and translation invariance. Initially, two curvature maps are computed for a given surface, namely the shape index map and the curvedness map. The shape index value $\mathrm{SI}(i)$ at surface point i can be computed from principal curvatures $\kappa_{\max}(i)$ and $\kappa_{\min}(i)$:

$$\mathrm{SI}(i) = \frac{1}{2} - \frac{1}{\pi}\tan^{-1}\frac{\kappa_{\max}(i) + \kappa_{\min}(i)}{\kappa_{\max}(i) - \kappa_{\min}(i)}. \tag{1}$$

As the scale-dependent counterpart of shape index, *curvedness* measures the rate of curvature at each point:

$$\mathrm{C}(i) = \sqrt{\frac{\kappa_{\min}(i)^2 + \kappa_{\max}(i)^2}{2}}. \tag{2}$$

The shape index map **SI** provides a smooth transition between concave $(0 < \mathrm{SI}(i) < 0.5)$ and convex $(0.5 < \mathrm{SI}(i) < 1)$ shapes. Since nose is a convex structure, thresholding is applied on the **SI** map (by 0.5) to eliminate concave regions. After the concave regions are eliminated, **SI** is weighted with curvedness to integrate scale-dependent and scale-independent components [25]:

$$\mathrm{WSI}(i) = \mathrm{SI}(i) * \mathrm{C}(i). \tag{3}$$

Here, **WSI** denotes the curvedness-weighted convex shape index. On this final map, the nose area is localized by employing template matching. For the construction of the nose template, an average face model is created by Thin Plate Spline warping,

where a set of registered nonoccluded neutral training images are used [33]. Then, the average nose model is obtained by manually selecting the nose region on the average face model, and the **WSI** map for the nose model is constructed to serve as the nose template. Given a test image, template matching is performed by normalized cross-correlation, where the region mostly resembling the nose structure is located.

B. Adaptive Model-Based Registration

To decrease the overall computational cost of ICP, an average face model is employed as a common registration reference [20]. However, when occlusions are present, the extraneous objects alter the surface. The altered information causes the correspondence matching and the registration steps to fail. Here, we employ a model-based registration approach which can cope with occlusion variations: For an occluded face, the nonoccluded patches are determined and a model with the corresponding average patches is selected. Therefore, only the nonoccluded facial points are considered in correspondence matching.

For the detection of distinctive facial features, such as eyes and mouth, we employ the relative geometry information of these features: On the average face model, the displacement vectors between the nose and the other fiducial patch centers are computed. After nose detection on a probe face (Section II-A), the probable patch centers are computed using the displacement vectors. Around the probable patch centers, the patches are sought in the predefined bounding boxes via template matching. The main patches we considered are the eyes and mouth; and we have an additional set of half patches: left-right nose halves, upper-lower nose halves, and left-right eyes. For each patch, the average model and template formation procedures are followed: The average patches are constructed by manually cropping the average face model, and the **WSI** map is computed to define the patch template. Due to occlusions over the facial surface, some patches will not be visible and cannot be located correctly by template matching. To determine the validity of each patch, thresholding is applied on template matching scores. The thresholds used for patch validity are calculated from the scores of a separate nonoccluded and neutral database. The probe patches that have dissimilarity scores below the threshold define the valid parts. The patch validity values are used for the model selection: The respective valid patches are selected from the average face model, and these average patches constitute the adaptive alignment model. The selected model is then used when registering the probe face.

For the alignment of facial surfaces, the coarse initialization is obtained by translating the center of the detected nose region to the nose area center of the average face. This translation provides a sufficient initialization for nearly frontal faces, where small pose deflections are corrected in the ICP-based fine registration. For fine registration, ICP is utilized where the adaptive model, selected specific to the probe, serves as the reference. Since the model employed in the fine alignment step is chosen according to the occlusion, the ICP algorithm automatically selects the respective nonoccluded parts as corresponding surface point sets and estimates the alignment parameters using only the nonoccluded regions. Hence, the overall registration approach becomes robust to occlusions.

C. Occlusion Detection Based on Difference From Average Face Model

After the facial surfaces have been registered, it is important to locate facial areas occluded by exterior objects. The most straightforward approach for occlusion detection is to analyze the difference of the input image from a mean face, as in [13]: If there is an exterior object covering a part of the facial surface, the difference for this specific area will be more evident. Therefore, occluded areas can be detected by thresholding the difference map obtained by computing the absolute difference between the average face and the input face. Prior to occlusion detection, the facial surfaces are resampled from a regular grid to construct the depth images. Regular resampling enables only the z coordinates, namely the depth values, to be considered for a sufficient comparison. If the depth images for the input face and the average face model are denoted by $\boldsymbol{x}$ and $\boldsymbol{x}_{avg}$, respectively, the occlusion mask is obtained by thresholding the absolute difference:

$$m(j) = \begin{cases} 1 & \text{if } |x(j) - x_{avg}(j)| > t \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where t denotes the predefined threshold value. On this initial mask, $\boldsymbol{m}$, postprocessing operations, namely morphological dilation and connected component analysis, are applied to obtain the final occlusion mask.

III. Masked Projection to Adaptive Subspaces

A. Global Classification Using Masked Projection

A useful property of the model-based registration scheme is that the extracted facial features, $\boldsymbol{x}_i$, are ordered vectors of the same size, enabling the use of subspace analysis techniques. However, subspace approaches assume complete facial feature vectors. Therefore standard subspace approaches cannot be applied directly on occlusion-free faces. The first idea to deal with incomplete data, would be to remove the pixels that are not present in the probe image from all of the training and gallery images, as in [36]. Using the masked training images, the subspace representing the partial surfaces can be learned by the Fisherfaces approach [7]. However, this approach is not feasible, since each probe face will have different pixels missing and a separate training phase is required. In this work, we propose a projection masking approach to obtain the adaptive subspace: The general projection matrix is learned using a set of nonoccluded complete training images. Then, the adaptive projection matrix is obtained by masking. The masked probe and gallery images are projected onto the subspace, and classification is performed. The algebraic details of the approach are given below.

Let $\boldsymbol{x}$ be the registered facial surface vector, and $\mathbf{W}$ represent a projection matrix[2]. The surface vector can be defined as $\boldsymbol{x} = \boldsymbol{\mu} + \mathbf{W}\boldsymbol{y}$, where $\boldsymbol{\mu}$ is the mean of the training images, and $\boldsymbol{y}$ is the coefficient vector residing in the subspace defined by $\mathbf{W}$. To simplify equations, we assume that $\boldsymbol{\mu}$ is zero. In practice,

[2]In our experiments, we have used the Fisherfaces projection.

this is assured by a change of coordinates. The coefficients are computed as

$$\boldsymbol{y} = \mathbf{W}'\boldsymbol{x}, \tag{5}$$

where $\mathbf{W}'$ is the transpose of $\mathbf{W}$. Now, suppose there is an incomplete version of $\boldsymbol{x}$, namely $\hat{\boldsymbol{x}}$, whose missing components are encoded in the occlusion mask $\mathbf{m}$. Let's assume that we have the coefficient vector $\tilde{\boldsymbol{y}}$, where the input image can be approximated as

$$\tilde{\boldsymbol{x}} = \mathbf{W}\tilde{\boldsymbol{y}} \tag{6}$$

where $\tilde{\boldsymbol{x}}$ is the approximated complete version of the input image $\hat{\boldsymbol{x}}$. Our objective is to find the coefficient vector $\tilde{\boldsymbol{y}}$, minimizing the error term $E = \|\hat{\boldsymbol{x}} - \tilde{\boldsymbol{x}}\|^2$. In this formulation, the missing components in $\hat{\boldsymbol{x}}$ will augment the total error term. To improve the error term, the masked norm [17] is used[3], where the information about the missing components is encoded in the mask $\mathbf{m}$. The masked norm for a vector $\mathbf{u}$ with the mask $\mathbf{m}$ is defined as $\|\mathbf{u}\|_m = \sqrt{(\mathbf{u},\mathbf{u})_m}$ where

$$(\mathbf{u},\mathbf{u})_m = \mathbf{u}'_m\mathbf{u}_m. \tag{7}$$

Here $\mathbf{u}_m$ is the masked version of $\mathbf{u}$, $\mathbf{u}_m = \mathbf{\Lambda}_m\mathbf{u}$, where $\mathbf{\Lambda}_m$ is a diagonal matrix, whose diagonal elements constitute the mask: $\mathbf{m} = diag(\mathbf{\Lambda}_m)$. When the masked data is inserted into E, we obtain:

$$\begin{aligned} E_m &= \|\hat{\boldsymbol{x}} - \tilde{\boldsymbol{x}}\|_m^2 \\ &= \hat{\boldsymbol{x}}'_m\hat{\boldsymbol{x}}_m - \hat{\boldsymbol{x}}'_m\mathbf{W}_m\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{y}}'\mathbf{W}'_m\hat{\boldsymbol{x}}_m \\ &\quad + \tilde{\boldsymbol{y}}'\mathbf{W}'_m\mathbf{W}_m\tilde{\boldsymbol{y}} \end{aligned} \tag{8}$$

where $\mathbf{W}_m = \mathbf{\Lambda}_m\mathbf{W}$. The error is minimized with respect to $\tilde{\boldsymbol{y}}$:

$$\frac{\partial E_m}{\partial \tilde{\boldsymbol{y}}} = -2\mathbf{W}'_m\hat{\boldsymbol{x}}_m + 2\mathbf{W}'_m\mathbf{W}_m\tilde{\boldsymbol{y}} = 0. \tag{9}$$

$$\hat{\boldsymbol{x}}_m = \mathbf{W}_m\tilde{\boldsymbol{y}}. \tag{10}$$

To calculate the coefficients $\tilde{\boldsymbol{y}}$, the inverse of $\mathbf{W}_m$ is needed. Since $\mathbf{W}_m$ is constructed from $\mathbf{W}$ by setting occluded regions to zero, $\mathbf{W}_m$ is no longer orthogonal. Since inverse of an orthogonal matrix is just the transpose it, for ease of calculations, we first orthogonalize the $\mathbf{W}_m$ matrix [18]:

$$\mathbf{W}_\perp = \mathbf{W}_m(\mathbf{W}'_m\mathbf{W}_m)^{-1/2}. \tag{11}$$

Then, the coefficient vector can be computed as:

$$\tilde{\boldsymbol{y}} = \mathbf{W}'_\perp\hat{\boldsymbol{x}}_m. \tag{12}$$

The masked projection matrix can be applied to the masked gallery image matrix $\boldsymbol{X}$, whose columns correspond to observations:

$$\boldsymbol{Y} = \mathbf{W}'_\perp\boldsymbol{X}. \tag{13}$$

Here, it should be noted that gallery vectors should also be projected using the masked projection matrix $\mathbf{W}_\perp$, rather than the original projection matrix $\mathbf{W}$, since these two matrices define different subspaces: The subspace of $\mathbf{W}$ is trained using a subset of complete facial surfaces. When $\mathbf{W}_m$ is constructed, parts corresponding to occlusions are eliminated from the original matrix. Therefore, the orthogonal vector sets defining the subspaces are different for $\mathbf{W}$ and $\mathbf{W}_m$ (hence for $\mathbf{W}$ and $\mathbf{W}_\perp$). The idea of projecting the gallery images with masked projection, in addition to the probe images, is the main difference between the proposed approach and the Gappy PCA method of [17].

After the projection to the adaptive subspace, the dissimilarity between the probe coefficients $\tilde{\boldsymbol{y}}$ and the coefficients of any gallery image; $\boldsymbol{y}_{G_k}$ which is the kth column of $\boldsymbol{Y}$; can be computed by the angular cosine distance measure:

$$D(\tilde{\boldsymbol{y}}, \boldsymbol{y}_{G_k}) = 1 - \frac{\tilde{\boldsymbol{y}} \cdot \boldsymbol{y}_{G_k}}{\|\tilde{\boldsymbol{y}}\| \cdot \|\boldsymbol{y}_{G_k}\|}. \tag{14}$$

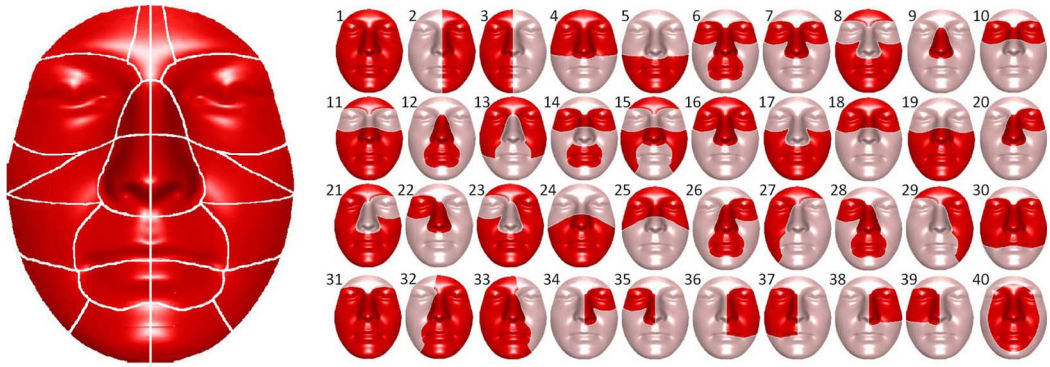
To obtain the final identification rates, the regional dissimilarity measures are fused by the product rule and 1-NN classification is employed.

B. Regional Classification Using Masked Projection

For further improvement in the classification phase, we propose to consider the 3-D surface as a combination of several regions. If the facial area is partially occluded by external objects, the incorrect information regarding the covered regions will cause the global classification approaches to fail. Therefore, in the presence of occlusions, it is beneficial to incorporate separate regional classifiers. In regional techniques, each region acts as an independent classifier, and the regional recognition results are fused to obtain an improved overall performance. For the construction of the regions, we have divided the facial surface into several nonoverlapping patches. Then, combination of these patches are merged to generate facial regions. The proposed regional division scheme consists of 40 regions as illustrated in Fig. 2. In Fig. 2(a), the 24 symmetrical patches defined on the average face model are given. The facial surface is partitioned considering both the semantic structure (eyes, mouth, forehead, cheeks) and the facial symmetry. When the patch sizes and locations are set, the extent of the local regions to be constructed are taken into account. For the determination of patch combinations, possible real life occlusion scenarios are considered. In Fig. 2(b), the regions created using different subsets of patches are visualized (except for the last region, which is obtained by eroding the global face model).

To incorporate regional classifiers with the proposed subspace method, a separate regional subspace should be learned [4]. Therefore, for each alignment model and for each region, a separate projection matrix, $\mathbf{W}$, is trained. Each projection matrix defines a separate subspace for the corresponding region, where the training images are registered with the corresponding alignment model. When a probe face is examined, all of the regional subspaces of the corresponding model are employed: Regional features are computed by regional masked projections, where the occlusion and regional masks are merged to obtain the final masks employed in the projection stage. Then, separate regional subspace features are compared against corresponding feature sets of gallery images, and the regional classification results are fused. Although training of separate subspaces appears

[3]In [17], this measure is referred to as the *gappy* norm.



(a) (b)

Fig. 2. Regional division scheme: (a) patches, (b) regions (in red). The regions in (b) are constructed as combinations of patches of (a) (except for region 40, which is obtained by eroding region 1).

as a time consuming process, it is handled in an offline manner and does not affect the duration of the classification phase.

IV. Experimental Results

A. Databases

Three databases are employed in this work: (1) FRGC v.2, (2) Bosphorus, and (3) UMB-DB. The FRGC v.2 [32] neutral subset, containing a total of 2365 images of 466 subjects, serves as a separate training set for: (i) the construction of the average face & patch models, (ii) the training phase of the Fisherfaces method, and (iii) the determination of threshold values. The Bosphorus [34] and UMB-DB [15] databases, containing occlusion variations, are utilized to evaluate the system performance. The Bosphorus database includes a total of 4666 scans collected from 105 subjects, including expression, pose, and occlusion variations. The database contains a total of 299 neutral scans. The first neutral scan of each subject is used to construct the gallery set (105 scans). The images with occlusion variations, consisting of 381 images, form the probe set. In Fig. 3, the four types of occlusions present in the Bosphorus database are shown: (1) occlusion of the eye area by eyeglasses, (2) occlusion of the eye area by a hand, (3) occlusion of the mouth area by a hand, (4) occlusion caused by hair. The expression and pose variations of the Bosphorus database are considered in the experiments handled in Section IV-G to evaluate the performance of the system under different acquisition scenarios. The UMB-DB database [15] is acquired from a total of 142 subjects, and there are a total of 1473 scans. In the experiments, the experimental protocol of [15] is considered: The gallery set contains first neutral scan of each subject, and the probe set consists of the occlusion subset. The gallery and probe sets contain 142 neutral and 590 occluded scans, respectively. The occlusions in this database are more challenging and they can be caused by hair, eyeglasses, hands, hats, scarves, and other objects. Furthermore, the location and amount of occlusion vary greatly. In Fig. 4, some examples from the UMB-DB are given to show how challenging the occlusions are.

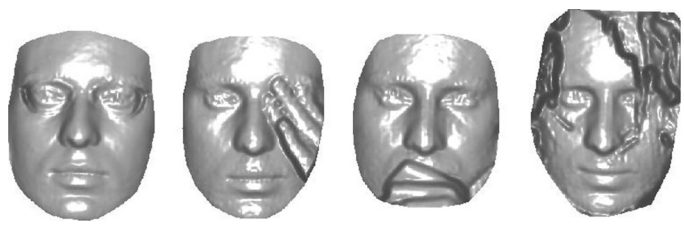


Fig. 3. Four occlusion types in the Bosphorus database: occlusion of the eye area by eyeglasses, occlusion of the eye area by hand, occlusion of the mouth area by hand, and occlusion caused by hair.

Fig. 4. Sample faces from the UMB-DB.

B. Automatic Nose Detection Accuracy

The automatic nose detection results are inspected on all of the databases. First, the performance of the nose detector is evaluated on the whole FRGC v.2 database. When inspected visually, the nose detector shows 100% accuracy both on the neutral subset (2365 scans) and the nonneutral subset (1642 scans with expression variations). For quantitative evaluation of the detection performance on the other databases, the ground truth nose landmarks are employed to estimate the nose area centers. Then, the distances between the estimated and automatically located nose area centers are thresholded. The detections that are distant from the ground truth centers within the predefined threshold value are counted as correct. The threshold value is set empirically on a separate database: Using the whole FRGC v.2 database, the distances between the estimated and automatically located nose area centers are computed. The maximum distance (after trimming the outlier distances) is set as the threshold value (11.5 mm). The automatic nose detection results indicate that the nose areas for nonoccluded scans can be successfully detected: For the Bosphorus neutrals (299 scans) and for the UMB-DB neutrals (441 scans), 100% nose detection accuracies are obtained. The nose detection performance on the occlusion subset

TABLE I
BASELINE DEPTH-BASED IDENTIFICATION PERFORMANCES TO INDICATE THE RELATIVE STANDING OF THE REGISTRATION APPROACHES

Database	Acquisition Type	Probe Count	Recognition Rates (%)		
			Face Model	*Nose Model*	*Adaptive Model*
Bosphorus	Neutral	194	100.00	97.94	100.00
	Occlusion	381	60.63	79.00	83.99
UMB-DB	Neutral	299	98.66	85.28	97.32
	Occlusion	590	47.29	46.27	65.25

Fig. 5. Correct and incorrect nose detections for the UMB-DB database are given in the first and second rows, respectively.

of the Bosphorus database is 98.69%. These results are verified by visually inspecting the detected noses on the respective subsets. For the UMB-DB database, the occlusions often cover the nose area partially. Therefore, manually labeled landmarks are incomplete, preventing a similar quantitative evaluation on this data set. Although the nose area is not fully visible for the UMB-DB occlusion subset, the nose localization performance, obtained by visual inspection, is still quite high: 93.90%. The performance of the nose detector is similar to the face detection performance (93.7%) reported on the UMB-DB database in [15]. When the erroneous detections are inspected, it is seen that the nose area is highly occluded for those scans. In Fig. 5, some correct and incorrect nose detection examples from the UMB-DB are given for challenging occlusions.

C. Patch Validation and Selection Accuracy

After the nose detection phase, the patches of a probe face are estimated and checked for validity and corresponding models are constructed adaptively. The thresholds used for patch validation are determined from the template matching scores of the FRGC v.2 neutral subset: For a specific patch, the scores are sorted and the smallest 10% of them are discarded and the smallest score of the remaining 90% is set as the threshold for that patch. The thresholds are used to set patch validity flags of the Bosphorus and the UMB-DB scans. When the model selection results are analyzed, it is seen that for the Bosphorus occlusions, 54 out of 381, and for the UMB-DB occlusions, 77 out of 590 scans, the model selection is erroneous. Most of the errors for UMB-DB (40 out of 77) are caused by the prior nose detection failures. Note that, even when patch selection is erroneous, faces may be registered well enough to be recognized.

D. Registration Accuracy

To evaluate the registration performance, a baseline recognition experiment is performed using depth information: Using the ground truth occlusion masks, the occluding parts on the registered images are discarded. The occlusion mask is applied both to the probe and to the gallery images. It should be noted that the depth-based identification performances reported here, with manually removed occlusions are provided to indicate the relative standing of the registration approaches.

To formally present the depth-based classifier, let the facial surface be represented by a vector of depth measurements: $\boldsymbol{x} = [z_1, z_2, \ldots, z_d]$, where each surface $\boldsymbol{x}$ contains d valid depth values obtained after regular resampling and occlusion masking. The dissimilarity between any two corresponding facial regions can be computed as:

$$D(\boldsymbol{x}^{(P)}, \boldsymbol{x}^{(G_k)}) = \frac{|\boldsymbol{x}^{(P)} - \boldsymbol{x}^{(G_k)}|}{d} \tag{15}$$

where P is a probe image and G_k is the kth gallery face, and $|.|$ denotes L_1-norm. For identification, a 1-NN classifier is employed on the masked images. Since in the previous registration phase, a specialized model is selected for each probe face, the adaptive approach should be imposed in the classification stage. Therefore, when the dissimilarities are computed, the probe face is compared against the gallery images registered using the corresponding model.

The depth-based identification experiment is conducted with three different registration approaches: (1) global face model-based ICP, as a baseline approach; (2) nose model-based ICP, which was previously used in [6]; and (3) the model-based ICP, where the model is selected adaptively, as initially proposed in [5]. In Table I, recognition rates for the Bosphorus and the UMB-DB databases are given.

When the identification results in Table I are compared, it is clear that using a bigger model is beneficial for the nonoccluded scans: For the neutral subsets, best performances are obtained when the whole face model is utilized. Nevertheless, the adaptive model-based registration has comparable results with the facial model, even though the adaptively selected model has at least 47.7% fewer surface points. This shows that the considered patch regions (eyes, nose, and mouth) provide sufficient information for registration. When the results on the occluded subsets are compared, it is clear that the face model-based registration is not applicable to occluded faces, and the advantage of the adaptive model over the nose model is clearly visible: For the Bosphorus database, the improvement is from 79.00% to 83.99%; and for the UMB-DB, the results are significantly improved from 46.27% to 65.25% with the baseline

TABLE II
GLOBAL IDENTIFICATION ACCURACIES WITH THE STANDARD AND THE MASKED FISHERFACES APPROACHES

Method	Fisherfaces Approach	Gallery Data	Probe Data	Bosphorus	UMB-DB
Fisherfaces [7]	Standard	Original	Original	53.28	43.56
Fisherfaces with Restoration [6]	Standard	Original	Restored	83.46	60.34
Fisherfaces with Gappy PCA [17]	Masked	Original	Masked	85.83	66.10
Proposed (Global)	Masked	Masked	Masked	87.40	69.15

depth-based classifier. The nose model-based registration fails on the UMB-DB, since in most of the occlusions in this data set, the nose area is covered. However, for the adaptive approach, the valid patches are used instead of using a single nose patch, and the identification rate is improved. On the other hand, in the Bosphorus database, occlusions over the nose area are small, yielding acceptable results even for the nose-model-based registration. It should be noted that this registration method assumes partial visibility of the nose area, since the initial alignment is based on nose detection. Nevertheless, the experimental results show that even the samples with over 50% nasal area occlusions are aligned. This is obtained by incorporating validity together with eye and mouth patches.

E. Regional Classification Performance Using Masked Projection

In this section, we evaluate the proposed *masked projection* classification approach. First, we start by considering the whole facial surface in an holistic manner. We compare two approaches to deal with missing data, where the occluded parts are either restored or removed. These approaches are evaluated in comparison with a baseline classifier, where the surfaces are considered without any preprocessing of the occluded parts. For removal, we have utilized global masked projection. For restoration, we have employed the partial Gappy PCA method of [6]: In partial Gappy PCA, the occluded parts are first removed from the surface. Then the whole facial surface is estimated using eigenvectors computed by PCA, where the estimated parts corresponding to the missing components are used to complete the facial surface.

In summary, we compare four different classification strategies in Table II: (1) The standard Fisherfaces [7] on original data, where no occlusion removal or restoration is applied (first row); (2) the standard Fisherfaces on restored data, where the missing parts are restored by partial Gappy PCA of [6] (second row); (3) the standard Fisherfaces applied on the *masked* probe features obtained by Gappy PCA [17] (third row); and (4) the proposed *masked* Fisherfaces, where the globally learned projection matrices are masked to obtain projections of both the gallery and the probe images (last row). As stated in Section III, the gallery images should also be projected using the masked projection approach, since the subspaces defined by $\mathbf{W}$ and $\mathbf{W}_m$ are different. This is the main difference between the proposed approach and the idea of the Gappy PCA, and a quantitative comparison between the two approaches are included in the last two rows of Table II. For the training of the Fisherfaces, the FRGC v.2 neutral subset is employed. As the results in Table II indicate, restoring occluded parts offers an improvement over original surfaces: For the Bosphorus, the performance is improved by 30%; for the more challenging UMB-DB, the improvement is 17%. However, we see that it is beneficial to remove the occluded parts, instead of restoring them: For the last two rows, the occlusions of the probe images are removed, whereas for the second row, restoration is employed: For the Bosphorus, 2–4% further increase is obtained; whereas for the UMB-DB, a more significant performance improvement (about 6–9%) is achieved. Furthermore, the proposed approach (last row) yields better results than of Gappy PCA (third row): For a fair comparison, the parameters used for the compared dimensionality reduction techniques are identically chosen (the dimensions used for PCA and Fisherfaces are 150 and 100, respectively). As these results indicate, instead of restoring occluded areas, it is beneficial to employ the masked projection, which incorporates only the nonoccluded surface regions. Moreover, the masked projection should be used to project both the gallery and the probe images.

Next, we evaluate the performances obtained by fusing the 40 separate regions, where the regional classifiers are fused at the score level by the product rule. Here, we compare three different classification strategies: (1) the standard regionally trained Fisherfaces on the restored data, which is included for comparative purposes; (2) the regionally trained masked Fisherfaces, where a set of regional projection matrices are learned and then masked by the occlusion mask; and (3) the globally trained masked Fisherfaces, where only a single projection matrix is learned and then masked by both the region and the occlusion mask. Once again, the FRGC v.2 neutral set is used for training of the Fisherfaces. For the analysis, we have conducted the experiments both with manually labeled occlusion masks, and the automatically detected masks. The performances are reported in Table III, where the results obtained by the ground truth occlusion masks (first three rows), and the automatically detected occlusion masks (last three rows) are given.

As the results in Table III indicate, better performances are obtained by the proposed masked projection approach. It is also clear that manual occlusion masking performs better than automatic occlusion masking. The difference is more significant for the UMB-DB, since it includes more challenging occlusion variations. The cumulative match characteristic (CMC) plots are given in Fig. 6 to verify the behavior of the considered classifiers. As these plots show, the CMC curves for the Bosphorus database are very similar, since the occlusions are relatively small and the regional division scheme can compensate for badly restored regions. For the UMB-DB database, the impact of masking is more visible: The standard Fisherfaces method on restored images performs poorly when compared with the masked approach. Moreover, these results indicate that employing masking to obtain regional projection matrices

TABLE III
REGIONAL IDENTIFICATION ACCURACIES WITH BOTH THE STANDARD AND THE NEWLY PROPOSED FISHERFACES APPROACHES (RESULTS ARE REPORTED WITH BOTH THE MANUAL AND THE AUTOMATIC OCCLUSION MASKS)

Method	Fisherfaces Approach	Training Approach	Gallery Data	Probe Data	Mask	Bosphorus	UMB-DB
Fisherfaces with Restoration	Standard	Regional	Original	Restored	Manual	93.44	71.19
Proposed (Occlusion Masking)	Masked	Regional	Masked	Masked	Manual	93.18	73.90
Proposed (Occlusion&Region Masking)	Masked	Global	Masked	Masked	Manual + Regional	93.18	73.56
Fisherfaces with Restoration	Standard	Regional	Original	Restored	Automatic	91.86	66.78
Proposed (Occlusion Masking)	Masked	Regional	Masked	Masked	Automatic	93.18	70.51
Proposed (Occlusion&Region Masking)	Masked	Global	Masked	Masked	Automatic + Regional	92.91	68.47

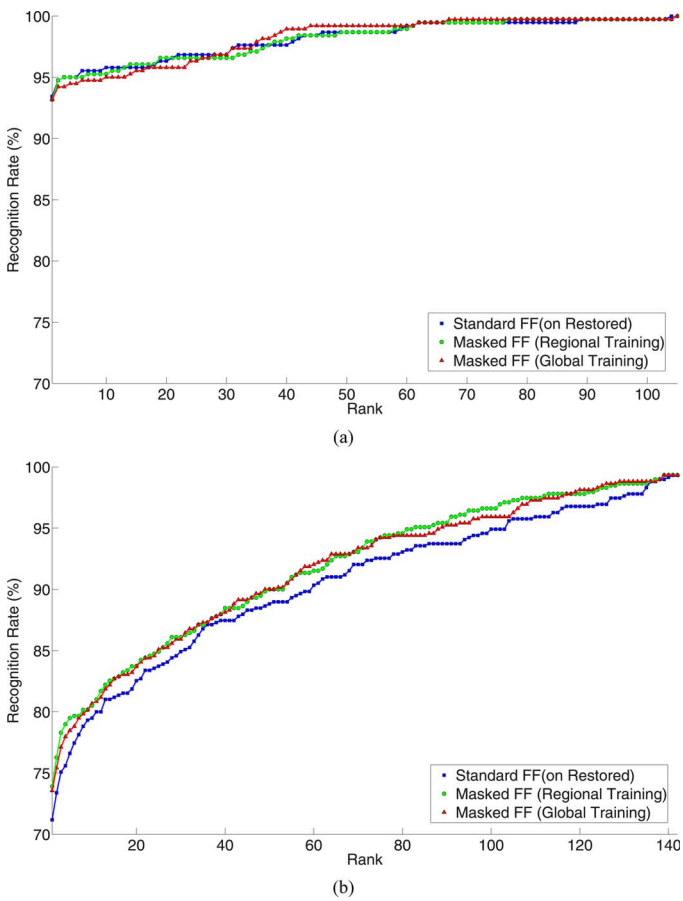


Fig. 6. Cumulative match characteristics (CMCs) plots for (a) the Bosphorus and (b) the UMB-DB databases, via three different approaches: 1) Standard Fisherfaces (FF) on restored images (blue squares), 2) masked Fisherfaces where regional projection matrices are obtained by regional training (green circles), and 3) masked Fisherfaces where regional projection matrices are also obtained by masking (red triangles).

by combining the occlusion and regional masks is a viable alternative. Furthermore, the performance of the proposed projection scheme is superior, when compared with the results reported in the literature on the UMB-DB database [15], where a PCA based classifier attains 56.50% identification rate on restored faces.

In Table III, the identification rates for the standard Fisherfaces over the restored images appear as comparable. However, further analysis show that this is a result of the successful regional division scheme considered: Since the regions are determined by considering possible facial occlusions, for an occluded (or restored) probe face, a number of regional classifiers consider only the nonoccluded parts. Therefore, they rectify the overall classification results. The regional classification results are reported using manual occlusion masks in Fig. 7(a) for the Bosphorus database, and in Fig. 7(b) for the UMB-DB, where three different classification approaches are compared[4]. The bars represent the performance improvement of the proposed masked Fisherfaces over the standard method, for different regions of the face. We observe that a performance improvement of 2 to 14% is obtained. When the globally trained masked Fisherfaces is compared with the regionally trained masked Fisherfaces, we see that performances are comparable and neither method performs better for all regions. We observe, however, that global training is superior for small regions due to the availability of more data in training. Global training is also preferred, since the learning is performed only once. The differences between the standard and the masked Fisherfaces performances are more prominent for the UMB-DB database, since this database contains more challenging occlusions.

As stated in Section III, a possible approach to deal with missing data in subspace analysis, is to remove the corresponding missing pixels from the training data and to learn the projection matrix from the masked training samples. Although this approach is not practical in occluded faces since each occlusion is unique, for comparative purposes, we have obtained recognition rates on the Bosphorus database using this masked training idea[5]. The regional results of masked training and masked projection approaches are compared in Fig. 8, where manually labeled occlusion masks are utilized. In contrast to our expectations, the newly proposed masked projection strategy gives better recognition results for all of the 40 different regions: Since in the masked projection approach, the regional projection matrix is learned from the complete training regions, the relation between the original face space and the lower-dimensional subspace is represented better. Therefore, instead of training the projection matrices separately for each probe face, it is beneficial to obtain a complete regional projection matrix in an offline manner and then to compute the corresponding projection matrix using the occlusion mask. In addition, the masked projection strategy is a more feasible method: Instead of retraining a projection matrix separately for

[4]The corresponding region for each region number is given in Fig. 2(b).

[5]The masked training results are included only for comparative purposes. Due to its high computational cost, the results are reported only on the Bosphorus database.

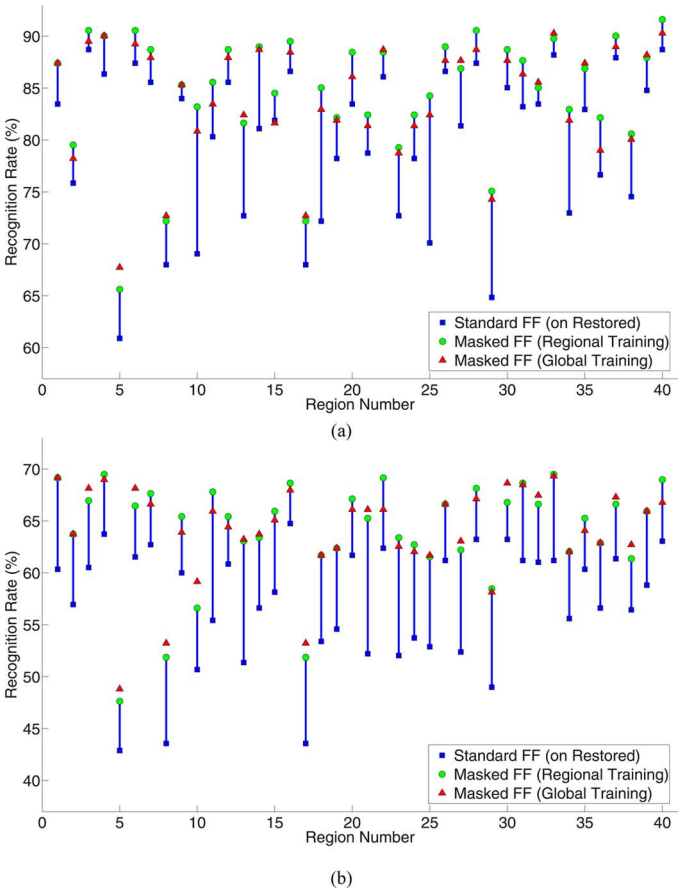


Fig. 7. Regional recognition rates using manual occlusion masks for (a) the Bosphorus and (b) the UMB-DB databases. The blue lines indicate the performance improvement of the masked regionally trained Fisherfaces (FF) approach (green circles) over the standard Fisherfaces on the restored images (blue squares). The results of the masked approach applied to obtain regional projections are also included (red triangles) for comparative purposes.

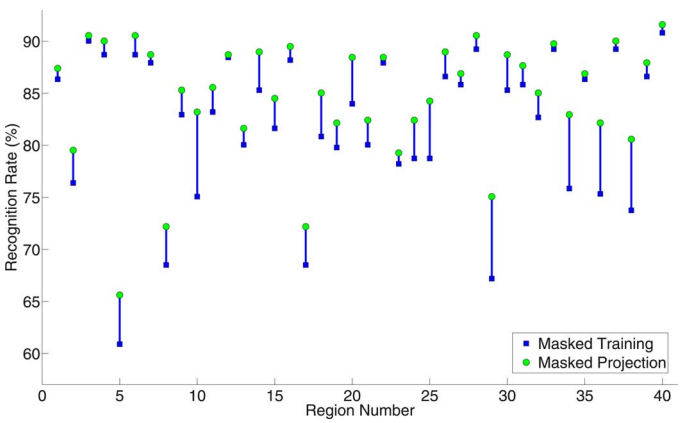


Fig. 8. Regional recognition rates using manual occlusion masks for masked training (blue squares) and masked projection (green circles). The results are obtained for the Bosphorus database using manual occlusion masks, where training is handled at the regional level.

each probe face, the corresponding masked projection matrix is computed from the complete projection matrix.

For evaluating the impact of occlusion on the recognition performance, we have analyzed the correctly and incorrectly identified samples for varying sizes of occluded areas. In Fig. 9,

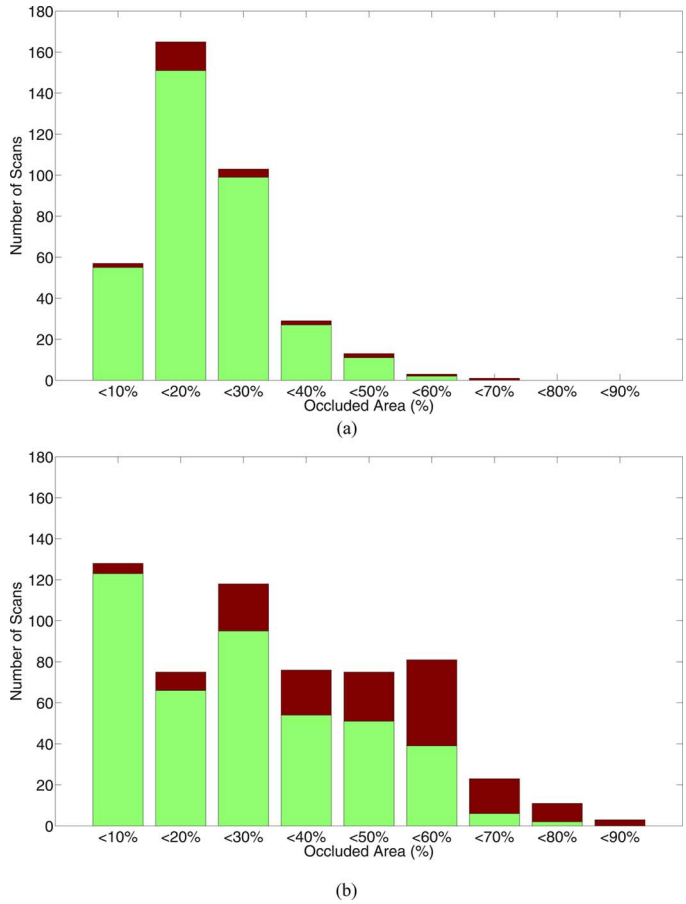


Fig. 9. Occluded area histogram for (a) the Bosphorus and (b) the UMB-DB databases. Correctly (green) and incorrectly (red) classified sample counts are shown for different occlusion percentage ranges.

the histograms of occlusion percentages are given for (a) the Bosphorus, and (b) the UMB-DB databases, where correctly and incorrectly classified sample counts are shown respectively in green and red. As these figures indicate, the UMB-DB database has more extensive occlusions when compared with the Bosphorus database. Nevertheless, both of the databases have some occlusions covering more than 50% of the facial area; and as the occluded areas expand, the recognition performances drop. However, when the highly occluded and correctly classified examples are investigated, it is clear that the proposed registration and recognition scheme serves as a viable approach: In Fig. 10, highly occluded and correctly classified examples are given (with both manually labeled and automatically detected occlusion masks) for the Bosphorus (Fig. 10(a)) and the UMB-DB (Fig. 10(b)) databases[6]. Even when the nose area is partially occluded or the facial surface has low visibility, the facial surface can still be classified correctly.

Next, we have experimented with the fusion scheme, to check if the overall performances can be improved: Until now the product rule was used to merge the regional dissimilarity measures, and all of the regional results were employed in the fusion. However, since the regional surfaces are occluded, some

[6]Note that, for the UMB-DB database, the occlusion masks are shown on the average face model, due to publishing constraints.

TABLE IV
REGIONAL IDENTIFICATION ACCURACIES WITH AND WITHOUT EMPLOYING CONFIDENCE THRESHOLDING IN FUSION

Method	Fusion Method	Bosphorus	UMB-DB
Proposed (Occlusion Masking)	Product Rule	93.18	73.90
Proposed (Occlusion&Region Masking)	Product Rule	93.18	73.56
Proposed (Occlusion Masking)	Product Rule with Confidence Thresholding	93.44	75.59
Proposed (Occlusion&Region Masking)	Product Rule with Confidence Thresholding	93.70	74.75

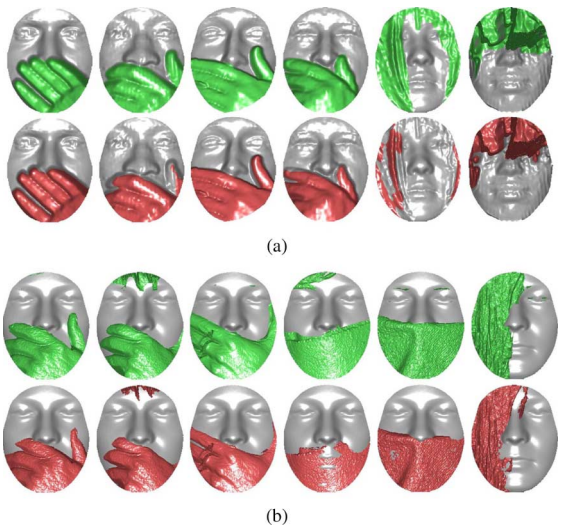


Fig. 10. Highly occluded sample images that are correctly classified by the proposed masked Fisherfaces approach: Examples from (a) the Bosphorus and (b) the UMB-DB databases, where top and bottom rows show the corresponding manually labeled (in green) and automatically detected (in red) occlusion masks.

regions will produce erroneous measures, and they can be discarded in the fusion stage for further improvement. Here, we have employed the confidence estimation technique of [19] to decide on the regional classifiers to be taken into account in fusion: For a probe face, the dissimilarity scores are first normalized and sorted in ascending order. Then a second normalization is performed such that the first score becomes one. After the second normalization, the second dissimilarity value denotes the slope between the normalized scores of the first two top-ranked classes. Therefore, this value defines the confidence of the classifier. Using this approach, separate confidence values are computed for regional classifiers for the considered probe face. In fusion, only the classifiers that have confidence values more than the preset threshold value are considered. In Table IV, masked projection results obtained with ground truth occlusion masks are given. For the results where confidence thresholding is employed, a threshold of 0.75 is used. For the Bosphorus database, the results are not affected significantly, since the occlusions are small in size. For the UMB-DB database, which has more challenging occlusion variations, we were able to obtain up to 1.69% improvement by employing confidence values in the fusion stage.

F. Time Complexity

It is worth noting the overall time complexity of the proposed registration and classification scheme: In the registration stage, a single model-based ICP is necessary, where the models include at most half of the whole facial surface. In the classification stage, the probe-specific projection is computed by simply masking the globally trained matrix. The most time-consuming part of the pipeline is the construction of the curvature map. The detailed average timings for processing a single test face, with an unoptimized MATLAB code running on a 64-bit Core i7 2.67 GHz PC with 12 GB RAM, are as follows: The nose detection stage, including the curvature map generation and template matching steps, takes about 21 seconds. Adaptive model selection and model-based registration takes a total of about 6 seconds. The subsequent occlusion detection stage is negligible (about 2 ms). The final masked projection and classification steps take a total of 3γ ms, where γ is the number of images in the gallery set, e.g., 315 and 423 ms for the Bosphorus and UMB-DB respectively[7].

Since most of the time is consumed at the nose detection stage, we further examined the timing measures and checked if any improvements can be obtained. The most important factor influencing the computation durations, is the number of model points, which can be reduced by downsampling. Therefore, we have further analyzed the timing measures of our system, where the resampling rate of the depth maps is altered. For the nose detector, the resampling rate can be lowered to obtain significant computational time improvement, while maintaining the exact nose detection performance: If the regular resampling rate is lowered by a factor of 16 (a grid with four times larger step is employed), time consumption of the system can be lowered significantly without any performance degradation: Curvature map generation and template matching durations drop from 21 seconds to four seconds, where the template matching takes about only 0.1 seconds. Although working on a low-resolution model is computationally beneficial without sacrificing the nose detection accuracy, we resort to a high resolution grid for subsequent stages. With a low resolution model, the alignment process cannot converge well and lower recognition rates are obtained due to worse registration, even when the resampling rate is lowered only by a factor of 1.5. Thus, sparser depth maps can be beneficial for the nose detection stage, whereas for the registration process, denser depth maps should be preferred.

G. Masked Projection for Other Acquisition Scenarios

The adaptive model-based registration and the masked projection approaches proposed in this paper are motivated by large occlusions causing a high proportion of missing points. In previous subsections, we have shown the viability of the proposed approach for occlusion scenarios. A natural question that arises is the applicability to other acquisition scenarios. To answer this question, we have conducted experiments on different subsets of the Bosphorus database: neutral, expression, and pose (up to

[7]The classification time is given in terms of the size of the gallery set. Nevertheless, its contribution to the overall time can be kept low by parallelizing the distance computation.

TABLE V
IDENTIFICATION ACCURACIES OF MASKED PROJECTION ON BOSPHORUS NEUTRAL, OCCLUSION, EXPRESSION, AND LIMITED POSE SUBSETS

Method	Training Approach	Mask	Classification Approach	Bosphorus Subsets			
				Neutral	Occlusion	Expression	Pose
Proposed (Global)	Global	Automatic	Global	100.00	83.73	88.24	85.83
Proposed (Occlusion&Region Masking)	Global	Automatic + Regional	Regional	100.00	92.91	95.04	88.15

30 degrees) variations. As in the previous Bosphorus experiments, the gallery contains the first neutral image of each subject, and has a total of 105 scans. The neutral probe subset consists of 194 scans, whereas the expression variations are a total of 2620 faces. The Bosphorus database includes 13 pose variations for each of the 105 subjects, and six of the variations have extreme poses (45 or 90 degrees of rotation). The other seven pose types include: three yaw rotations (10, 20, and 30 degrees) and four pitch variations (slightly up, slightly down, up, down). For the experiments, the extreme poses labeled as 45 and 90 degrees are discarded, and the pose subset contains the remaining 734 scans.

Table V summarizes the performance of the two proposed approaches, namely the global and the regional (occlusion and region masking) methods, on neutral, occlusion, expression, and pose subsets. On the neutral subsets, we obtain 100% identification accuracy. We repeat the occluded subset results here for comparative purposes: For global and regional methods, recognition accuracies of 83.73% and 92.91% are obtained respectively, where automatically detected occlusion masks are employed. The performance on the expression subset is higher: 88.24% and 95.04%. The last column shows the performance on the challenging pose subset: 85.83% and 88.15%. As these results indicate, our proposed system can be directly applied to neutral and expression scans. The results obtained on the expression subset are similar to the results of [4], for which the system is implemented directly for expression handling. The herein proposed approach is advantageous for expression variations, when both the registration and classification methodologies are considered: The idea of adaptive selection of the alignment model is beneficial for expression variations, since some patches can have extreme surface deformations and can mislead the alignment process. By using patch validity values, patches with expressive deformations are discarded from the registration process. Furthermore, occlusion detector automatically finds regions that do not resemble a neutral face. Therefore, expressive deformations are discarded from the classification process. When pose variations are considered, acceptable recognition rates are achieved. When the misclassified examples are examined, we saw that some of the faces exhibit pose variations greater than 30 degrees due to mislabeling during the database acquisition: For around 70% of the incorrect classifications, the face was altered from the frontal pose by more than 20 degrees. When handling *extreme* pose variations, the bottleneck of the system is the registration process: Since patch templates are obtained from the frontal average model, patch localization accuracy will be degraded for images with extreme rotations. Furthermore, the ICP algorithm will not be able to converge. If the facial surfaces are registered correctly to the adaptively selected alignment model, the classification stage can directly be applied: The mask detection procedure will easily and correctly locate the missing parts of the facial surface. With accurate occlusion masks defining the missing parts, the subsequent masked projection will perform well. As the results in Table V indicate, even though the system proposed is especially for occlusion variation handling, acceptable recognition results are obtained for other acquisition challenges such as expression and pose variations. Furthermore when results for each scenario are examined, it is clear that considering the faces as a combination of multiple regions further improves the recognition performance.

V. CONCLUSION

3-D face recognition has become an emerging biometric technique. However, especially in noncooperative scenarios, occlusion variations complicate the task of identifying subjects from their face images. In this paper, we present a fully automatic 3-D face recognizer, which is robust to facial occlusions.

For the alignment of occluded surfaces, we utilized a model-based registration scheme, where the model is selected adaptively to the facial occlusion. The alignment model is formed by automatically checking patches for validity and including only nonoccluded facial patches. By registering the occluded surface to the adaptively selected model, a one-to-one correspondence is obtained between the model and the nonoccluded facial points. Hence, occlusion insensitive facial registration can be achieved. When compared with the holistic face model-based registration strategy (which is often utilized for nonoccluded surfaces), about 20% improvement in identification rate is achieved with the adaptive-model approach for both the Bosphorus and the UMB-DB databases.

Following the occlusion detection stage, the facial parts detected as occluded are removed to obtain occlusion-free surfaces. Classification is handled on these occlusion-free faces. In this work, we propose *masked projection*, which incorporates a masking scheme into a subspace analysis technique, namely the Fisherfaces, to enable applicability to incomplete data. Subspace training is handled offline; and at the classification stage, the occlusion mask of the probe face is applied to the projection matrix. The masked projection matrix is used to project the gallery set and the probe face to the corresponding subspace, and identification is achieved by 1-nearest neighbor classifier. To further improve the overall identification performance, a regional classification scheme is employed: The facial surfaces are considered as a collection of overlapping regional parts; and each region acts as a separate classifier. In the regional level, the masked projection is applied in two different strategies: (1) each regional subspace is trained; and (2) the regional subspaces are obtained by applying the region mask to the global projection

matrix. As the experimental results indicate, by masked projection an improvement up to 14% can be achieved at the regional level. Furthermore, employing the masking approach to obtain regional subspaces appears as a viable alternative over regional training. Additionally, the proposed system can be directly applied to handle expression and small pose variations.

The proposed system is able to work with good performance under substantial occlusions, expressions, and small pose variations. When we examine the failures, we see that if occlusions are so large that the nose area is totally invisible, the initial alignment becomes impossible. Similarly, if the face is rotated by more than 30 degrees, it becomes difficult to accomplish the initial alignment. In our future work, we plan to develop alternative initial alignment techniques. Furthermore, the automatic occlusion detection stage can also be improved: As a future direction, we plan to model occlusions better, so that the overall performance of the system can be increased.

REFERENCES

[1] 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.

[2] 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, pp. 1–14.

[3] N. Alyuz, B. Gokberk, and L. Akarun, "A 3D face recognition system for expression and occlusion invariance," in *Proc. Int. Conf. Biometrics: Theory, Applications and Systems (BTAS)*, 2008, pp. 1–7.

[4] N. Alyuz, B. Gokberk, and L. Akarun, "Regional registration for expression resistant 3-D face recognition," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 3, pp. 425–440, Sep. 2010.

[5] N. Alyuz, B. Gokberk, and L. Akarun, "Adaptive model based 3D face registration for occlusion invariance," in *Proc. Eur. Conf. Computer Vision—Workshops—Benchmarking Facial Image Analysis Technologies (BeFIT)*, Florence, Italy, 2012.

[6] N. Alyuz, B. Gokberk, L. Spreeuwers, R. Veldhuis, and L. Akarun, "Robust 3D face recognition in the presence of realistic occlusions," in *Proc. Int. Conf. Biometrics (ICB)*, 2012, pp. 111–118.

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

[8] 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.

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

[10] K. Chang, W. Bowyer, and P. 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.

[11] A. Colombo, C. Cusano, and R. Schettini, "Detection and restoration of occlusions for 3D face recognition," in *Proc. Int. Conf. Multimedia and Expo*, 2006, pp. 1541–1544.

[12] A. Colombo, C. Cusano, and R. Schettini, "Recognizing faces in 3D images even in presence of occlusions," in *Proc. Int. Conf. Biometrics: Theory, Applications and Systems (BTAS)*, 2008, pp. 1–6.

[13] A. Colombo, C. Cusano, and R. Schettini, "Gappy PCA classification for occlusion tolerant 3D face detection," *J. Math. Imag. Vis.*, vol. 35, no. 3, pp. 193–207, 2009.

[14] A. Colombo, C. Cusano, and R. Schettini, "Three-dimensional occlusion detection and restoration of partially occluded faces," *J. Math. Imag. Vis.*, vol. 40, no. 1, pp. 105–119, 2011.

[15] A. Colombo, C. Cusano, and R. Schettini, "UMB-DB: A database of partially occluded 3D faces," in *Proc. Int. Conf. Computer Vision (ICCV)—Workshops*, 2011, pp. 2113–2119.

[16] M. De Smet, R. Fransens, and L. Van Gool, "A generalized EM approach for 3D model based face recognition under occlusions," in *Proc. Int. Conf. Computer Vision and Pattern Recognition (CVPR)*, 2006, vol. 2, pp. 1423–1430.

[17] R. Everson and L. Sirovich, "Karhunen-Loeve procedure for gappy data," *J. Opt. Soc. Amer. A*, vol. 12, no. 8, pp. 1657–1664, 1995.

[18] S. Fidler, D. Skocaj, and A. Leonardis, "Combining reconstructive and discriminative subspace methods for robust classification and regression by subsampling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 3, pp. 337–350, Mar. 2006.

[19] 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, Cyben.*, vol. 38, no. 1, pp. 155–173, Feb. 2008.

[20] B. Gokberk, M. O. Irfanoglu, and L. Akarun, "3D shape-based face representation and feature extraction for face recognition," *Image Vis. Comput.*, vol. 24, no. 8, pp. 857–869, Aug. 2006.

[21] 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. London, U.K.: Springer-Verlag, 2008, pp. 1–33.

[22] R. He, W. Zheng, and B. Hu, "Maximum correntropy criterion for robust face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1561–1576, Aug. 2011.

[23] J. Kim, J. Choi, J. Yi, and M. Turk, "Effective representation using ICA for face recognition robust to local distortion and partial occlusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 12, pp. 1977–1981, Dec. 2005.

[24] D. Lin and X. Tang, "Quality-driven face occlusion detection and recovery," in *Proc. Int. Conf. Computer Vision and Pattern Recognition (CVPR)*, 2007, pp. 1–7.

[25] T. Lo and J. Siebert, "SIFT keypoint descriptors for range image analysis," in *Annals of the BMVA X*, 2009, pp. 1–18.

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

[27] A. Martinez, "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 6, pp. 748–763, Jun. 2002.

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

[29] B. Park, K. Lee, and S. Lee, "Face recognition using Face-ARG matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 12, pp. 1982–1988, Dec. 2005.

[30] J. Park, Y. Oh, S. Ahn, and S. Lee, "Glasses removal from facial image using recursive error compensation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 805–811, May 2005.

[31] 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)*, vol. 7408, 2007.

[32] 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. Int. Conf. Computer Vision and Pattern Recognition (CVPR)*, 2005, vol. 1, pp. 954–954.

[33] 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.

[34] A. Savran, N. Alyuz, H. Dibeklioglu, O. Celiktutan, B. Gokberk, B. Sankur, and L. Akarun, "Bosphorus database for 3D face analysis," *Biometrics Identity Manage.*, pp. 47–56, 2008.

[35] A. Scheenstra, A. Ruifrok, and R. C. Veltkamp, "A Survey of 3D Face Recognition Methods," in *Audio- and Video-Based Biometric Person Authentication*, ser. Lecture Notes in Computer Science. New York, NY, USA: Springer, 2005, vol. 3546, pp. 891–899.

[36] F. Tarres, A. Rama, and L. Torres, "A novel method for face recognition under partial occlusion or facial expression variations," in *Proc. ELMAR Int. Symp.*, 2005, pp. 163–166.

[37] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[38] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local Gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition," in *Proc. Int. Conf. Computer Vision (ICCV)*, 2005, vol. 1, pp. 786–791.

[39] Z. Zhou, A. Wagner, H. Mobahi, J. Wright, and Y. Ma, "Face recognition with contiguous occlusion using Markov random fields," in *Proc. Int. Conf. Computer Vision (ICCV)*, 2009, pp. 1050–1057.

[40] L. Zou, S. Cheng, Z. Xiong, M. Lu, and K. Castleman, "3-D face recognition based on warped example faces," *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 3, pp. 513–528, Sep. 2007.

Nese Alyuz received the B.S. degree in computer engineering from Istanbul Technical University, Istanbul, Turkey, in 2005. She obtained the M.Sc. degree in computer engineering from Bogaziçi University, Istanbul, Turkey, in 2008. She is currently a Ph.D. candidate at the Perceptual Intelligence Laboratory of Bogaziçi University, working on "3D Face Recognition under Occlusion Variations." Her research areas are biometrics, computer vision, image processing, and pattern recognition.

Berk Gokberk received the B.S., M.Sc., and Ph.D. degrees in computer engineering from Bogaziçi 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 the Department of Electrical Engineering Mathematics and Computer Science, University of Twente. His research interests are in the areas of biometrics, computer vision, computer graphics, and pattern recognition.

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

From 1993 to the present, she has been a Faculty Member at, Bogazici University, where she serves as professor of Computer Engineering since 2001. Her research areas are biometrics; face recognition and hand vein recognition; and human computer interaction; with emphasis on human activity and gesture analysis.

Prof. Akarun has worked on the organization committees of the IEEE NSIP99, EUSIPCO 2005, eNTERFACE2007, and ICPR2010.