Robust 3D Face Recognition in the Presence of Realistic Occlusions

Neşe Alyüz[1], Berk Gökberk[2], Luuk Spreeuwers[2], Raymond Veldhuis[2], and Lale Akarun[1]

[1]Department of Computer Engineering, Boğaziçi University, Istanbul, Turkey
[nese.alyuz,akarun]@boun.edu.tr
[2]Department of Electrical Engineering Mathematics and Computer Science, University of Twente, The Netherlands
[b.gokberk,l.j.spreeuwers,r.n.j.veldhuis]@utwente.nl

Abstract

Facial occlusions pose significant problems for automatic face recognition systems. In this work, we propose a novel occlusion-resistant three-dimensional (3D) facial identification system. We show that, under extreme occlusions due to hair, hands, and eyeglasses, typical 3D face recognition systems exhibit poor performance. In order to deal with occlusions, our proposed system employs occlusion-resistant registration, occlusion detection, and regional classifiers. A two-step registration module first detects the nose region on the curvedness-weighted convex shape index map, and then performs fine alignment using nose-based Iterative Closest Point (ICP) algorithm. Occluded areas are determined automatically via a generic face model. After non-facial parts introduced by occlusions are removed, a variant of Gappy Principal Component Analysis (Gappy PCA) is used to restore the full face from occlusion-free facial surfaces. Experimental results obtained on realistically occluded facial images from the Bosphorus 3D face database shows that, with the use of score-level fusion of regional Linear Discriminant Analysis (LDA) classifiers, the proposed method improves rank-1 identification accuracy significantly: from 76.12% to 94.23%.

1. Introduction

As a biometric modality, face is preferred due to its contactless acquisition technology and its potential for non-cooperative scenarios. Recently, it has been shown that the face biometric can attain high performance under specific conditions, reaching the high level of security achieved by fingerprint or iris modalities [17]. However, face recognition remains a challenging task when adverse scenarios are considered such as the situations where illumination variations, pose changes, facial expressions, or occlusions covering the facial surface are present. In the three-dimensional (3D) domain, some of these challenges can be handled more effectively than in 2D: illumination differences can be avoided, small pose changes can be corrected. However, handling of extreme occlusions remains a challenging task, since occluding objects change the 3D surface information substantially.

Previous work on designing occlusion-resistant recognition systems usually involves 2D appearance-based techniques. In some of these 2D face recognition studies, the facial surface is divided into local regions, which are independently compared and the local classifiers are fused by basic voting [10], or by a probabilistic approach [14]. In [21], facial parts that are likely to have occlusions are eliminated. In [15], occlusions caused by eyeglasses are considered and a technique to compensate for the missing data is proposed. For the 3D modality, Colombo *et al.* proposed several approaches to detect the occluded regions [6], [7]. After the occlusions are detected, Gappy Principal Component Analysis (Gappy PCA) [8] is utilized to reconstruct the facial image. Reconstructed faces are then identified by a holistic approach. In [1], Alyuz *et al.* divide the facial surface into local parts and fuse the regional classifiers to cope with surface variations caused by expressions and occlusions. In that work, occlusions are not automatically detected and occlusion robustness is handled at the classifier fusion phase. An attempt to detect and restore occlusions in registered 3D facial surfaces has been made in [2]. However, registration of occluded surfaces remains an unsolved problem.

In this work, we aim to tackle the occlusion handling problem by i) designing an occlusion invariant 3D facial registration method, ii) detecting the occluded areas to obtain occlusion-free surfaces, iii) restoring or ignoring the missing parts, and iv) using multiple regional classifiers. For occlusion detection, accurate facial surface registration is a crucial step. In this work, we handle registration by a nose-based approach which assumes partial visibility of the nose. After automatically locating the nose area, a local rigid surface fitting with the Iterative Closest Point (ICP) method is carried out to globally align the. Estimation of alignment parameters using local ICP-based method makes the overall registration scheme occlusion insensitive. Once faces are transformed into a canonical coordinate system and aligned, non-facial surface regions caused by an occluding object are determined at the occlusion detection phase. At this phase, a generic face model is used to locate surface parts that belong to the occluding object. By the removal of these outlier regions, we can obtain an *occlusion-free*[1] face where occluded parts are removed from the original surface data. Occlusion-free faces contain actual facial surface data with missing regions. Therefore, it is possible to use them for identification purposes in two ways: 1) by restoring the missing regions as accurate as possible and 2) by ignoring the reconstruction process to use occlusion-free faces directly in classification. In our work, we studied both approaches. For restoration-based approach, we use face specific modeling via Gappy PCA that allows reconstruction of faces with missing parts. For the latter approach, using occlusion-free faces directly without restoration, we compute dissimilarity scores on mutually available depth image regions of gallery and probe faces. Lastly, we investigate the benefits of component-based representation schemes with respect to holistic approaches. We show that local component-based representations of restored faces which enable the use of statistical feature extraction schemes such as Linear Discriminant Analysis (LDA), can outperform holistic approaches. In our work, we used the Bosphorus 3D face database which is the largest publicly available database that contains realistic 3D facial occlusions. In the Bosphorus database, from each of the 105 subjects, several types of occluded scans were collected systematically.

2. Proposed System

Our occlusion robust 3D face recognition method consists of five main stages: (1) nose localization based on curvature maps, in order to provide an initialization for fine registration; (2) nose-based fine registration via ICP, to transform 3D faces into a common coordinate system; (3) occlusion detection and removal, where the extraneous objects are found automatically and discarded; (4) restoration, to fill in the parts that are labeled as occluded in the previous phase; and (5) classification, where local classifiers are combined to improve the performance of the recognizer. The overall diagram of the system is given in Figure 1.

3. Nose-based Registration

Before any comparison between two faces can be made, the facial surfaces should be brought into a common coordinate system. In 3D registration, the surfaces are aligned and a correspondence between the two sets of surface points is obtained. However, in the presence of facial occlusions, the geometry of the occluded face is altered by the exterior object, complicating the registration process. In this work, we propose an occlusion-robust 3D registration approach based on the assumption that the nose area will be mostly visible. First, the nose area is automatically detected for coarse alignment. Fine tuning of the initial alignment is handled by nose-based ICP registration. The ICP variant used in this work registers nasal facial area of the test face to a *generic nose model*, referred to as the *Average Region Model* (AvRM). The nose AvRM is utilized in both stages of the registration: (1) in nose detection, serving as a template; (2) in fine registration, defining the reference.

3.1. Nose Detection for Initial Registration

ICP algorithm [3] is widely used in geometric alignment of rigid 3D surfaces. However, the performance of iterative approaches like ICP rely heavily on the initial conditions. Hence, it is necessary to provide an initial registration, which will be improved iteratively. Most of the 3D face recognition systems are dependent on localization of facial landmarks [13], [4], [5]. Unfortunately, the presence of occluding objects over the facial surface complicate the task of locating such distinctive features, especially around nose borders and inner eye cavities. Therefore, in this work, we preferred to estimate the location of the whole nose area instead of localizing individual landmark points.

Our nose detection algorithm utilizes surface curvature information, which is advantageous due to its invariance to rotation and translation. We compute two curvature maps for a surface: (1) the shape index map; (2) the curvedness map. Introduced by [11], these two local shape measures can be defined as polar coordinate transformation of maximum (κ_{max}) and minimum (κ_{min}) curvatures. This transformation provides the separation of components that are scale-independent and scale-dependent [22]. Scale-independent components such as shape index provide the distinction between spherical and cylindrical surfaces, whereas scale-dependent components such as curvedness give the magnitude of curvature. Given κ_{max} and κ_{min}, shape index value $SI(p)$ at surface point p can be computed as follows:

$$SI(p) = \frac{1}{2} - \frac{1}{\pi} tan^{-1} \frac{\kappa_{max}(p) + \kappa_{min}(p)}{\kappa_{max}(p) - \kappa_{min}(p)} \qquad (1)$$

The shape index map SI takes values in $[0, 1]$ and provides a gradual transition between concave ($0 < SI < 0.5$) and convex ($0.5 < SI < 1$) shapes. The scale-dependent counterpart of shape index is the curvedness value. This measures the rate of the curvature at each point, giving the degree of how curved the surface is. Curvedness value $C(p)$

[1] In our work, we refer to occlusion-free faces as faces where extraneous occluding object's surface is removed from scanned 3D facial data.

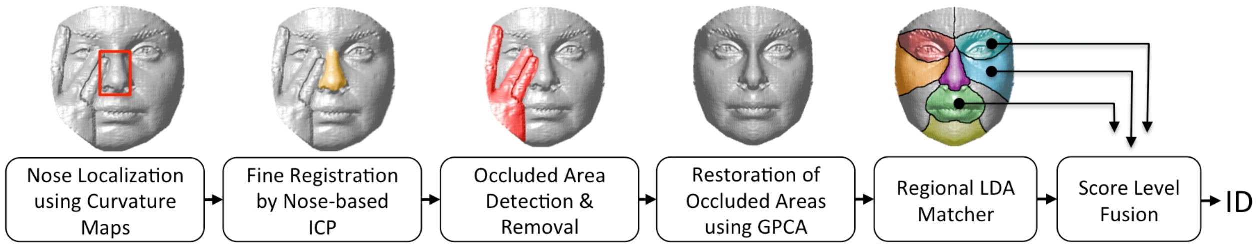

Figure 1. Overall diagram of the proposed method.

at point p is defined as

$$C(p) = \sqrt{\frac{\kappa_1(p)^2 + \kappa_2(p)^2}{2}}. \tag{2}$$

A planar surface will have a curvedness value of zero, whereas a non-planar surface will have a curvedness value proportional to its rate of curvature. Our nose detector first constructs shape index and curvedness maps. Since nose is a convex structure, thresholding is applied over the SI map (where th=0.5), eliminating concave regions. The convex SI map, referred to as SI_{cx}, is defined as

$$SI_{cx}(p) = \begin{cases} 0 & \text{if } SI(p) < 0.5 \\ SI(p) & \text{otherwise.} \end{cases} \tag{3}$$

After the construction of SI_{cx}, it is weighted with curvedness [12] to integrate scale-dependent components with scale-independent curvature information:

$$WSI(p) = SI_{cx}(p) * C(p) \tag{4}$$

Here, WSI is the curvedness-weighted convex shape index. Shape index, convex shape index, curvedness, and weighted convex shape index maps are illustrated in Figure 2 for an example face image.

(a) (b) (c) (d) (e)

Figure 2. Curvature maps utilized for nose detection are illustrated on an example image: (a) depth image, (b) shape index, (c) convex shape index, (d) curvedness, and (e) weighted convex shape index.

As the example in Figure 2 illustrates, the nose region appears as a distinguishable fork-shaped structure in the WSI map. We propose to localize the nose area via template matching, and for this purpose we construct a nose template. We first construct an average face model using a set of registered neutral and non-occluded training images using Thin Plate Spline warping algorithm [18]. Then, we manually crop the average face model to obtain the average nose region model. On this regional model, we construct the convex shape index and curvedness maps, and finally obtain the WSI map. The final regional WSI map constitutes the nose template. Given a test image, normalized cross-correlation based template matching is then performed to locate the most similar nose-like region.

3.2. ICP-based Fine Registration

The localization of the nose area provides a starting point for the fine registration process. Assuming that the facial surfaces are nearly frontal, faces can be coarsely aligned by translating the center of the detected nose region to the center of the reference nose template. This translation provides a sufficient initialization, whereas small pose deflections will be compensated in ICP-based fine registration. The aim of the fine registration step, is to find a rigid transformation that aligns the probe face, $P = \{\mathbf{p}_1, \ldots, \mathbf{p}_f\}$ to the nose AvRM, $N = \{\mathbf{n}_1, \ldots, \mathbf{n}_m\}$ where $\mathbf{n}$ and $\mathbf{p}$ are 3D points for the probe and nose AvRM surfaces, respectively. The transformation $\mathbf{T}$ can be defined by three rotations around x, y, z axes, $\mathbf{R}_x, \mathbf{R}_y, \mathbf{R}_z$ respectively, and a translation $\mathbf{t}$:

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

The error of transformation can then be computed as:

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

where $\mathbf{n}_j$ is model point corresponding to the surface point $\mathbf{p}_i$. The registration can be optimized by finding the transformation that minimizes the total error:

$$\hat{\mathbf{T}} = \underset{\mathbf{T}}{\operatorname{argmin}} \sum_{j=1}^{m} ||\mathbf{T}(\mathbf{p}_i) - \mathbf{n}_j|| \tag{7}$$

In this work, ICP method is utilized to solve the system of linear equations. After the convergence of the ICP, the final transformation is applied to the whole probe face, bringing it in alignment with the average face model.

4. Occlusion Detection

After the facial surfaces have been registered, it is important to locate facial areas occluded by exterior objects.

In this work, we have implemented a generic face model-based occlusion detection method. In this approach, occlusion detection is handled by thresholding the difference map obtained by computing the absolute difference between an average face model and the input face. Afterwards, a binary mask is obtained by thresholding the difference map. The binary occlusion mask is then post-processed by morphological dilation and connected component analysis operations.

5. Restoration with Gappy PCA

Gappy PCA [8] was proposed as a Principal Component Analysis (PCA) variant to handle data with missing components. With Gappy PCA, it is possible to reconstruct original signal up to a certain degree when the signal contains missing values. In order to estimate the unknown facial data by the Gappy PCA method, locations of the missing components are required. Prior to estimation, Gappy PCA method constructs a general model of facial data using a training set of non-occluded images. The basis vectors are determined using a training set of N non-occluded face images, $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\} \subset \mathbf{R}^n$. A face image $\mathbf{x}$ can be estimated using a subset ($M < N$) of these basis vectors:

$$\mathbf{x} = \boldsymbol{\mu} + \sum_{i=1}^{M} \alpha_i \mathbf{v}_i \tag{8}$$

where the vector $\boldsymbol{\mu}$ defines the mean, and $\mathbf{v}_i$ is an eigenvector whose coefficient is α_i. The eigenvector coefficients are obtained by the inner product of the input vector and the corresponding eigenvector. Suppose there is an incomplete version of $\mathbf{x}$, namely $\mathbf{y}$, whose missing components are encoded in the occlusion mask. In Gappy PCA, the aim is to find a similar expression that approximates the incomplete data as in (8):

$$\mathbf{y} \simeq \tilde{\mathbf{y}} = \boldsymbol{\mu} + \sum_{i=1}^{M} \beta_i \mathbf{v}_i \tag{9}$$

However the β_i coefficients cannot be computed by the simple inner product method. Instead, the coefficients minimizing the squared reconstruction error should be sought. A basic definition of the squared reconstruction error would be given as $E = ||\mathbf{y} - \tilde{\mathbf{y}}||^2$.

To improve the error term, only the available information should be involved in the calculations. To discard the missing components, the gappy norm must be used, where the information about the missing components is encoded in the mask $\mathbf{m}$. The gappy norm for a vector $\mathbf{u}$ with the mask $\mathbf{m}$ can be defined as $||\mathbf{u}|| = \sqrt{(\mathbf{u}, \mathbf{u})_m}$ where

$$(\mathbf{u}, \mathbf{u})_m = \sum_{i=1}^{n} u_i u_i m_i. \tag{10}$$

Using the gappy norm, the reconstruction error term can be redefined as: $E_m = ||\mathbf{y} - \tilde{\mathbf{y}}||^2_m$. If we rewrite the error term by opening the squared terms and differentiating with respect to each β_i coefficient, we obtain a linear system of M equations:

$$\frac{\partial E}{\partial \beta_i} = -\mathbf{z}_i + \sum_{j=1}^{M} \beta_j A_{ij} = 0. \tag{11}$$

where $\mathbf{z}_i = (\mathbf{y}, \mathbf{v}_i)_m$ and $A_{ij} = (\mathbf{v}_i, \mathbf{v}_j)_m$. The linear system can be rewritten as $\mathbf{A}\boldsymbol{\beta} = \mathbf{z}$ and the coefficients can be computed as follows: $\boldsymbol{\beta} = \mathbf{A}^{-1}\mathbf{z}$.

After the coefficients are computed, the incomplete image can be reconstructed by Equation 9. As opposed to the previous works that employ the Gappy PCA method for reconstruction [6], [7], we propose to use the reconstructed data only for the missing components and the original data for the non-occluded facial regions. We refer to this method as partial Gappy PCA (pGPCA).

6. Regional Classification

In the classification stage, we propose to consider the 3D surface as a combination of several local regions. If the facial area is partially occluded by external objects, information regarding the covered regions will not be available. Therefore it will be beneficial to incorporate separate local classifiers and then to fuse the regional results. In this work, we utilized the regional division scheme used in [20] and constructed a total of 30 overlapping regions as visualized in Figure 3.

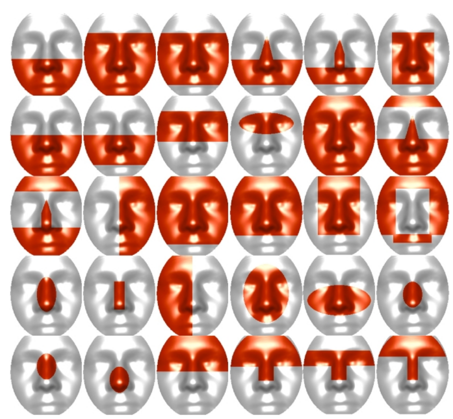

Figure 3. 30 regions used in regional classification system (in red).

For classification, we have applied Linear Discriminant Analysis (LDA) method to point set features independently for each region. The model based registration enables the use of LDA, since 3D faces are represented by ordered vectors of same the length after the alignment process. Since the facial surfaces are considered as a combination of several regions, separate LDA spaces are trained for each facial region. The LDA transformation matrix is computed on a separate face database in order to carry out independent evaluation of our face recognition system. The dissimilarity between corresponding regions of two facial surfaces can be measured by the angular cosine distance. The regional classifiers are then fused at the score-level by the product rule.

7. Experimental Results

7.1. Face Database

For the recognition experiments, we have used the Bosphorus 3D Face Database [19], which includes expression, pose, and occlusion variations. The database is acquired from a total of 105 subjects, 61 males and 44 females. The set consists of a total of 4666 scans, where each subject has approximately 34 expression, 13 pose, and 4 occlusion variations. Since we address the occlusion problem in this work, images containing expression and pose variations are discarded. Neutral images are used to construct the gallery set, which includes 1-4 scans per subject and has a total of 299 scans. The acquisitions that contain occlusion variations form the probe set, which consists of 381 images of four different types of occlusions. In Figure 4, the four types of occlusions 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. For the classification scenario, we have constructed an experimental setup where all neutral images constitute the gallery set, whereas the occlusion scans form the probe set. The gallery and probe sets contain a total of 299 and 381 scans, respectively.

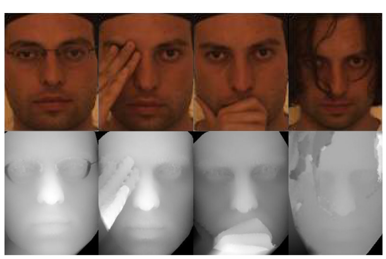

Figure 4. Four occlusion types in the Bosphorus database.

For the construction of the average face & nose models and for the training of the LDA method, we have utilized a separate dataset: The Face Recognition Grand Challenge version 2 (FRGCv2) [16]. FRGCv2 database consists of 4007 scans collected from 466 subjects. The scans are mostly frontal and contain facial expression variations. However, since the aim of this work is occlusion handling, we have utilized neutral subset of the FRGCv2 database. This subset consists of 2365 neutrals scans of 466 subjects.

7.2. Nose Detection Performance

We have located nose regions automatically in the gallery and probe set images of the Bosphorus 3D face database. By visual inspection of the detected regions, we see that all nose regions in the gallery images (without any occlusions) are successfully found. In the probe set that contains occlusions, six images out of 381 scans produced erroneous detections. In Figure 5, detection examples are illustrated on WSI maps: In the first row, some examples of challenging occluded scans are given where the nose region is accurately detected. As can be seen from these difficult examples, even though occluding objects such as hands and hair cover the borders of nose structure, it can still be localized sufficiently well. In the second row, all the erroneous detections are shown. As the erroneous examples show, the nose detection fails when the fork-shaped structure is not completely visible or the whole nose is covered by the occluding objects. We have also applied the nose detection algorithm to the FRGCv2 neutral subset which is required to train LDA classifiers. In the FRGCv2 database, only 2 out of 2365 scans have problematic detections, as illustrated in the last row of Figure 5. It can be seen that these errors are due to scale variations and acquisition errors, causing problematic detection of the nose area. Nevertheless, these problematic detections are not severe and can be corrected by the ICP method in the fine registration phase.

Figure 5. Nose detection results. (1) First row, correct detections on the Bosphorus database; (2) second row, erroneous detections on the Bosphorus database; and (3) third row, problematic detections on the FRGCv2.

7.3. Registration Performance

The second step of our registration approach is to perform ICP-based fine registration after coarse nose localization. To evaluate the performance of our automatic registration system, we have also constructed manual registration experiments where ground truth landmark locations are used for ICP initialization. In manual registration experiments, initial alignment is handled by the Procrustes Analysis [9] of five manually labeled landmark points around the nose area. The utilized landmark points are the nose tip, inner eye corners, and nose corners. However, some of these landmark points are not present in the ground truth set due to occlusions. In order to deal with the missing landmark problem, we employ pGPCA to estimate the unknown landmark positions, where pGPCA model of landmark coordinates is estimated from manual landmark positions of the FRGCv2 neutral subset. Using this approach, we are able to estimate the locations of one or two missing landmark points from known landmark coordinates.

In Figure 6, the first image illustrates the average face model together with the five landmark points. The last four images visualize the results of missing landmark estimation: the ground truth landmarks are visualized with green labels, whereas the estimated ones are shown in red.

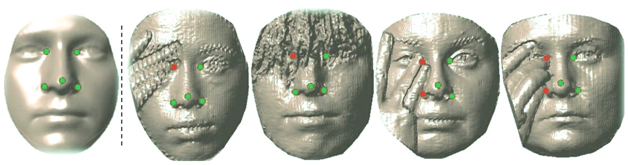

Figure 6. Manual landmark points: First image shows the average face model with five landmark points. The last four images are landmark estimation examples (ground truth and estimated points are shown with green and red labels, respectively).

For the performance evaluation of the proposed pGPCA-based landmark estimator, we have constructed landmark estimation experiments on the Bosphorus gallery set, which consists of 299 neutral scans with complete manual landmarks[2]. The gallery set is divided into two random groups, where 250 scans constitute the training set to train the PCA space and the remaining 49 scans form the test set. For each scan of the test set, we randomly selected a maximum number of two landmark points as missing and estimated them using pGPCA. Then, the Euclidean error between the estimated and the original landmarks are computed in 3D. For performance evaluation, four experimental setups are considered: Either one or two landmarks can be missing, and the nose tip point can appear in the missing set or not. Each of the four experiments are performed in 10 folds, where for each fold, the gallery is separated into training and test sets randomly. In Table 1, the mean Euclidean error values are provided. The ratio of the mean Euclidean distance to the average interocular distance (71.88mm) is also provided in the last column. As these results indicate, the missing landmarks can be estimated with sufficient accuracy. If the nose tip is visible, missing landmarks are estimated more accurately. Furthermore, the estimation performance is higher if fewer landmarks are missing.

[2]It should be noted that we use Bosphorus database here only for performance evaluation. In our actual system where we perform recognition experiments, FRGCv2 database is used to learn pGPCA model for landmark estimation

Table 1. pGPCA-based landmark estimation performance

Missing Landmark Count	Nose Tip Missing or Not	Average Euclidean Error (mm)	Error Ratio
1	Visible	5.56	7.73%
1	Missing	6.48	9.01%
2	Visible	6.85	9.52%
2	Missing	7.24	10.07%

To evaluate the performance of nose-based registration, we have constructed two experiments: (1) ICP-based registration with nose AvRM, (2) ICP-based registration with an average face model. In both of these experiments, manual landmark points are employed during the initial registration. The first experiment is constructed to evaluate the performance of the automatic nose detection. The second experiment provides a comparison between the use of a nose model in fine tuning, as opposed to using the whole facial surface. The experimental results for (1) automatic nose AvRM-based ICP, (2) manual nose AvRM-based ICP, and (3) manual average face model-based ICP are given in Table 2. In the classification phase, global depth-based $1-$nearest neighbor classifiers are employed. The average absolute distance (l_1-norm) between z coordinates is used as the dissimilarity measure in classification. In these experiments, identification is performed on occluded facial surfaces without any restoration or elimination.

Table 2. Comparison of identification results obtained for (1) automatic nose AvRM-based ICP, (2) manual nose AvRM-based ICP, and (3) manual average face model-based ICP.

Initialization	ICP Model	Recognition Rate (%)
Automatic	Nose	76.12
Manual	Nose	76.90
Manual	Whole Face	56.17

Comparing global versus nose-based registration, we see that nose-based ICP registration is superior to using the whole facial surface under occlusions: recognition rate improves from 56.12% to 76.90% with manual ICP initialization. This result validates the benefit of using nose region during the registration of occluded facial images. The second observation is related to the use of manual and automatic landmarking on the identification performance: We see that the performance decrease is quite small if landmarks are found automatically. With automatic landmarking, identification accuracy drops from 76.90% to 76.12%.

7.4. Impact of Occlusion Detection on Recognition

For the occlusion detection phase, we have utilized thresholding over the difference map between the original probe image and the mean face template. For comparison, we have manually labeled ground truth masks for the occluded areas. Identification results using depth image-based classifiers are given in Table 3, where the automatically detected and manually labeled occlusion masks are used to discard the regions respectively. Recognition performance of the original occluded faces using the global depth image-based classifier is also given for comparative purposes. Recognition rates are reported both for manual and automatic nose-based registration approaches.

As expected, occlusion-free faces where extraneous regions are manually removed (Table 3, Manual Masking) result in higher identification accuracy: rank-1 identification accuracy for depth-image based classifier with manually initialized registration improves from 76.90% to 83.73%. Similar improvement is also present for fully automatic registration. If we look at the identification accuracies obtained by automatic occlusion detection method (Table 3, Manual Masking), we see that they are very similar to the ones produced by manually labeled occlusion regions. For instance, with automatically registered facial surfaces, identification accuracy drops from 83.99% to 83.73% if occlusions are detected automatically. This finding reveals that our automatic occlusion detection method performs efficiently.

Table 3. Comparison of identification results (%) obtained by utilizing occlusion masks to discard occluded regions.

	Manual Registration	Automatic Registration
Occluded Face (No Masking)	76.90	76.12
Manual Masking (Ground Truth)	83.73	83.99
Automatic Masking	83.46	83.73

7.5. Impact of Restoration on Recognition

Here we provide the results of identification experiments where missing regions of the occlusion-free faces are restored using the pGPCA approach. For comparative purposes, we also provide the recognition performances of occlusion-free faces. In these experiments, the manually labeled occlusion masks are employed during the restoration process. In Table 4, identification results are given, where the surfaces are compared using the both global and local depth image-based classifiers. There are, in total, 30 local depth image-based classifiers, corresponding the regions shown in Figure 3. Their outputs are fused by the product rule.

If we look at the global classifier results in Table 4, we see that restoring missing regions actually decreases the identification rates for depth image-based classifiers. Similarly, although local classifiers are in general better than the global classifier, they achieve same identification accuracies for occlusion-free and restored faces. This means that occlusion-free faces contain more discriminatory information than faces with restored regions. In other words, restoration of missing regions might not always lead to an increase in person-specific surface characteristics. These results indicate that *occlusion removal without restoration* could also be a viable alternative for depth-based classifiers. However it should be emphasized that the more complex statistical classifiers such as LDA rely on restoration of occluded parts since they assume complete feature vectors. Identification results obtained by local LDA classifiers are given in the next section (Section 7.6).

Table 4. Identification accuracies with or without restoration for global and local depth image-based classifiers.

	Global Classifier	Local Classifier Fusion
Occlusion-free faces (no restoration)	83.99	85.04
With Restoration	81.63	85.04

7.6. Overall Face Recognition Performance: Fusion of Regional LDA Classifiers

Up to now, we have studied the effects of different registration, occlusion detection and restoration methods using depth image-based classifiers. In this section, we provide the comparative analysis of depth image-based classifiers and statistical classifiers that are based on LDA. To train the LDA based classifiers, a separate data set (FRGC v2 neutral subset) was used in order to get unbiased results. LDA projections are obtained on the restored surfaces. In Table 5, fusion results of both depth image-based and LDA classifiers are given where 30 local classifiers are fused by the product rule. For both of these classification methods, we provide identification results where the restoration is carried out using i) manually and ii) automatically found occlusion regions.

Table 5. Fusion performances (%) of depth image-based and LDA-based classifiers.

	Occlusion Masks	
Classifier	Ground Truth	Automatic
Depth Image-based	85.04	83.20
LDA-based	95.01	94.23

As the results presented in Table 5 show, using LDA-based classifiers improves the identification performance significantly. When ground truth masks are used to obtain the restored faces, this increase is from 85.04% to 95.01%. Similarly, with automatically detected occluded regions, performance boost is substantial: Our proposed scheme with automatic registration, occlusion detection and restoration modules, and fusion of LDA-based local classifiers achieves 94.23% rank-1 identification accuracy. Face recognition experiments were held on a 64-bit Core i7 2.67GHz PC with 12GB RAM. The timing details for processing a single test face using unoptimized MATLAB codes are as follows: Detecting nose takes approximately 10 seconds, whereas registration to the nose model takes about 8 seconds. Rest of the methods (occlusion detection, restoration and classification) are quite fast i.e., on the order of milliseconds.

8. Conclusion

In this work, we have proposed a 3D face recognition system which is robust to occlusions. For each aspect of the proposed system, namely *occlusion resistant registration*, *occluded area detection*, *restoration of original facial structures* and *the use of statistical regional classifiers*, we presented a detailed comparative analysis and measured their influence on the final recognition performance. To be able to reach valid conclusions that reflect real-world behavior of a 3D face recognizer, we have experimented on a 3D face database that contains realistic facial occlusions. We have shown that, with extreme occlusions covering significant amount of a facial surface, a specialized facial registration scheme should be employed. A popular 3D facial registration method, holistic ICP method, has limitations and can not handle erroneous regions during alignment. Therefore, we proposed to use regional nose-based alignment that is robust to outlier regions and achieve similar registration performance when compared to manual registration. We have also studied two possible scenarios after detecting the occlusions: 1) discarding extraneous regions to obtain occlusion-free faces and use them directly in classification, or 2) perform restoration on occlusion-free faces to estimate original facial structure. Our experiments reveal that classifiers that can operate on incomplete/missing facial features (i.e., depth image-based classifiers) attain better identification accuracies on occlusion-free faces. This finding reveals that restoration via a PCA-like model does not add discriminatory information useful for classification. However, with a complex statistical classifier (i.e., LDA) that requires complete feature vectors, reconstruction proves to be very beneficial. Therefore, we conclude that depending on the choice of classifier, you can choose to use occlusion-free or reconstructed faces.

References

[1] N. Alyuz, B. Gokberk, and L. Akarun. A 3D face recognition system for expression and occlusion invariance. In *2nd IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–7. IEEE, 2008.

[2] N. Alyuz, B. Gokberk, L. Spreeuwers, R. Veldhuis, and L. Akarun. Occlusion-robust 3D face recognition using restoration and local classifiers. In *Signal Processing and Communications Applications (SIU), 2011 IEEE 19th Conference on*, pages 750–753. IEEE, 2011.

[3] P. J. Besl and H. D. McKay. A method for registration of 3D shapes. *IEEE Trans. PAMI*, 14(2):239–256, 1992.

[4] K. Chang, W. Bowyer, and P. Flynn. Multiple nose region matching for 3d face recognition under varying facial expression. *IEEE Trans. PAMI*, 28(10):1695 –1700, oct. 2006.

[5] A. Colombo, C. Cusano, and R. Schettini. 3d face detection using curvature analysis. *Pattern Recogn.*, 39:444–455, March 2006.

[6] A. Colombo, C. Cusano, and R. Schettini. Detection and restoration of occlusions for 3D face recognition. In *2006 IEEE International Conference on Multimedia and Expo*, pages 1541–1544, 2006.

[7] A. Colombo, C. Cusano, and R. Schettini. Gappy PCA Classification for Occlusion Tolerant 3D Face Detection. *Journal of Mathematical Imaging and Vision*, 35(3):193–207, 2009.

[8] R. Everson and L. Sirovich. Karhunen–Loeve procedure for gappy data. *Journal of the Optical Society of America A*, 12(8):1657–1664, 1995.

[9] C. Goodall. Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society, Series B (Methodological)*, pages 285–339, 1991.

[10] 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. PAMI*, pages 1977–1981, 2005.

[11] J. Koenderink and A. van Doorn. Surface shape and curvature scales. *Image and vision computing*, 10(8):557–564, 1992.

[12] T. Lo and J. Siebert. Sift keypoint descriptors for range image analysis. *Annals of the BMVA X*, pages 1–18, 2009.

[13] X. Lu, A. Jain, and D. Colbry. Matching 2.5d face scans to 3d models. *IEEE Trans. PAMI*, 28(1):31 –43, jan. 2006.

[14] A. Martınez. Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *IEEE Trans. PAMI*, 24(6):748–763, 2002.

[15] J. Park, Y. Oh, S. Ahn, and S. Lee. Glasses removal from facial image using recursive error compensation. *IEEE Trans. PAMI*, pages 805–811, 2005.

[16] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 947–954. IEEE, 2005.

[17] P. Phillips, W. Scruggs, A. O'Toole, P. Flynn, K. Bowyer, C. Schott, and M. Sharpe. FRVT 2006 and ICE 2006 large-scale results. *National Institute of Standards and Technology, NISTIR*, 2007.

[18] A. A. Salah, N. Alyuz, and L. Akarun. Registration of 3D face scans with average face models. *Journal of Electronic Imaging*, 17(1), 2008.

[19] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun. Bosphorus database for 3D face analysis. *Biometrics and Identity Management*, pages 47–56, 2008.

[20] L. Spreeuwers. Fast and Accurate 3D Face Recognition. *International Journal of Computer Vision*, pages 1–26, 2011.

[21] F. Tarres, A. Rama, and L. Torres. A novel method for face recognition under partial occlusion or facial expression variations. In *ELMAR 47th International Symposium*, pages 163–166, 2005.

[22] J. Tittle and V. Perotti. The perception of shape and curvedness from binocular stereopsis and structure from motion. *Attention, Perception, & Psychophysics*, 59(8):1167–1179, 1997.