An Expression Deformation Approach to Non-rigid 3D Face Recognition

F. Al-Osaimi · M. Bennamoun · A. Mian

Abstract The accuracy of non-rigid 3D face recognition approaches is highly influenced by their capacity to differentiate between the deformations caused by facial expressions from the distinctive geometric attributes that uniquely characterize a 3D face, interpersonal disparities. We present an automatic 3D face recognition approach which can accurately differentiate between expression deformations and interpersonal disparities and hence recognize faces under any facial expression. The patterns of expression deformations are first learnt from training data in PCA eigenvectors. These patterns are then used to morph out the expression deformations. Similarity measures are extracted by matching the morphed 3D faces. PCA is performed in such a way it models only the facial expressions leaving out the interpersonal disparities. The approach was applied on the FRGC v2.0 dataset and superior recognition performance was achieved. The verification rates at 0.001 FAR were 98.35% and 97.73% for scans under neutral and non-neutral expressions, respectively.

F. Al-Osaimi (✉) · M. Bennamoun · A. Mian
School of Computer Science and Software Engineering, The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia
e-mail: faisal@csse.uwa.edu.au

M. Bennamoun
e-mail: bennamou@csse.uwa.edu.au

A. Mian
e-mail: ajmal@csse.uwa.edu.au

1 Introduction

Face recognition is highly desirable (due to its non-intrusive nature) in a wide range of applications including surveillance, access control and machine-human interaction. High recognition accuracy is crucial for the applicability of face recognition in most of its application areas, for example security related applications such as fraud prevention. One of the main objectives of research in face recognition is to increase its accuracy as the accuracies of the current face recognition technologies fall short before the requirements of many applications.

2D face recognition has been extensively researched in the last two decades. However, unlike 3D face recognition its accuracy is adversely affected by many factors such as illumination and scale variations. In addition, 2D images undergoes affine transformations during acquisition. Moreover, handling pose variations in 3D scans is more feasible than 2D images. It is believed that 3D face recognition has the potential for more accuracy than 2D face recognition (Bowyer et al. 2006). On the other hand, the acquisition of 2D images is less intrusive than 3D acquisition. However, 3D acquisition technologies are continuously becoming cheaper and less intrusive (The International Conference on 3D Digital Imaging and Modeling, 1997–2007).

Facial expression is one of the greatest challenges to face recognition. The geometry of the human face can drastically deform under facial expressions in a complex way (as they are driven by highly complex muscle mechanism), leading to deterioration in recognition accuracy. This challenge arises from the difficulty in differentiating between expression deformations and interpersonal disparities. Since expression deformations and the 3D shape of a face are more accurately captured by its 3D scans, this challenge can be

better addressed in the 3D domain compared to the 2D domain.

Approaches to 3D face recognition that address the expression challenge can be classified into two broad categories; rigid (Chang et al. 2006; Faltemier et al. 2007; Mian et al. 2007) and non-rigid (Bronstein et al. 2005; Lu and Jain 2008; Kakadiaris et al. 2007). The rigid approaches treat the human face as a rigid object while the non-rigid approaches apply deformations to the 3D facial scans to counteract expression deformations or to reduce their influence on the recognition performance. Most of the approaches in the literature are rigid. Despite the fact that currently high performance approaches can be found in both categories, the non-rigid approaches are more promising in handling the expression challenge because it is possible to robustly model facial expression deformations (Lu and Jain 2008) and consequently extract more information from the face. Robust modeling of facial expressions is possible because these deformations follow patterns governed by the underling anatomy of the human face.

The main contribution of this paper is a non-rigid 3D face recognition approach. This approach robustly models the expression patterns of the human face and applies the model to morph out facial expressions from a 3D scan of a probe face before matching. Robust expression modeling and subsequent morphing gives our approach a better ability in differentiating between expression deformations and interpersonal disparities. Consequently, more interpersonal disparities are preserved for the matching stage leading to better recognition performance.

The paper is organized as follows. Related literature is reviewed in Sect. 1.1. Then an overview of the approach is presented in Sect. 1.2. Section 2 describes the Expression Deformation Model and the non-rigid matching approach. Section 3 describes several experiments and presents their results. Conclusions are drawn in Sect. 4.

1.1 Related Work

Many approaches to 3D face recognition have been proposed e.g (Lu et al. 2006; Samir et al. 2006; Gokberk et al. 2005; Hesher et al. 2003; Pan et al. 2005). Recent and comprehensive reviews of 3D face recognition literature can be found in the surveys by Bowyer et al. (2006) and Zhao et al. (2003). This section reviews prominent rigid and non-rigid approaches that address the expression challenge.

Rigid matching of sub-facial regions with least expression deformations has shown higher recognition performance under expressions than when the whole face is considered. In the work by Mian et al. (2007), the nose and the forehead regions (semi-rigid regions) were separately matched in a rigid fashion then their scores were fused.

Matching multiple overlapping regions around the nose was investigated by Chang et al. (2006). In fact, expression deformations remain a source of error as all the regions of the human face deform with expressions. Another rigid approach is to enroll multiple gallery scans under different facial expressions (Faltemier et al. 2007). Since there is an unlimited number of facial expressions, that approach is still prone to errors when the expression of a probe matches that of a gallery scan belonging to a different subject, adding to that the increased computational complexity of a larger gallery. In the work by Chua et al. (2000) point signatures are used to recognize 3D facial scans under expressions. To avoid adverse effects of expressions, only point signatures from the upper part of the face are matched. Li et al. (2007) extract multiple geometric attributes (features) like the geodesic distances and curvatures and adapt them to expression variations by combining the geometric attributes using different weights. Chang et al. (2005a) devised an approach for selection of multiple rigid regions. They believe that there are some regions that remain relatively rigid between any two 3D facial scans under different expressions.

Non-rigid approaches have been applied to both 2D and 3D face recognition. Examples of 2D non-rigid approaches are as follow. Waters' face animation model (Parke and Waters 1996) which is based on muscle-models is used by Lu et al. (2004) to synthesize 2D images with various expression variations. The use of the animation model provided large training data. Then an affine subspace is constructed for each gallery face and probe image. The subspaces are then matched for recognition. The CANDIDE animation approach (Rydfalk 1987) is used to convert non-neutral 2D facial images to neutral ones in a preprocessing stage then recognition is performed on the processed images (Ramachandran et al. 2005). The well-known 3D Morphable Model (Blanz and Vetter 2003) uses statistical modeling to estimate the 3D shape and texture from a single 2D image. The 3D Morphable Model is used to reanimate a 2D face by varying the texture and shape parameters (Blanz et al. 2003; Blanz and Vetter 1999). The 3D morphable model recognizes a 2D face across different poses and illumination variations given that the images have similar expressions (Blanz and Vetter 2003). A common feature of the non-rigid approaches that adopt animation models is that they tend to handle expression deformations in a preprocessing stage and/or by modeling their variations followed by matching. Such approaches have the inherent drawback that while removing expression deformations, they may introduce artifacts leading to lower recognition accuracy. On the other hand, other approaches usually formulate face recognition as a fitting problem between a probe and gallery faces. Since the starting and final 3D scans (which might be under different expressions) are always known (given) for the probe

and gallery, these approaches are more immune to drift errors.

Among the non-rigid approaches in 3D is the approach by Bronstein et al. (2005) who demonstrate that facial expressions can be modeled as isometries of the facial surface. The surface of a 3D face under various facial expressions is deformed (to a Canonical Form) by embedding the surface into a multi-dimensional space using Multi-dimensional Scaling (MDS) while largely preserving the geodesic distances among the points of the 3D facial surface (isometric embedding). This preserves the intrinsic geometries of the face and gets rid of the extrinsic geometries. Thus an expression invariant representation of the face is achieved as the facial expressions are mainly extrinsic geometries. Finally, the Canonical Forms are rigidly matched against each other. In an extension of that work, Bronstein et al. (2006a, 2006b) isometrically embed two 3D face surfaces into one another (called Generalized Multi-dimensional Scaling, GDMS). The embedding error is used as a similarity measure between the two surfaces as it is zero for isometric surfaces (and low for matching surfaces), avoiding the need for computation of Canonical Forms. The GMDS has better flexibility in handling partial surface matching than the MDS approach. In addition, the GMDS makes use of some metric distortion that serves as a dissimilarity measure resulting in a small number of surface samples for accurate matching. Thin Plate Spline (TPS) is used to learn expression deformations from a control group of neutral and non-neutral scans (Lu and Jain 2008). Then, the learned deformations along with ICP (Chen and Medioni 1991; Besl and McKay 1992) are used iteratively to fit probes to gallery scans. Improved performance over rigid ICP was reported. The deformations are learned actually from sparse landmarks on the face which do not cover the whole face. In the work by Kakadiaris et al. (2007) and Passalis et al. (2007), the 3D face is parametrized using the Annotated Face Model (AFM) followed by elastic deformations. Significant recognition accuracy was reported for that approach. In the work by Wang et al. (2007), fields are applied to deform the non-neutral 3D scans to neutral ones. Constraints are applied to avoid overfitting. Although a performance improvements was also reported for that approach, some interpersonal disparities may be morphed as well since the 3D face scans are not deformed according to expression patterns. Our approach also falls into this category i.e. non-rigid 3D face recognition.

1.2 Overview

In our approach, expression deformations are learned in a PCA subspace, called Expression Deformation Model (EDM). The PCA subspace is built from shape residues between non-neutral and neutral scan pairs (training data), each pair belongs to the same subject. Initially, the scan pairs are cropped and pose-corrected. Then the scans of each pair are finely registered according to the semi-rigid region of the face (the forehead and nose) to get more consistent registration over the various facial expressions. As PCA is applied only on expression deformations, it models the facial expression deformations (the shapes of the top eigenvectors encode the patterns of the expression deformations) while it lacks the capacity to model the human face itself. In the recognition stage, the shape residues between the probe and gallery scans are found in the same way. One of these shape residues is mainly expression deformations, the one from the matching probe and gallery scans (same subject).

The residues are minimized using the PCA subspace (expressions are morphed out). The projection of the residues on the PCA subspace retains expression deformations and lose interpersonal disparities. A non-orthonormal projection that accounts for the artifacts in the learnt expression patterns and outliers in the residues was devised (see Sects. 2.4.1 and 2.4.2). The non-orthonormal projection also limits excessive morphing by applying a cost proportional to the extent an eigenvector is used. The minimized residues are then reconstructed. Finally, the similarity measures are extracted from the errors between the original and the minimized residues (see Fig. 1).

PCA is a well-known dimensionality reduction tool. It has been used by Kirby and Sirovich (1987, 1990) then by Turk and Pentland (1991) for compact 2D face representation and similarly applied to 3D face recognition (Chang et al. 2005b). In the work by Russ et al. (2006), a 3D face is registered to a generic 3D face using ICP. PCA was then applied to the registered faces. It is known that PCA based approaches are sensitive to expression variations as they merely represent the higher dimensional face data by a compact eigen-coefficient representation which is still adversely affected by expressions. Note that our approach does not use PCA merely as a means of dimensionality reduction but also for splitting the higher dimensional data (3D face residues) into an expression part (preserved during projection) and a non-expression part (lost during projection). In our approach, the emphasis is on the eigenvectors rather than the eigen-coefficients.

Fig. 1 Block diagram showing how the expression deformation model is built

2 Details of Non-rigid 3D Face Recognition

As PCA is used for modeling the expression deformations, the 3D facial scans need to be converted into a suitable range image representation. In addition, PCA requires that the 3D scans are pose-corrected and cropped. In contrast, the pointcloud representation suits the ICP. At different stages, the approach switches between the two representations. This section details the two main modules of the approach, the Expression Deformation Model and the nonrigid face matching. It also describes the preprocessing of the data and the automatic pose-correction and cropping of the 3D facial scans.

2.1 Data Preprocessesing

The approach was applied on the frontal facial scans of the FRGC v.20 dataset (Phillips et al. 2005) which were acquired using Minolta vivid 3D scanner. The 3D scans are dense pointclouds and also contain spikes and holes. The 3D scans are decimated to reduce the computational complexity of the approach. The spikes are removed and the holes are filled.

The dense pointclouds of the FRGC dataset (Phillips et al. 2005) are in the form of three matrices x, y and z, the three coordinates of the pointclouds are in mm. The spikes are removed by dropping the outlier points from the three matrices based on local statistics. The matrices are then smoothed using a mean filter which averages the non-missing points in the local neighborhood (neglects the missing points). The three matrices are bi-cubically interpolated at the holes. The bi-cubic interpolation is used here because it can accurately fill in the holes which are to some extent large such as the eye-brows which are occasionally missing in the pointclouds.

The range image is then computed from the three matrices by interpolating at the integral x and y coordinates and storing the corresponding z coordinates in the range image matrix using x as a horizontal index and y as a vertical index. This step also decimates the range image as the points become equally spaced at a 1 mm distance along x and y axes as opposed to the irregular sub-millimeter accuracy of the raw data. Finally, the range image is smoothed using a Gaussian filter.

2.2 Automatic Cropping and Pose Correction

The facial scans are cropped and their poses are corrected to frontal view using the approach by Mian et al. (2006). That approach gives sufficient cropping and pose correction accuracy such that it can be used for modeling expression deformations using PCA and initial coarse registration for the ICP

algorithm. The cropping and pose-correction are also important for the coarse localization of facial regions by means of fixed binary mask images.

The tip of the nose is initially detected by slicing the 3D facial scan horizontally and searching the slices for the point which has maximum distance to the line segment defined by the intersection of the slice and a sphere centered at that point. Once the tip of the nose is detected the 3D facial scan is cropped around the nose tip by dropping the points which are at a distance more than 80 mm from the tip of the nose. The cropped points are then pose corrected according to their principal directions. The holes which might appear due to self-occlusion are filled using bi-cubic interpolation. The last two steps are iterated until the pose stabilizes. The facial scan is shifted so that nose tip is at the center of the range image and its depth (z coordinate) is zero.

2.3 Expression Deformation Model

The expression deformations of the human face is modeled by PCA subspace. PCA is chosen for modeling expression deformations of the human face for two reasons. The expression deformations have similar patterns, consequently they are expected to reside in a lower dimensional PCA subspace. Deforming a 3D face using PCA is computationally efficient as it can be achieved using a closed form solution (see Sect. 2.4). The Expression Deformation Model is chosen to be generic to the human face so the approach can perform non-rigid face recognition based on only a single gallery enrollment per subject.

The generic Expression Deformation Model is trained using non-neutral scans of different people. Each non-neutral scan is paired with a neutral scan of the same subject. The training data contains multiple non-neutral scans per subject but one neutral scan per subject is sufficient. Ideally, it should contain a wide range of facial expressions and sufficient number of subjects with balanced gender and race ratios.

The training scan pairs are used to generate the Expression Deformation Model according to the following steps. The training scan pairs are cropped and pose-corrected as described in Sect. 2.2. Then the scan pairs are finely registered using the ICP algorithm according to the forehead and nose region. Next, the shape residues of the finely registered scan pairs are computed, each scan pair produces a shape residue. Finally, PCA is applied on the shape residues.

The fine registration step is needed here mainly for the computation of accurate shape residues. As the scan pairs are pose-corrected with respect to the principal directions of the whole cropped face, the scan pairs can have pose errors due to the fact that the principal directions can slightly vary with expression variations. Pose-correction errors in the range of $\pm 2°$, $\pm 6°$ and $\pm 1.66°$ respectively around x, y and

z axes of the principal direction method are reported in Mian et al. (2006). Such error ranges can affect the accuracy of the shape residues. However, their projections on the $x-y$ plane are not largely affected in such error ranges. Fine registration according to the forehead and nose region produces accurate shape residues as this region is the least affected by facial expressions. For example, the global shape of the forehead is largely defined by the forehead bone more than the muscle movements. When the eyebrows are raised, the skin of the forehead slides over the skull bone but the global shape of the forehead largely remains unchanged except locally near the eyebrows. Small local variations of the 3D surfaces do not tangibly affect the registration accuracy of the ICP. Some previous papers performed 3D face recognition relying solely on the semi rigid regions of the face, e.g. Mian et al. (2007). Furthermore, the shape of that region is also appropriate for correct ICP convergence. The forehead and nose regions from two 3D scans of the same subject grove into each other leading to a 6DOF accurate registration.

The shape residues are computed from the finely registered scan pairs simply by taking the difference between the non-neutral and the neutral scans. Because each training scan pair belongs to the same subject, the shape residues are expected to be rich in information that is relevant to expression deformations while they are low in person-specific facial shape information. Although the shape residues of a given facial expression are generally similar for different people, variations can take place. For that reason the training data has scan pairs belonging to a sufficiently large number of people with diversity in gender and race. Based on these properties of the training data, the application of PCA on the shape residues models generic expression deformations of the human face while it lacks the capacity to model the shape of the human face itself. Consequently, when the model is used in face matching (Sect. 2.4), the expression deformations can be removed (morphed out) without overfitting the 3D facial scans of different people.

2.3.1 Fine Scan Pair Registration

The region of the forehead and nose in one scan of the pair is determined using the binary mask shown in Fig. 2 (top left). Given that the scan pair is cropped, pose-corrected and the nose tip is brought to the center of the range image (as described in Sect. 2.2), the shape of that mask covers substantial parts of the forehead and the whole nose of every face. As the two facial scans are registered according to subregion, the points which are far away from the reference region are considered to be redundant. The redundancy is removed by applying a dilated version of the first mask to the other scan.

Then the two masked regions are converted from the range image to pointcloud representation (list of 3D tuples).

Fig. 2 3D facial scan pairs registered according to the forehead and nose region of the face (semi-rigid region). The 1st row shows the binary masks which were used to detect the region in the pose corrected range images and an example range image. The solid regions in the scan pairs are registered using ICP and the same rigid transformations are applied to the other regions of the face

Each range pixel produces an (x, y, z) tuple, where x and y are the horizontal and vertical indices of the pixel respectively and z is its value (depth). The pointclouds from the first and second facial scans are called $\mathcal{P}$ and $\mathcal{Q}$ (the one with the extended region), respectively. The ICP algorithm is used next to register the two pointclouds. In the ICP algorithm the closest point search is performed for the points in $\mathcal{P}$ (i.e. the searched pointcloud is $\mathcal{Q}$). Because $\mathcal{Q}$ covers larger region, it is more likely to find proper closest-points for the points near the mask edges in $\mathcal{P}$. Consequently, the accuracy of the registration is not affected by the edges. In the ICP algorithm, $\mathcal{P}$ is iteratively transformed until it is registered to $\mathcal{Q}$, denoted as $\mathcal{R}$. The total rigid transformations

which are equivalent to the rigid transformations in all iterations are computed. The total rotation matrix and translation vector are called respectively $\mathbf{R_t}$ and $\mathbf{t_t}$.

The remaining points of the first facial scan (which are in the unmasked region) are also converted to pointcloud representation ($\mathcal{P}'$) and then the total rigid transformations are applied to them according to (1).

$$\mathcal{R}'_i = \mathbf{R_t}\mathcal{P}'_i + \mathbf{t_t} \tag{1}$$

The registered pointclouds of the masked and unmasked regions ($\mathcal{R}$ and $\mathcal{R}'$) are combined in one pointcloud representing the whole first face, $\mathcal{F} = \mathcal{R} \cup \mathcal{R}'$, and converted to range image representation.

The pointcloud ($\mathcal{F}$) is converted to range a image representation as follows. The x and y coordinates of the pointcloud tuples are rounded to the nearest integers. Then their z coordinates are stored in the pixels of the range image indexed by the rounded x and y coordinates. If a pixel is indexed by more than one tuple the one with the maximum z is stored in the pixel (deemed self-occluded). Since the scan pairs are initially pose-corrected, the fine registration is not expected to cause significant holes or self-occluded pixels. A line of missing data in the range image (crack) might appear. The range at the crack is linearly interpolated. This conversion approach is computationally cheap. The rounding errors are intangible if the approach is not used iteratively (which is the case).

Figure 2 shows examples of scan pairs finely registered according to the forehead and nose region. It is noticed from the figure that the registration is consistent despite expression variations.

2.3.2 PCA on Shape Residues

Let $\mathbf{I}_{e_i}$ and $\mathbf{I}_{n_i}$ denote respectively the finely registered non-neutral and neutral range images of the i-th training scan pair. The i-th shape residue is the difference between the two range images as in (2).

$$\mathbf{R}_i = \mathbf{I}_{e_i} - \mathbf{I}_{n_i} \tag{2}$$

PCA (Jollife 1986) is applied to the shape residues and a lower dimensional PCA subspace is found. The shape residues $\mathbf{R}_i$ are vectorized into $m \times 1$ vectors $\mathbf{r_i}$. The eigenvectors $\mathbf{e_1}, \dots, \mathbf{e_k}$ with the top k eigenvalues of the covariance matrix Ω (3) are found, where $k \ll m$.

$$\Omega = \sum_{i=1}^{n} (\mathbf{r_i} - \bar{\mathbf{r}})(\mathbf{r_i} - \bar{\mathbf{r}})^{\top} \tag{3}$$

Where $\bar{\mathbf{r}}$ is the average shape residue and n is their number. The subspace spanned by the k eigenvectors is represented by a matrix $\mathbf{E}$ in which the columns are the eigenvectors.

Fig. 3 Some eigenvectors which span the shape residue PCA subspace (our expression deformation model). The images in the *top row* are the 1st, 2nd and 3rd eigenvectors, respectively *from the left*. The 9th, 10th and 11th are in the *middle row* and the 25th, 26th and 27th are in the *bottom row*. Notice that the top eigenvectors are more relevant to the largely deformable regions (e.g. the mouth) and the deformations in the other regions gradually appear in the lower eigenvectors

Some of the eigenvectors of the matrix Ω are shown in Fig. 3. From their shapes, the relevance to facial expressions can be perceived to some extent. The eigenvectors with higher eigenvalues are largely related to the large expression deformations. For example, it can be seen in the figure that the first eigenvectors are largely relevant to the open mouth expression and those which involve large cheek deformations. As the eigenvalue decreases, smaller expression deformations appears in the eigenvector. For example, the deformations at the eyebrows, eyes, nose and forehead appear clearly in the eigenvectors with the 9th eigenvalue and beyond.

Another aspect of the eigenvectors is that the deformations appearing in one facial region are generally in accord with the other deformations in the other facial regions. For example, the first eigenvectors shows that opening the mouth is concurring with lowered cheeks as they are stretched due to the lowering of the jaw. In the second eigenvector the mouth is stretched to the sides and at the same time, the outer sides of the cheeks pop up (due to displacement) and also the sides of the nose are stretched.

To test how the expression model can model unseen facial expressions, novel shape residues ($\mathbf{r}_{n_{1,\dots,n}}$) are projected on the subspace $\mathbf{E}$ according to the following equation

$$\mathbf{r}_{p_i} = \mathbf{E}(\mathbf{E}^\top \mathbf{E})^{-1} \mathbf{E}^\top \mathbf{r}_{n_i} \qquad (4)$$

The projection of unseen shape residues on the subspace retains most of the shape residues if they are computed from the same subject. When projecting shape residues computed from facial scans that belong to different people (i.e. the difference between two different faces), the residues in this case are not merely expression deformations. These shape residues are expression deformations plus other shape disparities between the subjects. The projection of such shape residues on the subspace lose much more shape than the residues of the same subjects. The average ratio of retained shape (computed as the squared sum of $\mathbf{r}_p$) in the two cases is about 4 : 1. The projections of facial scans on the subspace are severely distorted, indicating that the subspace models the expression deformations but not the human face. We conclude that simply the projection of shape residues on the subspace can be used for non-rigid face matching (model fitting), see Sect. 2.4. The projection is computationally cheap in comparison to the iterative numerical minimization techniques used in model fitting.

In (4), the term $(\mathbf{E}^\top \mathbf{E})^{-1}$ equals to the identity matrix for projection on an orthonormal subspace. However, during the matching phase (Sect. 2.4), the matrix of the eignevectors $\mathbf{E}$ is modified to account for the artifacts at the borders of the eigenvectors (see Fig. 3). The matrix is further modified to cater for the outliers in the expression residues (Sect. 2.4.1) and to restrict the projections of the residues to appropriate regions of the subspace (Sect. 2.4.2). Consequently, after modification, $(\mathbf{E}^\top \mathbf{E})^{-1}$ does not equal to the identity matrix and the lower subspace no longer remains orthonormal. Empirical tests show that the average expression residue is insignificant and on that basis it was not included in (4). Tests also confirmed that the omission of the average residue from (4) did not affect the recognition accuracy.

2.4 Face Matching

Non-rigid 3D face matching of a probe $\mathbf{P}$ to a number of gallery 3D facial scans $\mathbf{G}_{1,\dots,n}$ is performed as follows. The probe p and the gallery scans are pose-corrected and cropped as described in Sect. 2.2. Then $\mathbf{P}$ is finely registered to each gallery scan $\mathbf{G}_i$, in a similar way to the registration of the training scan pairs (Sect. 2.3.1). From each registered probe and gallery, a shape residue $\mathbf{R}_i$ is computed. One of these shape residues represent purely expression deformations, the one in which the probe and the gallery scans belong to the same subject. The shape residues are then vectorized $\mathbf{r}_i$ and they are projected on the expression deformation subspace $\mathbf{r}_{p_i}$ according (4), except that the projection is modified to avoid the effects of the borders and the

Fig. 4 The images in the *first row* are the shape residues of the registered scan pairs shown in Fig. 2 followed by their projections on the PCA subspace (*second row*). Their projection on the PCA subspace retains most of the shape residue. The shape residues in the *third row* are not found from scan pairs belonging to the same subject. Instead, the non-neutral scans are registered to the neutral scans of different subjects which are respectively the 1st with the 2nd, the 2nd with the 3rd and the 3rd with the 1st. The shape residues in this case are not merely shape deformations. In this case, their projections on the PCA subspace lost more of the shape residue (4th *row*). The *last row* shows that the projections of the facial scans are not modeled by the PCA subspace as their projections are severely distorted

outliers in the data and also the projection is restricted to the portions of the subspace $\mathbf{E}$ in which realistic expression residues can exist (see Sects. 2.4.1 and 2.4.2). Negative sim-

ilarity measures are extracted from the error vector $\mathbf{e}_i$ between the shape residues and their projections on the expression deformation subspace s_i according to (5) and (6). The gallery scan with the minimum value of s_i is considered to be the probe's match. The elements of the error vector are checked against a threshold e_t and those which exceed e_t are truncated. This truncation helps in mitigating the effects of outliers (which might be caused by hair or any other error source) in $\mathbf{e}_i$ as the s_i will not be over-penalized when the identity of the probe is matching the gallery. A value of 3 mm is chosen for the threshold e_t. Let e_i^j denote the j-th element of the vector $\mathbf{e}_i$.

$$e_i^j = \begin{cases} r_i^j - r_{p_i}^j & \text{for } r_i^j - r_{p_i}^j \leq e_t \\ e_t & \text{for } r_i^j - r_{p_i}^j > e_t \end{cases} \tag{5}$$

$$s_i = e_i^\top e_i \tag{6}$$

In theory neither the probe scan nor the gallery ones need be neutral. In fact, the projection of the shape residue on the expression deformation subspace can morph out expression deformations between any two non-neutral expressions. The projection (4) is split into the following two equations.

$$\mathbf{c}_i = (\mathbf{E}^\top \mathbf{E})^{-1} \mathbf{E}^\top \mathbf{r}_i \tag{7}$$

$$\mathbf{r}_{p_i} = \mathbf{E}\mathbf{c}_i \tag{8}$$

Where $\mathbf{c}_i$ is a vector of eigen-coefficients. Equation (7) is the solution to the over-determined system, $\mathbf{r}_i = \mathbf{E}\mathbf{c}_i$. Then the projected shape is reconstructed according to (8). Let $\mathbf{r}_{ab}$ denote the shape residue between two scans under non-neutral expressions a and b. Suppose that there is a hypothetical neutral scan. Then morphing the facial expression from expression a to b is equivalent to morphing the facial expression from the neutral face to expression b, $\mathbf{E}\mathbf{c}_b$, minus the morph from the neutral expression to expression a, $\mathbf{E}\mathbf{c}_a$. See Fig. 5.

$$\mathbf{E}(\mathbf{E}^\top \mathbf{E})^{-1} \mathbf{E}^\top \mathbf{r}_{ab} = \mathbf{E}(\mathbf{c}_b - \mathbf{c}_a) \tag{9}$$

Figure 6 shows examples of 3D facial scans with expression morphing between non-neutral probes and gallery scans, $\mathbf{g} + \mathbf{r}_{ab}$.

2.4.1 Outlier Tolerant Projection

The solution of (7) is the least square minimization which can be sensitive to the outliers in $\mathbf{r}_i$. To avoid introducing outliers at non-overlapping regions at the borders, those elements are removed from $\mathbf{r}_i$ and their corresponding elements in all the columns of $\mathbf{E}$ are also removed (or simply they can be set to a value of zero). Also the artifacts at the borders of the eigenvectors which are caused by the
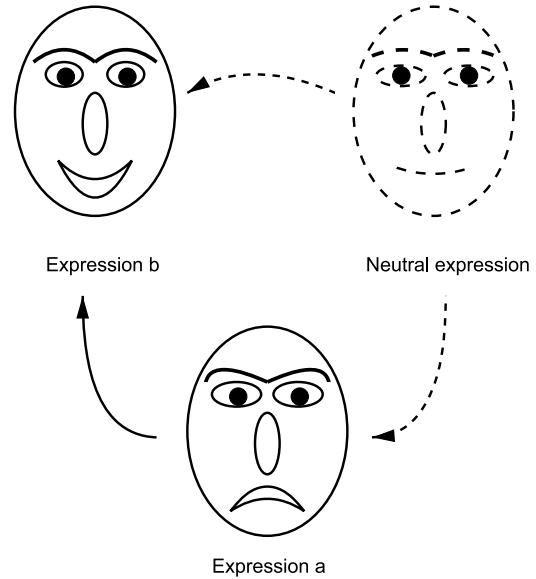

Fig. 5 Although the expression deformation model is trained using 3D facial scan pairs under non-neutral and neutral expressions, it can be used in morphing expressions in 3D facial scans that are both non-neutral. Thus the approach can match a non-neutral probe to a non-neutral gallery. In our model, morphing the expression in the non-neutral scans a to b is equivalent to expression morphing from a hypothetical neutral face (which is not in physically required) to b minus expression morphing from the neutral face to a

same reason (borders) are similarly removed and their corresponding elements in $\mathbf{r}_i$, producing a residue $\mathbf{r}_{b_i}$ and matrix $\mathbf{E}_b$. The borders are cropped out by means of a fixed binary mask. Any other elements which can be deemed as outliers such as those which unrealistically have high shape residue values are also removed, producing $\mathbf{r}_{bo_i}$ and $\mathbf{E}_{bo}$. Then the projection is computed using (10) and the similarity measure is extracted from $\mathbf{r}_{bp_i}$ and $\mathbf{r}_{b_i}$ according to (5) and (6).

$$\mathbf{r}_{bp_i} = \mathbf{E}_b (\mathbf{E}_{bo}^\top \mathbf{E}_{bo})^{-1} \mathbf{E}_{bo}^\top \mathbf{r}_{bo_i} \tag{10}$$

2.4.2 Restrictive Projection

The projection described in (10) has shown a high recognition performance. However, by imposing restrictions on the eigen-coefficients of the projection to their natural ranges, it was shown that the performance can be further improved. This is because not all the combinations of the eigenvectors represent realistic expression deformations. Consequently, the restriction reduces the chance of over-fitting the 3D faces belonging to different people.

To learn what are the natural ranges of each eigen-coefficient c_i, the training residues are projected onto $\mathbf{E}$. Then, the mean and the standard deviations of the eigen-coefficients are computed, μ_i and σ_i, respectively. An ad-

Fig. 6 These are examples of morphing expressions from non-neutral 3D facial scans (*left column*) to different non-neutral expressions in other scans (*middle column*). The 3D scans on the left are deformed according to the expression deformation model so that they better fit the middle scans (shown in the *right column*)

ditional cost proportional to the deviations from μ_i and inversely proportional to σ_i is added to the implied objective function of the projection (sum of least squares). It turned out that the mean values are close to zero, considering their corresponding standard deviations. For simplicity, we assumed they are zeros.

The $k \times k$ matrix (**M**) shown in (11) is concatenated to the end of the matrix $\mathbf{E}_{bo}$ in (10), giving a new matrix $\mathbf{E}_r = [\mathbf{E}_{bo}^\top \, \mathbf{M}]^\top$. The vector $\mathbf{r}_{bo_i}$ is padded with k zeros at the end,

giving $\mathbf{r}_r$ (k is the lower subspace dimension).

$$
\mathbf{M} = \begin{bmatrix} \frac{S}{\sigma_1} & 0 & \dots & 0 \\ 0 & \frac{S}{\sigma_2} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \frac{S}{\sigma_k} \end{bmatrix} \tag{11}
$$

S in the matrix is called the stiffness factor. The residue is then projected as in (12)

$$
\mathbf{r}_{rp_i} = \mathbf{E}_b (\mathbf{E}_r^\top \mathbf{E}_r)^{-1} \mathbf{E}_r^\top \mathbf{r}_{r_i} \tag{12}
$$

The total cost incurred by the restriction C is expressed by (13).

$$
C = \sum_{i=1}^{k} \left(\frac{S}{\sigma_i} c_i \right)^2 \tag{13}
$$

Where $\mathbf{c}_i$ is the i-th eigen-coefficient (see (7) and (8)).

The eigenvalues equivalently can be used instead of the standard deviations σ_i. However, using the standard deviations gives us some intuition about the determination of an appropriate value of the stiffness factor. Knowing that an average fitting cost (least square sum) of matched (same subject) probes and galleries is roughly about 15000. The stiffness factor S which makes the restriction cost C for eigen-coefficients $c_{i=1,\dots,k}$ that are deviated from their means (zeros in our case) by one standard deviation (k is assumed about 50) sums up to one third of the non-restricted average cost is about 10. By performing recognition using $S = \{2.5, 5, 10, 20 \text{ and } 40\}$, it was shown empirically that $S = 20$ achieves the best performance. Note that the proposed approach has a limited number of free parameters which ensures that it does not suffer from over-fitting. Moreover, it has been tested on the FRGC validation set as well as an additional dataset acquired in our laboratory. Our results in the next section indicate that the proposed approach generalizes.

3 Experiments

A number of experiments were conducted on the proposed approach. This section describes the experiments, the dataset on which the experiments were conducted and presents their results.

3.1 Dataset Description

The FRGC v2.0 is currently the largest publicly available 3D face dataset. It contains about 5000 3D facial scans under neutral and non-neutral expressions. The dataset is composed of two partitions: the training partition (943 scans) and the evaluation partition (4007 scans). The evaluation

partition scans belong to 466 subjects and contains about 2410 neutral scans and about 1597 that have various facial expressions.

The FRGC dataset was augmented by 3006 scans that were acquired using a Minolta vivid scanner in our laboratory.[1] The 3006 scans belong only to three subjects (1000 non-neutral scans and 2 neutral per subject). The non-neutral scans were acquired while the subject was talking or reading loudly and with the intention to produce as diverse facial expressions as possible. The FRGC dataset have scans for a large number of subjects but we also need a large number of non-neutral scans per subject for experiment no. 3 (Sect. 3.5). In addition, some of them are used in model training (Sect. 3.2) as it turned out that the expression deformation model requires large training data.

3.2 Model Training Data

The training partition of the FRGC was not sufficient for training our Deformation Expression Model. Insufficient training data can result in noisy eigenvectors (of the model, Sect. 2.3), especially those with lower eigenvalues. Also, a smaller training data may lack enough instances of facial expressions of different people. Consequently, the model may not perform optimally during face recognition.

To test how the size of the training data affects the performance of the deformation model, three Expression Deformation Models were trained using training data of 400, 800 and 1700 scan pairs. Then, they were used in non-rigid face recognition of 400 unseen probes under non-neutral expressions with an appropriate subspace dimension of the deformation model (the dimension is 55, see Sect. 3.3). The identification rates in the three cases were 89%, 93%, and 95%, respectively. The rates have increased for larger training data sizes.

In the following experiments, the Expression Deformation Model which gave best results (the one which is trained by 1700 pairs) was used as the generic deformation model. The 1700 pairs were formed among the training partition of the FRGC dataset (943 scans), 597 non-neutral scans from the evaluation partition (leaving about 1000 non-neutral scans for testing) and 500 scans from our acquired data. Although we had 2500 more non-neutral scans in our dataset, we used only 500 of them so that the deformation model is not biased toward the expressions of the 3 people in our dataset.

3.3 Experiment 1: Optimal Subspace Dimension

This experiment aims at finding empirically the most appropriate dimension of the subspace of the Expression Defor-
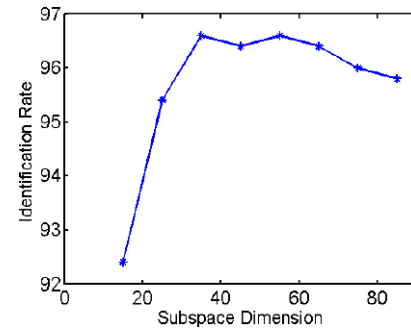

Fig. 7 Identification rate versus the dimension of the subspace in the deformation model

mation Model. The dimension of the deformation model influences the recognition performance positively in one way and negatively in another way. On one hand the higher the dimension of the subspace is, the more it is expected to model fine deformations (as it appears from the eigenvectors shown in Fig. 3 in Sect. 2.3.2). This factor pushes for a higher recognition performance. On the other hand, at higher dimensions there are more chances for over-fitting probe and gallery scans which belong to different subjects.

Non-Rigid 3D face matching was applied on 500 unseen probes under non-neutral expressions using deformation models that have subspace dimensions ranging from 15 to 85. Their first rank recognition rate (as shown in Fig. 7) has noticeably increased at the very low dimensions. Then it is almost leveled at its best performance at the range of dimensions from 35 to 65. For dimensions more than that, minor performance degradations are noticed when the dimension increases. In the following experiments, we use the deformation model with a dimension of 55.

3.4 Experiment 2: Performance on FRGC Dataset

One neutral scan per subject is enrolled in the gallery dataset (466 subjects). Then the remaining FRGC validation scans are used as probes. Non-rigid recognition is performed using both restricted ($S = 20$) and non-restricted projections (the model dimension is 55). The recognition performance for the unseen non-neutral probes and those which were used to augment the training data are found separately. The results in both cases are compared to those when non-rigid recognition is used (plain ICP, faces are registered according to the semi-rigid region of the face).

The scores in the similarity matrices are normalized row-wise to range from zero to one according to the following formula (14).

$$s_i' = \frac{s_i - s_{min}}{s_{max} - s_{min}} \tag{14}$$

where s_{min} and s_{max} are the minimum and maximum scores in the row of the similarity matrix and s_i and s_i' are the
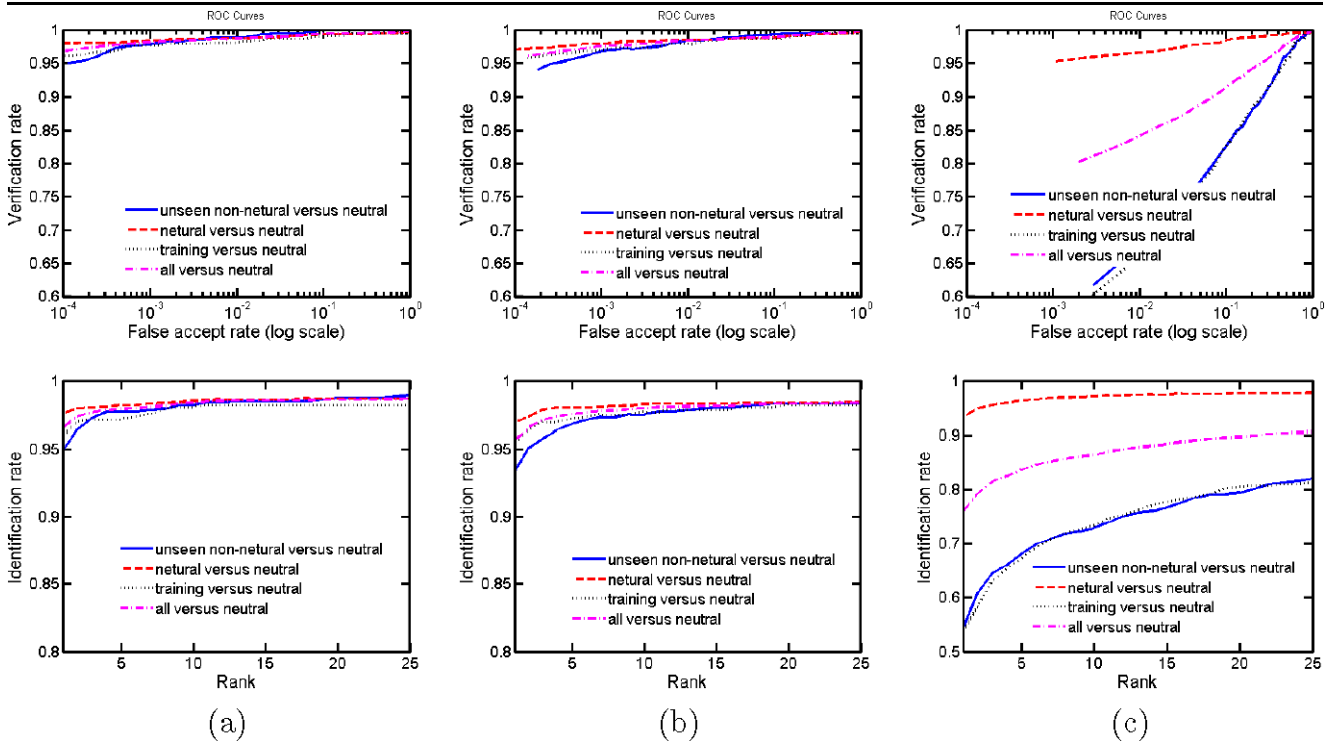
Fig. 8 The ROC and CMC curves of the non-rigid recognition approach on the FRGC dataset when restricted projection is used (**a**) and when it is not used (**b**). Our approach gave significantly higher recognition performance in comparison to the rigid ICP approach (**c**), in particular for facial scans under non-neutral expression

i-th non-normalized and normalized scores in the matrix row, respectively.

Figure 8 shows the ROC and Cumulative Matching Characteristic (CMC) curves for the three cases. The two non-rigid approaches gave a significantly higher recognition performance in comparison to the rigid approach in both identification and verification. Also, the performance of the non-rigid approach with restricted projection is noticeably better than that with non-restricted projection. Another aspect of the results is that the performance of the two non-rigid approaches does not deteriorate with facial expressions as it is the case with the rigid approach.

The verification rates at 0.001 FAR in the case of the non-rigid approach with restricted projection are 97.8% and 98.35% for the unseen non-neutral probes and neutral probes, respectively. For the probes which are used to augment the training data, the verification rate was 97.68%. At that FAR, the training probes seem to have no advantage over the unseen ones. The verification rate for all the non-neutral probes is 97.73%. A verification rate of 98.14% was achieved for all the probes (neutral and non-neutral). Its first rank identifications are 94.8%, 95.95% and 95.2% respectively for the unseen, training and combined non-neutral probes. The neutral versus neutral identification rate is 97.58%. The identification rate for all the probes is 96.52%.

In comparison to that, the non-restricted projection achieved verification rate at 0.001 FAR of 96.8% for the case of non-neutral probes and 97.94% for the neutral ones. As shown in Fig. 8 at lower FAR the restricted approach maintains high verification rates even better than the non-restricted approach. Its first rank recognition rate is 94.07% for the non-neutral probes and 95.63% for the neutral ones. The CMC curves of the restricted approach rise and settle more quickly than the non-restricted approach. The rigid ICP approach achieved reasonable performance on the neutral scans (93.3% first rank identification) but low performance for the non-neutral scans (about 54%).

In addition, the FRGC ROC I, II and III curves were computed (see Fig. 9). A 4007 × 4007 similarity matrix was computed by matching all the 3D scans in the evaluation data and normalized according to a target set as illustrated in (14). The sub-matrices for ROC I, II and III were then extracted from the 4007 × 4007 matrix using the masks provided by the FRGC protocol. At 0.001 FAR, we obtained the verification rates of 94.55%, 94.12% and 94.05%, respectively for ROC I, II and III. Since these three ROCs involve non-neutral to non-neutral matching, their verification rates are to some extent lower than the non-neutral versus neutral rate (Fig. 8). This is an indication that our approach has better performance when it matches non-neutral 3D scans against neutral gallery scans.
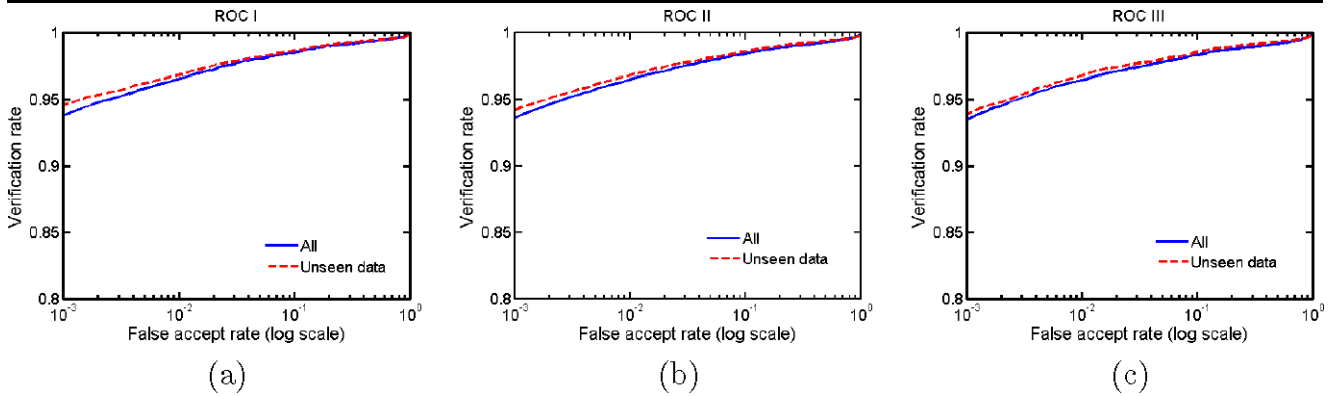
Fig. 9 FRGC ROC I, II and III curves, respectively. The *solid line* is for all the scans defined by the FRGC protocol for ROC I, II and III whereas the *dashed line* is for the same scans excluding the training data
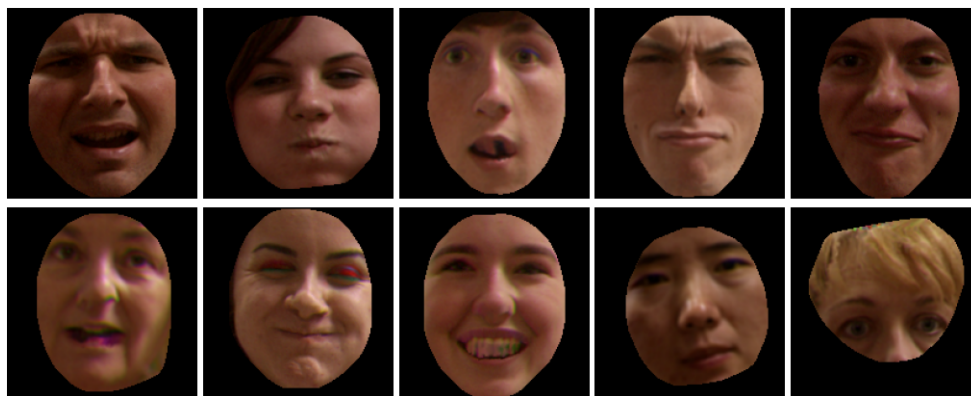

Fig. 10 Examples of challenging probes which were recognized (*top row*) and misrecognized (*bottom row*). One can notice that misrecognition occurs due to data distortion and failure in 3D face detection, cropping and/or pose correction. Misrecognition also occurs due to in-

correct convergence of ICP and when some interpersonal disparities are removed during the removal of expression deformations. However, as indicated by the recognition performance, the effects of these reasons appear to be insignificant

This effect is possibly influenced by the nature of our training data which measures the difference between non-neutral and neutral scans. See Table 1 for the comparative results.

The failures in recognition occur due to four reasons. The first reason is due to an incorrect automatic 3D face detection and pose correction. This reason affects about 2% of the probe scans. The reported results throughout the paper include these probes and no intervention was attempted to correct their pose manually. The performance of the approach can be further improved by performing iterative nose tip detection while pose-correcting the 3D face as nose tip detection is more accurate for pose-corrected 3D faces and/or considering more fiducial points. The second reason is due to the non-convergence of the ICP algorithm which is usually caused by an inaccurate coarse registration. The third reason is due to distortions in the input 3D scans (see Fig. 10). The fourth reason occurs when some interpersonal disparities are also removed during the removal of expression deformations. This however appears to be insignificant.

These four reasons usually occur concurrently. Figure 10 shows examples of challenging probes which were misrecognized and other challenging probes which were correctly recognized.

A comparison of the verification rates at 0.001 FAR of our proposed approach to other high performing approaches in the literature and the PCA baseline performance is presented in Table 1. The proposed approach achieves superior performance compared to other approaches that use only the 3D modality in the case of non-neutral versus neutral 3D face matching. The verification rate for the neutral 3D scans is slightly less than the approach proposed in Mian et al. (2007) but the verification rate of the non-neutral ones is the highest compared to all other considered approaches. The over all verification performance (neutral + non-neutral) of the proposed approach is very close to the Mian et al. (2007). However, our approach has a higher overall (all versus neutral) identification performance than the 3D unimodal performance of Mian et al. (2007) (96.52% compared to 96.2%).

Table 1 Comparison of verification rates are at 0.001 FAR on the FRGC dataset. The results quoted in the table are for the 3D modality alone

	Modality	Neutral vs. neutral	Non-neutral vs. neutral	All vs. neutral	ROC I	ROC II	ROC III
Maurer et al. (2005)	3D/2D	99.2%	NA	95.8%	NA	NA	NA
Husken et al. (2005)	3D	NA	NA	89.5%	NA	NA	NA
Kakadiaris et al. (2007)	3D	NA	NA	NA	97.3%	97.2%	97%
Passalis et al. (2005)	3D	NA	NA	NA	89%	88%	87%
Mian et al. (2007)	3D/2D	99.4%	97%	98.5%	NA	NA	NA
Mian et al. (2008)	3D/2D	97.4%	92.1%	99.9%	NA	NA	NA
FRGC Baseline	3D	45%	NA	40%	NA	NA	NA
This paper	3D	98.35%	97.8%	98.14%	94.55%	94.12%	94.05%

To compare our approach to the GMDS approach by Bronstein et al. (2006a, 2006b), we measure the recognition performance of our approach in a similar experimental set up to theirs and compare the two recognition performances. In Bronstein et al. (2006b), a gallery of 30 subjects was constructed by enrolling a neutral 3D scan per subject. 3D face recognition was performed on 5 non-neutral scans per subject (a total of 180 probes). The 3D scans were taken from the FRGC dataset. They applied mild occlusions (cropping) to the probes by means of a mask which they described as narrow and covers most of the relevant part of the face. In our case we closely simulated their mild occlusions as shown in their example image on p. 11 Bronstein et al. (2006b). As we do not know which probes they used, we repeated the experiment five times. Compared to their verification rate of about 77% at 0.001 FAR (as shown on their ROC curve), our approach achieved a minimum in the five experiments of 91.6%, a maximum of 95.28% and a mean of 93.38% verification rate at the same FAR.

Our approach stores a single cropped and pose-corrected 3D facial range image (161×161 pixels) per gallery face as a signature. This signature extraction time on an Intel Core 2 Quad machine is performed in 4 seconds using a Matlab implementation. Most of the computational complexity of our approach comes from the application of the ICP during the matching stage. The matching part was implemented in C. We used an optimized variant of the ICP algorithm. The used ICP algorithm uses a k-d tree structure to speed up the process of the closest point search as used in Greenspan et al. (2003). The matching time is about 100 ms per gallery face.

3.5 Experiment 3: Person-specific Deformation Model

This experiment aims at measuring the performance of the non-rigid approach when a person-specific deformation model is used instead of the generic one. In this case the gallery contains a neutral image per subject and a deformation model per subject as well. The person-specific deformation models are generated from training data belonging to the same subject.

This approach requires model training data for every person-specific model. However, there are practical application scenarios for this approach. For example, the recognition system can start with a generic model then as it performs recognition during the deployment stage of the system it builds the person-specific models for the repeatedly recognized subjects. The computational complexity of the person specific deformation is the same as when the generic model is used. This gives the approach an advantage over the systems which use multiple enrollment per subjects to increase their recognition performance as verification or identification for a small number of subjects can be achieved in real-time which is needed for applications like robot-human interaction for instance.

The acquired non-neutral data of each of the three subjects are divided into two parts, one part is for the person-specific model training and the other one is for recognition performance evaluation. The part which is used for model training is about 650 non-neutral scans (leaving about 350 probes per subject for recognition performance evaluation). The person-specific models are computed in the same way as the generic ones (see Sect. 2.3). Then a neutral image per each one of the three subjects were added to the gallery which was used for FRGC performance evaluation (Sect. 3.4), making a gallery of 469 subjects.

The ROC and CMC curves of the performance of this approach on only the evaluation probes of the three subjects are shown in Fig. 11. A very high verification and identification rates were achieved even at very low FAR (0.0001) reaching 99.3%, especially when considering that the probes are under facial expressions.

Fig. 11 Person-specific
deformation models gave very
high recognition performance
(all the probes are under facial
expressions)
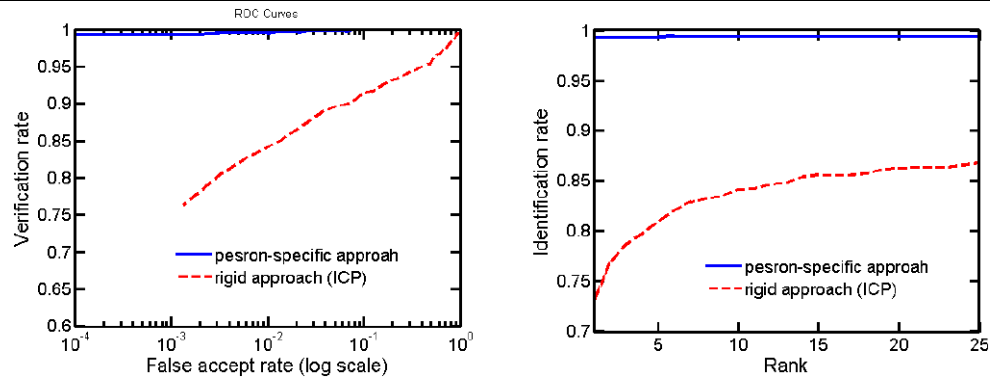
4 Conclusion

By morphing expressions from 3D facial scans using the
proposed Expression Deformation Model, significant recog-
nition performance was achieved especially for scans under
facial expressions. To the best of our knowledge, the per-
formance of our generic restrictive projection approach on
the FRGC dataset is the best 3D (unimodal) face recogni-
tion performance in the literature. In addition, the system
can further improve the recognition performance by adapt-
ing person-specific deformation models during the deploy-
ment stage of the system as shown in one of our experi-
ments.

References

Besl, P., & McKay, N. (1992). A method for registration of 3-D shapes.
    IEEE Transaction on Pattern Analysis and Machine Intelligence,
    14, 239–256.
Blanz, V., & Vetter, T. (1999). A morphable model for the synthesis of
    3D faces. In *Proceedings of siggraph*.
Blanz, V., Basso, C., Poggio, T., & Vetter, T. (2003). Reanimating faces
    in images and video. In *Computer graphics forum*.
Blanz, V., & Vetter, T. (2003). Face recognition based on fitting a 3D
    morphable model. *IEEE Transactions on Pattern Analysis and
    Machine Intelligence*, *25*(9), 1063–1074.
Bowyer, K., Chang, K., & Flynn, P. (2006). A survey of approaches
    and challenges in 3D and multi-modal 3D + 2D face recognition.
    Journal of Computer Vision and Image Understanding, *101*(1),
    1–15.
Bronstein, A., Bronstein, M., & Kimmel, R. (2005). Three-dimensional
    face recognition. *International Journal on Computer Vision*,
    64(1), 5–30.
Bronstein, A., Bronstein, M., & Kimmel, R. (2006a). Generalized mul-
    tidimensional scaling: a framework for isometry-invariant par-
    tial surface matching. *Proceedings National Academy of Sciences
    (PNAS)*, *103*(5), 1168–1172.
Bronstein, A., Bronstein, M., & Kimmel, R. (2006b). Robust
    expression-invariant face recognition from partially missing data.
    In *Proceedings of the European conference on computer vision*.

Chang, K., Bowyer, K., & Flynn, P. (2005a). Adaptive rigid multi-
    region selection for handling expression variation in 3D face
    recognition. In *IEEE conference on computer vision and pattern
    recognition*.
Chang, K., Bowyer, K., & Flynn, P. (2005b). An evaluation of mul-
    timodal 2D + 3D face biometrics. *IEEE Transaction on Pattern
    Analysis and Machine Intelligence*, *27*(4), 619–624.
Chang, K., Bowyer, K., & Flynn, P. (2006). Multiple nose region
    matching for 3D face recognition under varying facial expression.
    IEEE Transaction on Pattern Analysis and Machine Intelligence,
    28(10), 1695–1700.
Chen, Y., & Medioni, G. (1991). Object modeling by registration of
    multiple range images. In *IEEE international conference on ro-
    botics and automation* (Vol. 3, pp. 2724–2729).
Chua, C., Han, F., & Ho, Y. (2000). 3D human face recognition using
    point signature. In *IEEE conference on automatic face and gesture
    recognition*.
Faltemier, T., Bowyer, K., & Flynn, P. (2007). Using a multi-instance
    enrollment representation to improve 3D face recognition. In
    *IEEE conference on biometrics: theory, applications and systems
    (BTAS)*.
Gokberk, B., Salah, A., & Akarun, L. (2005). Rank-based decision fu-
    sion for 3D shape-based face recognition. In *International confer-
    ence on audio-based biometric person authentication*.
Greenspan, M., & Yurick, M. (2003) Approximate K-D tree search
    for efficient ICP. In *IEEE international conference on recent ad-
    vances in 3D digital imaging and modeling*.
Hesher, C., Srivastava, A., & Erlebacher, G. (2003). A novel technique
    for face recognition using range images. In *International sympo-
    sium on signal processing and its applications*.
Husken, M., Brauckmann, M., Gehlen, S., & Malsburg, C. (2005).
    Strategies and benefits of fusion of 2D and 3D face recognition.
    In *IEEE workshop on face recognition grand challenge experi-
    ments*.
Jollife, I. (1986). *Principal component analysis*. Berlin: Springer.
Kakadiaris, I., Passalis, G., Toderici, G., Murtuza, M., Lu, Y., Karam-
    patziakis, N., & Theoharis, T. (2007). Three-dimensional face
    recognition in the presence of facial expressions: an annotated de-
    formable model approach. *IEEE Transaction on Pattern Analysis
    and Machine Intelligence*, *29*(4), 640–649.
Kirby, M., & Sirovich, L. (1990). Application of the KL procedure for
    characterization of human faces. *IEEE Transactions on Pattern
    Analysis and Machine Intelligence*, *12*(1), 103–108.
Li, X., & Zhang, H. (2007). Adapting geometric attributes for
    expression-invariant 3D face recognition. In *IEEE conference on
    shape modeling and applications*.
Lu, X., Hsu, R., Jain, A., Kamgar-Parsi, B., & Kamgar, B. (2004). Face
    recognition with 3D model-based synthesis. In *Proceedings of the
    international conference on biometric authentication*.

Lu, X., & Jain, A. (2008). Deformation modeling for robust 3D face matching. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, *30*(8), 1346–1357.

Lu, X., Jain, A., & Colbry, D. (2006). Matching 2.5D face scans to 3D models. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, *28*(1), 31–43.

Maurer, T., Guigonis, D., Maslov, I., Pesenti, B., Tsaregorodtev, A., West, D., & Medioni, G. (2005). Performance of geometrix activeID 3D face recognition engine on the FRGC data. In *IEEE conference on computer vision and pattern recognition*.

Mian, A., Bennamoun, M., & Owens, R. (2006). Automatic 3D face detection, normalization and recognition. In *International symposium on 3D data processing, visualization and transmission (3DPVT)*.

Mian, A., Bennamoun, M., & Owens, R. (2007). An efficient multimodal 2D–3D hybrid approach to automatic face recognition. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, *29*(11), 1927–1943.

Mian, A., Bennamoun, M., & Owens, R. (2008). Keypoint detection and local feature matching for textured 3D face recognition. *International Journal of Computer Vision*.

Pan, G., Han, S., Wu, Z., & Wang, Y. (2005). 3D face recognition using mapped depth images. In *IEEE workshop on face recognition grand challenge experiments*.

Parke, P., & Waters, K. (1996). *Computer facial animation*. Natik: A.K. Peters Ltd.

Passalis, G., Kakadiaris, I., & Theoharis, T. (2007). Intra-class retrieval of non-rigid 3D objects: application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*(2), 218–229.

Passalis, G., Kakadiaris, I., Theoharis, T., Tederici, G., & Murtaza, N. (2005). Evaluation of 3D face recognition in the presence of facial expressions: an annotated deformable model approach. In *FRGC workshop, in IEEE conference on computer vision and pattern recognition*.

Phillips, P., Flynn, P., Scruggs, T., Bowyer, K., Chang, J., Hoffman, K., Marques, J., Min, J., & Worek, W. (2005). Overview of the face recognition grand challenge. In *IEEE conference on computer vision and pattern recognition*.

Ramachandran, M., Zhou, S., Jhalani, D., & Chellappa, R. (2005). A method for converting a smiling face to a neutral face with applications to face recognition. In *IEEE conference on acoustics, speech, and signal processing*.

Russ, T., Boehnen, C., & Peters, T. (2006). 3D face recognition using 3D alignment for PCA. In *IEEE conference on computer vision and pattern recognition*.

Rydfalk, M. (1987). *CANDIDE, a parameterized face* (Report No. LiTH-ISY-I-866). Consulting Psychologists Press Inc.

Samir, C., Srivastava, A., & Daoudi, M. (2006). Three-dimensional face recognition using shapes of facial curves. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, *28*(11), 1858–1863.

Sirovich, L., & Kirby, M. (1987). Low-dimensional procedure for characterization of human faces. *Journal of the Optical Society of America*, *2*, 519–524.

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, *3*(1), 71–86.

Wang, Y., Pan, G., & Wu, Z. (2007). 3D face recognition in the presence of expression: a guidance-based constraint deformation approach constraint deformation approach. In *IEEE conference on computer vision and pattern recognition*.

Zhao, W., Chellappa, R., Phillips, P., & Rosenfeld, A. (2003). Face recognition: a literature survey. *ACM Computing Surveys*, *35*(4), 399–458.