Computers & Graphics 34 (2010) 231–241

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Technical Section

Expression modeling for expression-invariant face recognition

Frank B. ter Haar [a,*], Remco C. Veltkamp [b]

[a] Electro-Optics, TNO Defence Security and Safety, The Hague, The Netherlands
[b] Information and Computing Sciences, Utrecht University, The Netherlands

ARTICLE INFO

Article history:
Received 4 December 2009
Received in revised form
12 March 2010
Accepted 15 March 2010

Keywords:
Morphable models
Expression modeling
Face recognition
Correspondence estimation

ABSTRACT

Morphable face models have proven to be an effective tool for 3D face modeling and face recognition, but the extension to 3D face scans with expressions is still a challenge. The two main difficulties are (1) how to build a new morphable face model that deals with expressions, and (2) how to fit this morphable face model automatically to new 3D face scans with unknown expressions. This work presents a multi-resolution approach to semi-automatically build seven morphable expression models, and one morphable identity model from scratch. We propose an algorithm that automatically selects the proper pose, identity, and expression such that the final model instance accurately fits the 3D face scan. To prove high fitting accuracy and its use for face recognition, we perform experiments on the publicly available UND, GAVAB, BU-3DFE, FRGC v.2 datasets. Our results show high recognition rates of respectively 99%, 98%, 100%, and 97% after the automatic removal of the expressions.

1. Introduction

Statistical models of the human face have proven to be an effective tool for person identification using 3D face scans. To build a statistical model, a set of example faces is required with face features in full correspondence. With such a model, a new face instance can be constructed as a linear combination of the example faces. For 3D face identification, the idea is to use the statistical model to construct a face instance that resembles an input image. The way these example faces are combined linearly to represent an input face, provides both global and local information about the input face, that can be used to classify and identify different input faces. Expressions are a problem, because they change the resemblance of the input faces.

1.1. Related work

Most of the early 3D face recognition methods focused on variants of the iterative closest point (ICP) [2] algorithm to find similarity between two 3D face scans. As 3D face recognition became more challenging with larger datasets and expression scans, the ICP-based methods showed two main disadvantages. The non-rigid expression deformations forced the ICP-based methods to rely on smaller face regions such as the nose and forehead, and the computational expensive face matching lowered its practical use. Methods of Faltemier et al. [8] and Mian et al. [12] reported high recognition rates based on nose regions in combinations with ICP.

For efficient face matching, the extraction of person specific features became the new area of interest. For instance, the re-parameterization of each facial surface to an indexed collection of curves enables the direct comparison of these curves during face matching [16]. With high recognition rates, low computational costs during face matching, and high robustness to noise and missing data, *3D morphable face model* based methods prove to perform well. To build a 3D morphable face model, dense correspondence are required among a set of 3D example faces. The mean face and the statistical variation of these faces can be computed using principal component analysis (PCA). Using the statistical face variations, the mean face can be deformed to fit the noisy scan data. The way such a model is deformed (larger, wider, longer nose, etc.), provides information on the geometric shape properties of the input face. The coefficients that induce these deformations form a relatively small feature vector for efficient face matching. For reliable model coefficients, the model deformation must be independent of changes in the face pose. Therefore, the model fitting is often combined with an ICP algorithm to compensate for the rigid transformation between closest point features. Because both the model fitting and the ICP algorithm are local optimization methods, a coarse alignment between the scan data and the model should be automatically established first.

In [4], Blanz et al. fit a morphable model to 3D scan data and use the deformation weights (or model coefficients) to recognize faces with neutral expressions. In each iteration of their stochastic Newton algorithm, the current model instance is projected to 2D image space and the model coefficients are adjusted according to

* Corresponding author.
E-mail addresses: frank.ter.haar@live.com, frank.terhaar@tno.nl (F.B. ter Haar).

0097-8493/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cag.2010.03.010

the difference in texture and depth values. For the coarse alignment and to initiate the morphable model, they manually select seven corresponding face features on their model and in the depth scan.

Amberg et al. [1] have built a PCA model from 270 identity vectors and a PCA model from 135 expression vectors and combined the two into a single morphable face model. Their method fits this model to 3D scan data by iteratively finding closest point pairs to improve on the alignment, the identity deformation, and the expression deformation at the same time. Their local optimization method, which does not guarantee convergence to the global minimum, returns a set of identity coefficients that perform well in terms of face recognition.

Lu and Jain [11] train a morphable expression model for each expression in their test set. Starting from an existing neutral scan, they fit each of their expression models separately to adjust the vertices in a small region around the nose to lower the ICP error between that particular neutral scan and an expression scan. The expression model that produces the most accurate fit is used to deform the neutral scan. For the initial alignment they use three automatically detected feature points. For the fitting, they combine the accurate ICP alignment for the rigid transformation with the fast *eigenspace projection* [19] for the expression deformation. This process is iterated until convergence and the lowest residual error is used as the dissimilarity score between the neutral scan and the new scan. Although the authors use PCA models, their method can be classified as an ICP based method, because the fitting procedure has to be repeated for every pair of face scans in the dataset. The expression models are merely used to improve on the ICP fitting procedure.

Mpiperis et al. [14] build a bilinear PCA model for the BU-3DFE dataset suitable for both expression and identity recognition *after* a face scan is brought into full correspondence with the model. To establish full correspondence, they detect the boundary of the mouth, elastically deform a low resolution face mesh to the scan data (considering the mouth), and subdivide the mesh for denser correspondences. The bilinear PCA model is solely used to map the full correspondence to expression and identity coefficients that are either used for expression classification or person identification.

Kakadiaris et al. [9] deform an annotated subdivision face model to scan data. Their non-statistical deformation is driven by triangles of the scan data attracting the vertices of the model. The deformation is restrained by a stiffness, mass and damping matrix, which control the resistance, velocity and acceleration of the model's vertices. They use the newly created geometry for wavelet analysis and achieve state of the art recognition results on the face recognition grand challenge (FRGC) [15].

1.2. Contribution

Starting with a dataset of neutral scans, expression scans, and a small set of facial landmarks, we describe how to build a strong multi-resolution PCA model for both identity and expression variations of the human face. We build one morphable identity model and seven separate morphable expression models, for the 'expressions' anger, disgust, fear, happiness, sadness, surprise, and inflated cheeks. For expression invariant face recognition it is important to have the identity and expression models separated, but it makes the model-to-scan fitting more complex. Our fitting algorithm automatically searches for the best identity and expression combination, such that the newly created face instance accurately fits the input face scan. It combines eigenspace sampling to avoid local minima, eigenspace projection for fast local convergence, and predefined face components for higher accuracy. The shape priors captured in the morphable face model allows for the robust handling of noise and holes in the face scan. After the fitting process, the final model instance and its PCA coefficients can be used as; the filled and noiseless representation of the expression scan, to retrieve the defined landmarks, to bootstrap the face model, to remove the expression, and for expression invariant face recognition. In our work, face recognition is not only an interesting application, it also quantifies the uniqueness of the model coefficients and indirectly the fitting accuracy as well.

The contributions are:

- An easy way to build different identity and expression models.
- The decoupling of the rigid model-to-scan alignment and the non-rigid model-to-scan deformation for efficient face modeling.
- An enhanced automatic fitting algorithm to establish dense correspondences among faces with expressions and to extract landmarks.
- New face recognition with the use of multiple local minima in the identity space.
- Performance evaluation of three different coefficient vectors for the recognition.

2. Datasets

We use the 3D face scans of the UND [5], the GAVAB [13,17], the BU-3DFE [21], the FRGC v.2 [15], and the USF Human ID 3D [20] databases. The UND set, from the University of Notre Dame, contains 953 frontal range scans of 277 different subjects with mostly neutral expression. The GAVAB set consists of nine low quality scans of which we use seven for each of its 61 subjects as in [17]. The BU-3DFE set, from the Binghamton University, was developed for facial expression classification. This set contains one neutral scan and 24 expression scans having different intensity levels for each of its 100 subjects. The FRGC v.2 set, of the Face Recognition Grand Challenge contains 4007 high quality 3D face scans of 466 different subjects. Almost half of these scans show an expression varying from a smile or frown to a pronounced laugh or a surprised look. The USF Human ID 3D database, from the University of South Florida, contains 136 high quality full head scans without expressions.

We aim at 3D face modeling and recognition, and therefore we need to segment the face from each scan. For that, we employ the pose normalization method described in [18] that takes as input a triangular surface mesh and outputs the normalized pose of the face with the tip of the nose in the origin. The face is segmented by removing the scan data with a Euclidean distance larger than 130 mm from the nose tip. In several scans of the FRGC v.2, the frontal pose was not completely recovered due to hair covering the face. To further improve on the face's pose, an average nose template (shown in Fig. 1) is aligned to each segmented face and the inverse transformation applied to the scan. This template was selected for its high expression invariance as described in [12]. Qualitative evaluation showed that the tip of the nose was found in all 3D scans, except for two scans of the FRGC v.2 which did not have a nose (2 failures out of 8023 scans).

3. Morphable face model

In this work, we use a new morphable face model built from both 3D neutral and expression scans of the human face. We fit this model to 3D scan data in such a way that expressions can be removed and subjects identified in an expression invariant manner. To build a morphable face model with expressions, an example set of subjects showing various expressions is required.

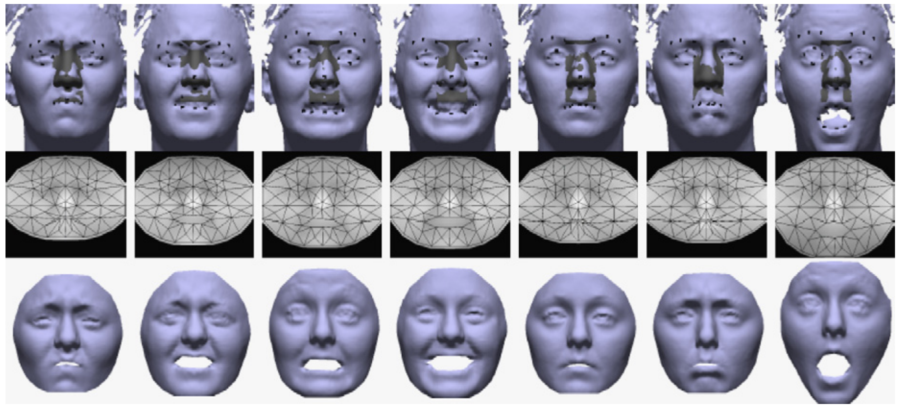

Fig. 1. Semi-automatic model building. The pose normalized faces are annotated with landmarks (first row) that correspond among different expressions and different subjects to construct an initial face mesh as a layer over the cylindrical depth image (second row). A subdivision scheme is applied to acquire dense 3D face meshes (third row).

For that, we use the BU-3DFE [3] dataset, from which we select the 100 neutral scans and 600 expression scans at their highest intensity level. The BU-3DFE set was developed for facial expression classification. This set contains one neutral scan and 24 expression scans having different intensity levels, for each of its 100 subjects. From this set we selected the neutral scans and the highest intensity level expression scans (anger, disgust, fear, happiness, sadness, surprise at level 4). The goal is to model a neutral face model from a dense set of correspondences, and a neutral-to-expression model for each of the expressions anger, disgust, fear, happiness, sadness and surprise. The neutral face model, which is built from the 100 neutral scans, captures the identity of different subjects, whereas a neutral-to-expression model captures the facial changes caused by a certain expression.

A morphable face model is a type of statistical point distribution model (PDM) [7], where the points are facial features that have a different distribution among different faces. Building a morphable face model, requires n dense correspondences $S=(x_1,y_1,z_1,\ldots,x_n,y_n,z_n)^T \in \Re^{3n}$ among the input face scans of the training set. Principal component analysis (PCA) is used to capture the statistical distribution of these correspondences among the input faces. Because the automatic estimation of reliable dense correspondences among noisy face scans with expressions is still unsolved, we propose a semi-automatic correspondence estimation that requires 26 facial landmarks. With the use of these 26 landmarks, we construct a low resolution mesh that is projected to the cylindrical depth image of a 3D face scan. By subdividing the triangles of the low resolution mesh, a multi-resolution representation of the face is constructed. At each level, we assume that the vertices between different subjects or expressions correspond. The correspondences at the highest level are used to build a neutral 3D morphable face model as well as a morphable expression model for each of the expressions. Because the manual annotation of facial landmarks in 3D face scans is often a major disadvantage in statistical modeling, we explain in Section 3.6 how our initial morphable face model can be used to enhance itself with new 3D scan data. This *automatic bootstrapping* is a useful tool to limit the user input. We explain the semi-automatic construction of the morphable identity and expression models using the following steps:

(1) Manual annotation of facial landmarks, including nose, eyes, eyebrows, and mouth.
(2) Cylindrical depth image construction.
(3) Multi-resolution face mesh construction.
(4) Building the morphable identity model.
(5) Building the morphable expression models.
(6) Automatic bootstrapping the morphable model.
(7) Data reduction.
(8) Component selection.

3.1. Landmark annotation

In each of the 700 pose normalized (raw) BU-3DFE scans, we manually selected the same sequence of 26 facial landmarks as an initial set of correspondences. These landmarks include locations on the nose, mouth, eyes, and eyebrows, and provide a coarse notion of facial changes among different identities and expressions. This is the only user input throughout this work. In fact, most of these landmarks were already annotated in the BU-3DFE set and the nose tip was detected automatically.

3.2. Cylindrical depth image

Knowing that almost all face scans (even with facial hair and expressions) can be correctly pose normalized after the final alignment to an average nose template (Section 2), it makes sense to build the morphable face model based on face scans in the coordinate system of this nose template. Each BU-3DFE scan was brought into alignment with the reference nose template, which has the desired pose and its nose tip in the origin. Although the nose template was accurately fitted to the face scans, this does not mean that the nose tip of the face scan is aligned to the nose tip in the template. A smaller nose, for instance, has its tip behind the template (lower z-value) and a larger nose in front of the template (higher z-value). To produce a cylindrical depth image $d(\theta,y)$ for each of the face scans, we simulate a cylindrical laser range scanner. To cover most of the face, the nose template and the aligned face scans are moved 80 mm along the positive z-axis. A surface sample is acquired for each angle θ at each height y with radius distance d to the y-axis of the coordinate system. Basically, we cast a horizontal ray at height y with an angle θ in the xz-plane from the y-axis to the face scan, and store the distance to the furthest ray–triangle intersection. Because we model the face only, we scan the front half of the cylinder i.e. the angles $\theta=[180^\circ,360^\circ]$. The step size for θ is 0.4° (450 angles) and the step size for y is 0.5 mm, producing a high resolution 2D cylindrical depth image. The 26 annotated landmarks are projected to the cylindrical depth image, by assigning them to the closest ray. Note that the cylindrical depth image can be converted to a 3D triangle mesh by connecting the adjacent samples and projecting the cylindrical coordinates to 3D.

3.3. Multi-resolution face mesh

To construct the low resolution face mesh, we extend the initial set of landmarks using predefined locations on the cylindrical depth map, such as the location (270°, nose tip$_y$+180) on the forehead, the location (270°, lower lip$_y-70$) on the chin, leftmost location (180°, upper lip$_y$) and rightmost location (360°, upper lip$_y$). This way a coarse 2D mesh is constructed as an overlay on the cylindrical depth image (Fig. 1). By using relative locations, we ensure that all necessary face features (chin, forehead, cheeks) are captured, whereas the cylindrical depth images provide radial depth variation among different subjects. Alternatively, a more geometry guided approach could be used instead. Large triangles are avoided in the coarse mesh, by adding extra vertices in sparse density areas. The final low resolution mesh consists of 68 vertices and 110 triangles. To improve on the cylindrical depth map quality, depth values outside the face polygon and inside the mouth polygon are removed, and depth values are interpolated and extrapolated to fill the gaps. With the use of this underlying depth map, the face mesh can be projected to 3D. To construct a higher resolution face mesh, we subdivide each triangle in the low resolution mesh into four smaller congruent triangles. In an iterative manner, we construct five meshes with a resolution up to 28 160 (110×4^4) triangles. The highest resolution mesh is projected in 3D using the cylindrical depth image and speckle noise is removed by a single iteration of Laplacian smoothing. The advantage of the subdivision scheme is that each vertex in a lower resolution mesh has the same index number in the highest resolution mesh, which means that the highest resolution mesh can be used as the final multi-resolution face mesh.

In the end we have acquired for each input face a set of dense correspondences $S=(x_1,y_1,z_1,\ldots,x_n,y_n,z_n)^T\in\Re^{3n}$, with 28 160 triangles and $n=14\,288$ vertices of which the first 26 vertices were manually annotated. Valid transitions for lower resolution faces are in this case, $n=68$, $n=246$, $n=932$, $n=3624$. To distinguish between a face with an expression and a neutral face, we use E_i for an expression face and S_i for a neutral face in full correspondence. The multi-resolution mesh construction is shown in Fig. 1. In addition to the multi-resolution mesh, we also construct a mirrored version. Therefore, we interchange the coordinates of the left and right side landmarks, mirror the cylindrical depth map in the y-axis, and redo the multi-resolution mesh construction. With these additional faces, the variability of the morphable face model increases. Also, the statistical mean face becomes fully symmetric around the y-axis, because facial asymmetry is modeled in both directions.

3.4. Morphable identity model

Building an identity based face model requires a training set of neutral faces of different subjects in full correspondence. For that we use the $m=200$ (original and mirrored) neutral face instances $S=(x_1,y_1,z_1,\ldots,x_n,y_n,z_n)^T\in\Re^{3n}$, with $n=14\,288$. Principal component analysis (PCA) is applied to these neutral face instances S_i to acquire an orthogonal coordinate system in which each face can be described as a linear combination of principal shape vectors, i.e. the eigenvectors of the eigenspace. Turk and Pentland [19] also described how to compute the 'eigenfaces' that define this 'face space', for 2D intensity images.

Each of the $m=200$ face instances S_i is described as a one-dimensional vector of size $3n$, and the average of these vectors is the mean face shape $\overline{S}$. The mean shape $\overline{S}$ is extracted from each face instance S_i, and these shape deformation vectors are stored in a matrix $A[S_1-\overline{S},S_2-\overline{S},\ldots,S_m-\overline{S}]$ of size $3n\times m$. To compute the eigenfaces, a covariance matrix $C=A\,A^T$ is constructed from which the eigenvectors and eigenvalues are extracted which can be done efficiently as described in [19]. The eigenvectors $\mathbf{s_i}=(\Delta x_1,\Delta y_1,\Delta z_1,\ldots,\Delta x_n,\Delta y_n,\Delta z_n)^T$, the eigenvalues λ_i and identity coefficients α_i are used to model an identity vector according to

$$S_{id}=\sum_{i=1}^{m-1}\alpha_i\sqrt{\lambda_i}\cdot\mathbf{s_i}.$$

Adding this identity vector to the mean face $\overline{S}$ results in a 3D face with a new identity, $S_{inst}=\overline{S}+S_{id}$. The identity coefficient α_i represents the number of standard deviations $\sigma_i=\sqrt{\lambda_i}$ that a face instance morphs along eigenvector $\mathbf{s_i}$. To determine the coefficient α_i for face instances S_i, one can subtract $\overline{S}$ and project its residual identity vector S_{id} into face space:

$$\alpha_i=\frac{1}{\sqrt{\lambda_i}}(\mathbf{s_i}^T S_{id}).$$

The projection of the identity vector onto each eigenvector returns the least-squares solution defined as $\boldsymbol{\alpha}=(\tilde{S}^T\tilde{S})^{-1}(\tilde{S}^T S_{id})$, because the columns in matrix $\tilde{S}=[\mathbf{s_1},\mathbf{s_2},\ldots,\mathbf{s_{m-1}}]$ are orthogonal [10]. Without the use of PCA (as in [11]), one must solve the least-squares solution for the (linearly independent) face instances S_i according to $\boldsymbol{\alpha}=(A^TA)^{-1}(A^TS_{id})$, which is computationally more expensive. In the end, the vector $\boldsymbol{\alpha}$ can be used to describe a subject in face space, and as a feature vector for the recognition of 3D faces.

3.5. Morphable expression model

Building an expression model requires full correspondence between all the neutral faces and the sets of expression faces, which we established in Section 3.3. Matrix A is now initiated with the difference between the expression face E_i and neutral face S_i of subject i. The computation of the eigenvalues and eigenvectors for the matrix $A[E_1-S_1,\,E_2-S_2,\,\ldots,\,E_m-S_m]$ remains the same. The eigenvectors $\mathbf{e_i}=(\Delta x_1,\Delta y_1,\Delta z_1,\ldots,\Delta x_n,\Delta y_n,\Delta z_n)^T$, the eigenvalues μ_i ($\sigma_i^2=\mu_i$) and weights β_i are used to model an expression vector according to

$$S_{expr}=\sum_{i=1}^{m-1}\beta_i\sqrt{\mu_i}\cdot\mathbf{e_i}.$$

Adding an expression vector S_{expr} to a neutral face instance S_{inst}, results in a 3D face with a certain expression, $S_{expr.inst}=\overline{S}+S_{id}+S_{expr}$.

After the correspondence estimation and mirroring, our training set consists of 200 neutral faces and 1200 expression faces in full correspondence. At this point, we could either build a generic model including all expressions or a specific model for each of the expressions, anger, disgust, fear, happiness, sadness, and surprise. In the work of Lu and Jain [11], experiments with an expression-generic and expression-specific models show that the latter outperforms the former. Although their example faces are different and the expressions were only modeled in a small area around the nose, we decided to use expression-specific models as well. For each of the expressions we build a new model ($\beta_i,\mu_i,\mathbf{e_i}$), which we use to add an expression to the neutral face instance S_{inst}, but also to remove an expression from $S_{expr.inst}$.

3.6. Automatic bootstrapping

For face recognition purposes it is important to have an identity model S_{id} that describes a large human population. The face space allows for the interpolation between example faces and the extrapolation outside its statistical boundary, but only to some extend. In case a subject cannot be sufficiently described in face space, its identity coefficients α_i become unreliable. However, manually annotating more 3D face scans is not desired. Instead, we automatically enhance the identity model with 134 (beardless) scans of the USF Human ID 3D database. For that, the morphable face model is fitted to each scan as described in Section 4 and the set of 26 facial landmarks extracted. Then the cylindrical depth images are constructed, the multi-resolution face meshes S_i created, and the identity model S_{id} rebuild with the 468 original and mirrored sets of correspondences S_i. These steps are shown in the flow chart of Fig. 2.

Because the FRGC v.2 dataset contains not only scans with the aforementioned expressions, but also scans with inflated cheeks, we select twenty subjects with inflated cheeks and their neutral faces, and build an 'expression' model of these scans. For that, we again use our model fitting algorithm to establish full correspondences between morphable identity model and these forty FRGC scans. Afterwards, PCA is applied to the 20 regular and 20 mirrored expression vectors to build the expression model S_{expr} for cheek inflation.

The automatic bootstrapping method that we use here, does not require a perfect fit of the entire model, but just for the 26 annotated landmarks. However, to get these landmarks in place, it helps to fit the full morphable model and not just the vertices that correspond to such landmarks.

3.7. Data reduction

In Fig. 3, the mean face $\overline{S}$ is deformed along the first two eigenvectors of either the identity model or an expression model. Because each of the expressions causes a similar face deformation among the training subjects, the first eigenvector is the main vector to move from the cluster of neutral faces to the cluster of expression faces. A negative coefficient β_1 for eigenvector $\mathbf{e_1}$ means that we move away from an expression cluster, causing unrealistic changes. Note that a weight of $\beta_1 = 2$ is already an exaggeration of the expression. The first eigenvector of all expression models except the sad model causes a larger shape deformation than the first eigenvector of the identity model, which in turn causes a larger shape deformation than the second eigenvector of the expression models. This is reflected by the eigenvalues λ_i and μ_i of the deformation models, which are larger in case of a larger shape deformation. During the morphable model fitting it is important to optimize the large shape deformations before the smaller shape deformations. Because the smallest eigenvectors are the least significant and add merely noise to a model instance, we reduce the number of eigenvectors for the identity model to m_s=80 and for each expression model to m_e=6.

3.8. Component selection

Each eigenvector of a morphable model defines a translation vector for each vertex in the model. With the use of a mask vector one can simply turn a vertex in the model on or off. A vertex that is not selected, is not adjusted nor evaluated during the model fitting. So, the use of a lower resolution for the *multi-resolution* face model speeds up the fitting process. Additionally, we can select *predefined components* such as the nose, eyes, mouth, and the rest of the face and refine each of these components individually to have a larger face variety with the same model [18]. With our expression deformation models, we can find regions of the face that are *invariant to expressions*. When the expression coefficients β_i of an expression model are all set to one, each vertex is translated to a new position. These translation vectors have different lengths, depending on the selected expression model and the expression invariance of that particular

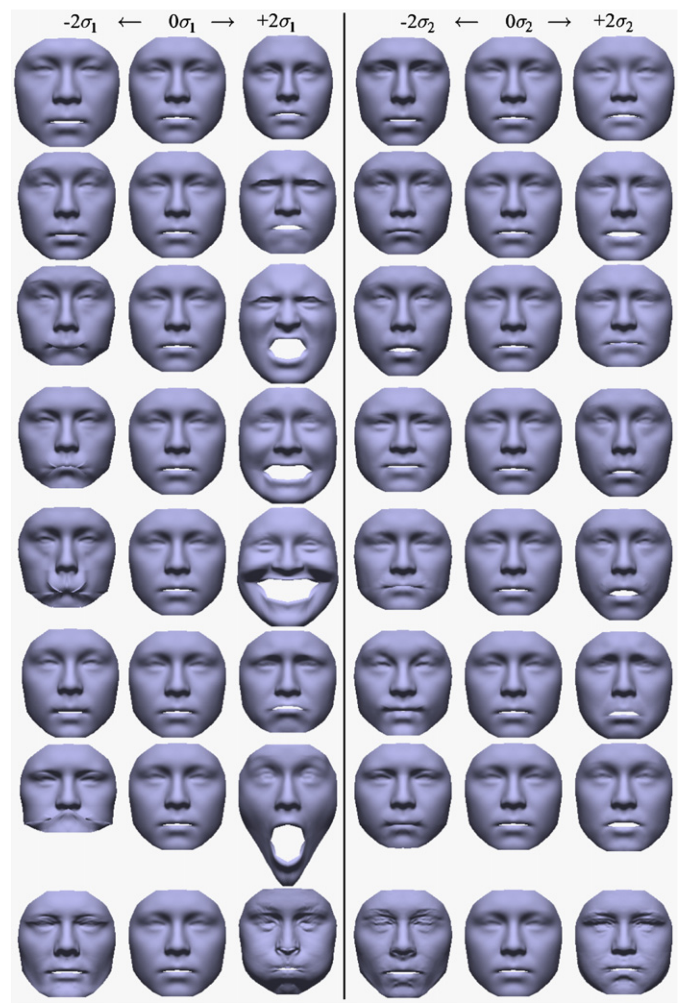

Fig. 3. Face deformation along the first eigenvector (left) and the second eigenvector (right). Starting from the mean face, a model based shape deformation is applied by changing the coefficients α_1, α_2, β_1, or β_2 to -2 or +2. From top to bottom, the deformation is based on the identity model ($\sigma = \alpha$), and the expression models ($\sigma = \beta$) anger, disgust, fear, happiness, sadness, surprise, and cheek inflation. The deformation $\beta_1 = -2$ results in unrealistic faces.

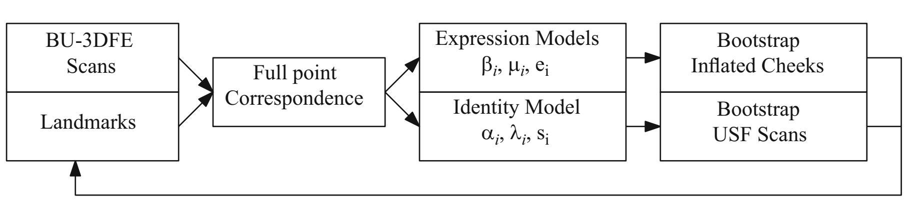

Fig. 2. Flow chart of the semi-automatic model building.

vertex. If the maximum displacement, over all expression models, is stored for each vertex, we can determine face regions that are more static under different expressions. We select 60% of the vertices with the smallest maximum vertex displacements as the static face component. A static face component can be used to coarsely estimate the identity coefficients before estimating the expression coefficients. Fig. 4 shows these sets of selected vertex indices.

4. Morphable model fitting

The task of the model fitting algorithm is to find the face instance $S_{expr.inst}$ in the high dimensional face space that produces the best point-to-point correspondence with a new face scan. Additionally, the model fitting algorithm should be robust to noise and perform well even when large areas of the face are missing. To regulate the scan density, each face surface is cylindrically rescanned (as in Section 3) to a uniform resolution ($\Delta\theta = 1.3, \Delta y = 1.3$) with approximately 16 000 vertices. Slender triangles and small connected components are removed. When accurate point-to-point correspondences are established between the morphable face model and the scan data, then the identity coefficient vector $\boldsymbol{\alpha}$ can be used for face recognition, or the expression deformation S_{expr} can be subtracted from $S_{expr.inst}$ to produce a neutral face instance S_{inst} for geometry based face recognition. In this section, we describe a fully automatic method that efficiently optimizes the weights $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ to obtain a model instance from the high dimensional face space that accurately fits the face scan. To evaluate if an instance of the morphable face model is a good approximation of the 3D face scan, we use the root mean square (RMS) distance of closest point pairs in Euclidean space $e_{\min}$,

$$d_{rms}(S_{expr.inst}, scan) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_{\min}(p_i, scan)^2}$$

using n vertices of $S_{expr.inst}$. Closest point pairs (p, p') for which p' belongs to the boundary (including holes) of the face scan are not used in the distance measure.

Several methods for 3D morphable model fitting have been proposed in the literature [1,4,11,18]. These methods consider two transformations, a rigid transformation to align the model with the scan data and a non-rigid deformation to deform the model to the scan data. As described in Section 2, our method computes the rigid transformation only once, with the use of a pose normalization method that aligns the scan data with an average nose template. This transformation is kept constant during the fitting process for fast model fitting. The model fitting is separated into an *expression optimization* step to select the best expression model and an *identity optimization* step to find the global minimum (and local minima) in the identity space. The morphable model fitting is shown in Fig. 5 as a flow chart and in Fig. 6 based on an example face. Both the expression and identity fitting use the two *coefficient selection* algorithms described below.

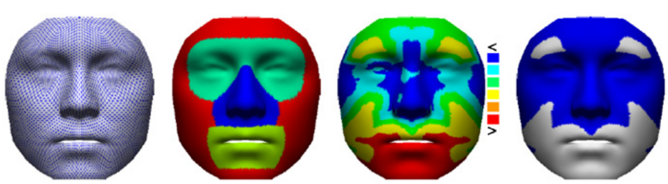

Fig. 4. Component selection. From left to right, $n = 3624$ lower resolution vertices, four face components, expression deformation intensity, and 60% most static vertices under various expressions.

4.1. Coefficient selection

With the face scan aligned to the average nose template, we can compute the closest point pairs between the mean instances $\overline{S}$ and the scan. These closest point correspondences will only be reliable when the scan closely resembles the mean face. To improve on the set of correspondences, the model coefficients of the principal identity vectors and the expression vectors are adjusted iteratively using *eigenspace sampling* (algorithm ESSamp). After a number of iterations, the correspondences of $S_{expr.inst}$ are reliable enough to apply *eigenspace projection* (algorithm ESProj) and to evaluate the fit using d_{rms}. Fig. 7 shows a schematic view of the iterative search through (2D) coefficient space. To optimize expressions, algorithms ESSamp and ESProj use β_i, μ_i, $\mathbf{e_i}$ instead.

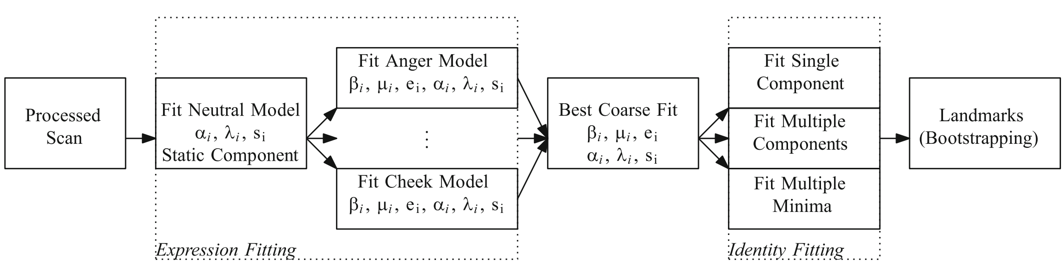

Fig. 5. Flow chart of the combined identity and expression model fitting.

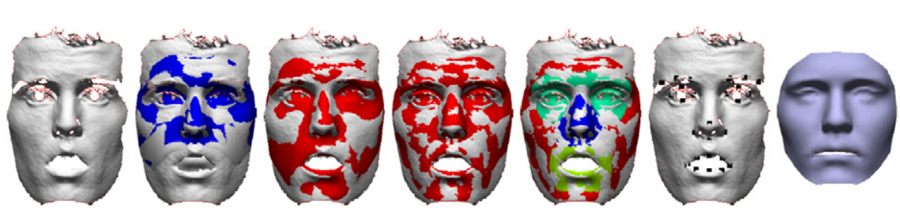

Fig. 6. Example faces of the morphable model fitting. From left to right, the processed scan, the static region fit, the selected coarse fit (surprised), the full face refinement, the multiple component refinement, the automatically acquired landmarks, and the neutralized face model.

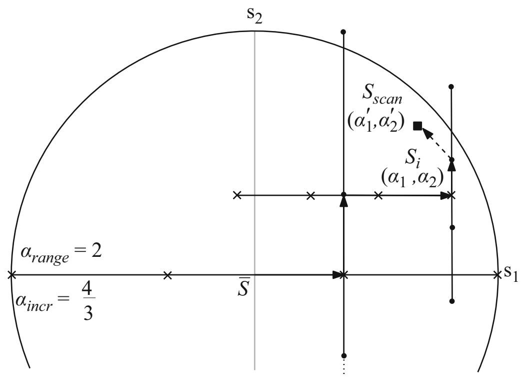

Fig. 7. Searching the coefficient space using eigenspace sampling (solid arrows) and eigenspace projection (dashed arrow). At each mark, the model instance S_i is updated and correspondence with the *scan* estimated.

Algorithm 1. ESSamp (S,$\boldsymbol{\alpha}$,$scan$)

for $k \leftarrow 1$ **to** $k_{\max}$ **do**
  $\alpha_{incr} = \frac{2}{3}\alpha_{range}$ (4 samples in full range)
  for $i \leftarrow i_{\min}$ **to** $i_{\max}$ **do**
    for $\alpha'_i \leftarrow \alpha_i - \alpha_{range}$ **to** $\alpha_i + \alpha_{range}$ **do**
      update $S' = S + ((\alpha'_i - \alpha_i)\sqrt{\lambda_i} \cdot \mathbf{S})$
      $d_{rms}(S', scan)$ smaller $\rightarrow \alpha_{opt} = \alpha'_i$
      $\alpha'_i = \alpha'_i + \alpha_{incr}$
    update $S = S + ((\alpha_{opt} - \alpha_i)\sqrt{\lambda_i} \cdot \mathbf{S})$
  $\alpha_{range} = \frac{5}{4}\alpha_{incr}$ (slight overlap)
return $d_{rms}(S, scan)$

Algorithm 2. ESProj (S,$\boldsymbol{\alpha}$,$scan$)

for $k \leftarrow 1$ **to** $k_{\max}$ **do**
  select sets of correspondences S and S_{scan}
  compute residual deformation vector S-S_{scan}
  for $i \leftarrow i_{\min}$ **to** $i_{\max}$ **do**
    $\alpha'_i = \alpha_i + \frac{1}{\sqrt{\lambda_i}}((S - S_{scan})^T \mathbf{S})$
  update $S = \sum_{i=1}^{m} \alpha'_i \sqrt{\lambda_i} \cdot \mathbf{S}$
return $d_{rms}(S, scan)$

The *eigenspace sampling* iteratively selects model coefficients, that morphs the face model closer to the scan data. This algorithm simply tries four new coefficients for each sequential eigenvector $\mathbf{s_i}$, and keeps the one that produces the smallest RMS distance. By reducing the search space α_{range} in each iteration, the algorithm produces a more accurate fit in each of the iterations. Because the first eigenvectors induce the fitting of global face properties and the last eigenvectors change local face properties, each iteration follows a global to local fitting scheme. To avoid local minima in face space, we try four new coefficient values in each iteration and we use in following iterations a slightly larger range α_{range} than the latest increment α_{incr}.

The *eigenspace projection* refines the set of correspondences that are selected with the eigenspace sampling method. Before the projection, we have to establish $n' \le n$ closest point correspondences from instance $S_{expr.inst}$ to the scan data, where each point-pair describes the direction to which the vertex of $S_{expr.inst}$ should move for a tighter fit. The number of correspondences n' is usually smaller than the number of vertices of the morphable face model n, because we use a multi-resolution scheme and not every vertex in the model has a closest compatible point in the scan data. The sets of n' correspondences $S_{expr.inst}$ and S_{scan} are subtracted and in case $n' < n$, the missing correspondences are replaced with zero vectors to retain full correspondence with the morphable model. The residual deformation vector can be projected either into the eigenspace of the expression model or into the eigenspace of identities:

$$\beta'_i = \beta_i + \frac{1}{\sqrt{\mu_i}}((S_{expr.inst} - S_{scan})^T \mathbf{e_i}),$$

$$\alpha'_i = \alpha_i + \frac{1}{\sqrt{\lambda_i}}((S_{expr.inst} - S_{scan})^T \mathbf{s_i}).$$

Projecting the residual deformation vector onto the eigenvectors of the orthogonal eigenspace is the fastest and easiest way to obtain the least-squares solution for the given set of correspondences (Section 3.4). Afterwards, a new set of closest point correspondences can be selected and the residual deformation is projected into the eigenspaces again. This process converges to a local optimum within a few iterations ($k_{\max}$). As a result, the model coefficients are refined.

4.2. Expression fitting

The main difficulty in model fitting is that neither the expression coefficients nor the identity can be optimized without optimizing the other. When the model is fitted to a new scan with an unknown expression, it makes sense to coarsely estimate the identity based on expression invariant regions and then to select the best expression model and search for its optimal expression parameters.

Starting with the morphable mean face $\bar{S}$, the identity coefficients $\boldsymbol{\alpha}$ are improved using algorithms ESSamp and ESProj based on the static face component (Fig. 4). The former algorithm iteratively improves coefficients ($i_{\min}$,$i_{\max}$) α_1 up to α_4, α_5 up to α_8, and α_9 up to α_{12}. To cover a large range of facial variety, we use a large range of coefficients $\alpha_{range} = 2$, and $k_{\max}$=4 iterations. The established correspondences are refined with algorithm ESProj to obtain the coarsely fitted face instance S_{coarse}.

To find the expression instance $S_{expr.inst}$ that fits the scan data best, we need to find the optimal combination of identity and expression coefficient vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Moreover, to select the appropriate expression model, we need an optimized fit for each of the expression models in combination with S_{coarse}. For that, we select an expression model and three different coefficients $\beta_1 = \{0.0, 0.5, 1.0\}$ for its first expression vector $\mathbf{e_1}$ and apply this deformation to S_{coarse}. Note that this first expression vector causes the largest shape deformation, and that its weight should be positive (Section 3.7). Starting from each of these instances $S_{expr.inst}$, the coefficients β_2 up to β_6 are refined with algorithm ESSamp using $\alpha_{range} = 2$ and $k_{\max}$=4. Then α_1 to α_4, β_1 to β_6, α_5 to α_8, and α_9 to α_{12} are refined with algorithm ESSamp using $\alpha_{range} = \frac{1}{2}$ and $k_{\max}$=4.

For each expression model and for each coefficient β_1, a combined identity/expression fit is acquired with ESSamp. Each of these coarse fits is then projected onto the eigenspace of expressions and then to the eigenspace of identities to refine the established correspondences with all model coefficients in algorithm ESProj ($k_{\max}$=1). The best fit, i.e. the instance with the smallest d_{rms} distance, is selected as the *best coarse fit* as shown in Fig. 5. To further refine the expression parameters $\boldsymbol{\beta}$ of the selected fit, the residuals are projected onto the expression space with algorithm ESProj using $k_{\max}$=5 iterations. This gives us the final expression vector S_{expr}.

4.3. Identity fitting

After the expression fitting, we have obtained a coarse identity vector that, in combination with the final expression vector, produces a relatively good fit to the scan data. For the purpose of face recognition, each subject needs a unique expression invariant identity vector $\boldsymbol{\alpha}$. Amberg et al. [1] proposed to produce the best possible fit and to use the decoupled identity vector for face recognition. In [18] a more accurate fit was produced by fitting predefined face components individually. Here, we use both methods and propose a new descriptor.

To produce the best possible fit for the entire face as one component, we use algorithm ESProj to refine the identity coefficients in $k_{\max}=5$ iterations. This gives us the final identity vector S_{id} and its coefficient vector $\boldsymbol{\alpha}$. This *single-component vector* is used as feature vector for the face matching.

For the multiple component method, we define a subset of vertices for the nose, eyes, mouth and remainder regions and project the residual vector for each component to the identity space using ESProj. This gives us an identity vector S_{id} and a coefficient vector $\boldsymbol{\alpha}$ per component. The coefficient vectors are concatenated to produce a single feature vector for the face matching, which we refer to as the *multi-component vector*. The identity vector S_{id} can be used to bootstrap the model directly, or its facial landmarks can be mapped to the scan data and used to model a new set of correspondences as we did in Section 3.

Since there is no guarantee that the fitting process finds the global optimum in the identity space, we propose to search for a number of local optima and concatenate their coefficient vectors $\boldsymbol{\alpha}$. These locations in m_s-dimensional space should form a unique pattern for each subject usable for face recognition. To find four local minima, we initiate a face instance as the combination of the mean face with the final expression vector ($\overline{S}+S_{expr}$), and adjust its first two coefficients α_1 and α_2 with $\{-2,2\}$. Then algorithm ESSamp is applied using a relatively small $\alpha_{range}=\frac{1}{2}$, and $k_{\max}=4$, to iteratively refine coefficients α_1 up to α_{12} as we did before. Each of the four coarsely fitted $S_{expr.inst}$ is then refined using algorithm ESProj in $k_{\max}=5$ iterations (also without $\boldsymbol{\beta}$). This gives us an identity vector S_{id} and its coefficient vector $\boldsymbol{\alpha}$ for each of the four initializations. The coefficient vectors are concatenated to produce a single feature vector for the face matching, which we refer to as *multi-minima vector*. More local minima can be used at the cost of a larger feature vector.

4.4. Implementation

The difficulty in model fitting is the high dimensional face space in which each location represents a detailed face mesh. Exhaustive search for the global optimum is an intractable task and scan deficiencies and expressions cause local minima in the face space. During the expression model fitting (Section 4.2), we iteratively fit each expression model using three different values β_1 to be able to select the proper intensity of the expression. Algorithm ESSamp, iteratively improves a small set of coefficients (e.g. α_1 up to α_4) at a time, which allows the algorithm to recover from an incorrect choice for a coefficient at an early stage. Refining either one coefficient iteratively at a time or all coefficients per iteration are two variants that performed less well. For the selection of the best expression model, we fit each model in combination with no more than twelve coefficients of the identity model. With all identity coefficients, several expression models may produce a tight fit to the scan data, so with limited resources the distance d_{rms} becomes more reliable.

Mpiperis et al. [14] also experienced that especially the mouth region causes local minima in the face space that require additional effort to avoid. One common local minimum is a model instance with a closed mouth while the face scan shows an open mouth. Instead of a dedicated mouth detection algorithm that Mpiperis proposed, we allow the model's vertices in the mouth area to pair up with boundary points of the scan data. In case the face scan has an open mouth and the face model has not, these point pairs are automatically penalized by the distance measure d_{rms}.

For the ESProj algorithm we use closest point-to-point correspondences from the model to the scan data and from the scan data to the model. This results in a higher fitting accuracy [14]. With the use of a kD-tree the closest point correspondences can be found efficiently. For high efficiency, we compute in algorithm ESSamp only the correspondences from model to scan, because the model and its kD-tree change in each iteration. For the eigenspace sampling we consider a close point-pair to be valid if their distance is smaller than 50 mm, for the eigenspace projection we use a distance of 10 mm. We stop traversing a kD-tree, when this criterion can no longer be met.

For time efficiency, algorithm ESSamp is applied using the low resolution face model of $n=932$ vertices. Algorithm ESProj is applied to a coarse fit using $n=3624$ vertices and to the final expression and identity vectors using $n=14\,288$ vertices. In the end, the time to process a raw scan requires ca. 3 s for the face segmentation, ca. 10 s to fit all expression models, less than 1 s to improve the coarse identity fit, and ca. 4 s to find four local minima on a Pentium IV 2.8 GHz. Note that the fitting of each expression model as well as the search for multiple minima can be done in parallel to further speed up the process.

5. Face matching

We automatically fit the morphable model to all scans in the UND, GAVAB, BU-3DFE, and FRGC v.2 datasets. Note that a small subset of the BU-3DFE and the FRGC v.2 were added to the models. After the fitting we have obtained three feature vectors of model coefficients, namely, the single-component vector, the multi-component vector, and the multi-minima vector. For the face matching we use each of these vectors individually to do 3D face recognition. To determine the similarity of faces with these coefficient vectors, we use the L_1 distance between the normalized coefficient vectors. So, the matching of two faces requires only the comparison of either m_s or $4\,m_s$ float values, which is extremely time-efficient. For each query, we compute its similarity to all other models in the training set, generating a ranked list of face models sorted on decreasing similarity values in the process. Based on these ranked lists, we compute the recognition rate (RR), the mean average precision (MAP), and according to the FRGC benchmark the verification rate at 0.1% false acceptance rate (VR@0.1%FAR) [18,15]. The VR@0.1%FAR is a measure for the authentication scenario, because it distinguishes between clients to accept and imposters to rejected.

6. Results

6.1. Morphable model fitting

In this section we evaluate the accuracy of the final expression instances $S_{expr.inst}$ both quantitatively and qualitatively. After the identity fitting in Section 4.3 we have two final face instances, a single component fit (SC) and a multiple component fit (MC).

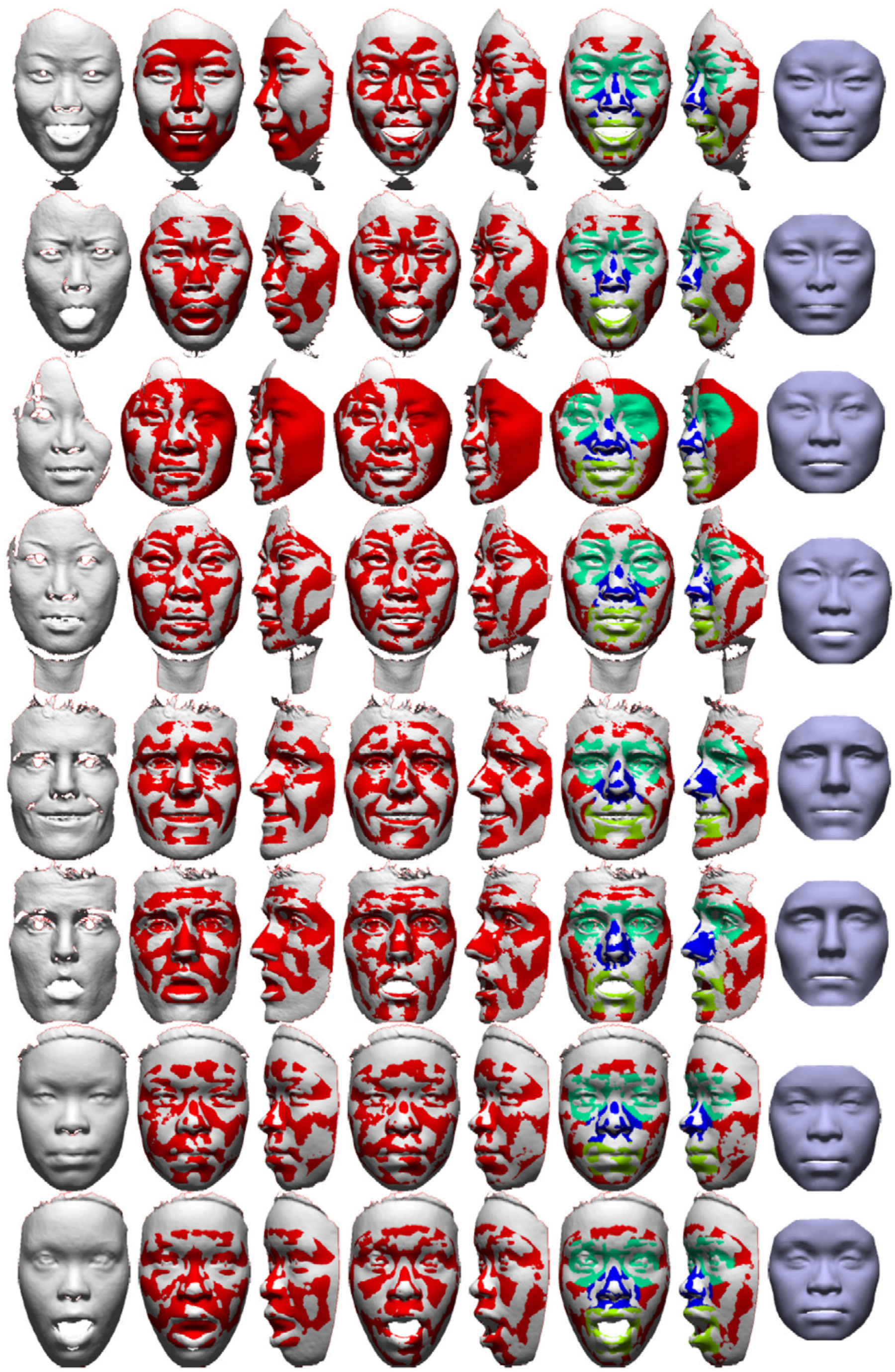

Fig. 8. Model fitting to processed scans (1st column) using the neutral model only (2nd and 3rd column), a single component (4th and 5th column) and multiple components (6th and 7th column). The last column shows the neutralized face instances $S_{expr.inst} - S_{expr}$ of the 6th column.

We can evaluate the fits qualitatively by looking at the more frequent surface interpenetration of the fitted model and face scan (Fig. 8), which means a tighter fit. Note that our fitting method is robust to missing data and even creates accurate face instances when half of the face is missing. A quantitative evaluation can be done by comparing the residual d_{rms} distances. Table 1 shows a decrease in residual error for the multiple components.

To prove that the expression modeling improves the fitting process, we also fitted the neutral model as a whole without the additional expression models. These fitting results are shown in the 2nd and 3rd column of Fig. 8, which show a consistent failure in case of expression scans. The higher residual d_{rms} distances are listed in Table 1 (SC neutral).

6.2. Face matching

After the fitting process, the expression deformation S_{expr} can be subtracted to neutralize the expression, $S_{expr.inst} - S_{expr} = \overline{S} + S_{id}$. The identity coefficients $\boldsymbol{\alpha}$ that model the identity deformation S_{id} are used for the face matching as explained in Section 5. For perfect retrieval results, the acquired coefficient vector $\boldsymbol{\alpha}$ for each scan of the same subject should point to the same unique position in the m_s-dimensional coefficient space. To get an impression of the identity clustering in coefficient space, we show in Fig. 9 ten random subjects of the UND dataset having more than four scans. The projected coefficient vectors are those acquired with the single component fit. These graphs show that not only the principal eigenvectors are useful to distinguish between different subjects, but that even the fiftieth coefficient contributes to the clustering of subjects.

The face matching results of the single-component (SC), multi-component (MC), and multi-minima (MM) vectors are listed in Table 2. Results show that (1) our method can be applied with considerable success to a large range of datasets, (2) the use of expression models is essential for high performance, (3) the use of multiple components (MC) improves on the single component (SC) results, (4) in case of scan data with lower quality, as in the GAVAB dataset, the multiple minima (MM) approach can improve the system's performance. In Fig. 10, we show the *cumulative match characteristic* (CMC) curves for both the multiple component and multiple minima results on the four datasets. Notice that the recognition rates for the BU-3DFE are very high, this is because for each face there is at least one other face that is very similar due to the slightly different 'expression intensities' in this set.

Comparison UND. Several authors report recognition rates for the UND dataset. Blanz et al. [4] achieved a 96% RR for 150 queries in a set of 150 faces. Amberg et al. [1] used all 953 scans and achieved 100% RR.

Comparison GAVAB. The GAVAB dataset has been used in the Shape Retrieval Contest 2008 [17] to compare 3D face retrieval methods. Results of different approaches vary between 60% and 100% RR. Recently, Amberg et al. [1] achieved a recognition rate of 99.7% on this dataset. They use a morphable head model that covers the neck and ears as well, features that may aid the person identification.

Comparison BU-3DFE. Mpiperis et al. [14] performed experiments on the BU-3DFE dataset. They used two methods for the expression invariant face matching, a symmetric bilinear model and geodesic polar coordinates, with respectively 86% and 84% RR.

Comparison FRGC Lu et al. [11] applied their expression-specific deformation models to only 100 subjects of the FRGC v.2 and report 92% recognition rate and 0.7 VR@0.1%FAR, which is considerably lower than the results with our expression-specific deformation models. Moreover, we do not need a neutral face scan for the deformation nor the computational expensive ICP algorithm for the matching. Other 3D shape based methods that report the VR@0.1%FAR for the all-to-all face matching experiment are, Mian et al. [12] with 0.87 VR, Cook et al. [6] with 0.92 VR, and Faltemier et al. [8] with 0.93 VR. Most of them use the computational expensive ICP algorithm during face matching and simply neglect data in regions with expressions. Kakadiaris et al. [9] reported a 97% RR and 0.97 VR@0.1%FAR for slightly different experiments.

Table 1
The residual RMS error (mm) is determined for each model fit to its scan.

Dataset	Fit	min	max	mean	sd
UND	SC neutral	0.64	2.34	0.75	0.13
	SC	0.63	2.32	0.75	0.12
	MC	0.62	2.27	0.72	
GAVAB	SC neutral	0.70	1.98	0.89	0.16
	SC	0.70	1.53	0.85	0.10
	MC	0.67	1.36	0.78	0.08
BU-3DFE	SC neutral	0.60	2.32	0.79	0.14
	SC	0.60	1.12	0.71	0.06
	MC	0.59	0.99	0.67	0.05
FRGC v.2	SC neutral	0.64	3.55	0.81	0.19
	SC	0.65	3.39	0.79	0.17
	MC	0.63	3.40	0.75	0.16

The most accurate (min) and least accurate (max) fit, the mean and the standard deviation are reported for each dataset.

Table 2
All to all face matching.

Dataset	Fit	RR	MAP	VR@0.1%FAR
UND	SC neutral	0.99	0.99	0.99
	SC	0.99	0.98	0.97
	MC	0.99	0.98	0.98
	MM	0.99	0.99	0.97
GAVAB	SC neutral	0.94	0.80	0.53
	SC	0.97	0.90	0.73
	MC	0.97	0.92	0.77
	MM	0.98	0.93	0.80
BU-3DFE	SC neutral	0.98	0.59	0.29
	SC	1.00	0.91	0.80
	MC	1.00	0.92	0.82
	MM	1.00	0.91	0.75
FRGC v.2	SC neutral	0.91	0.80	0.73
	SC	0.97	0.89	0.84
	MC	0.97	0.91	0.87
	MM	0.97	0.91	0.82

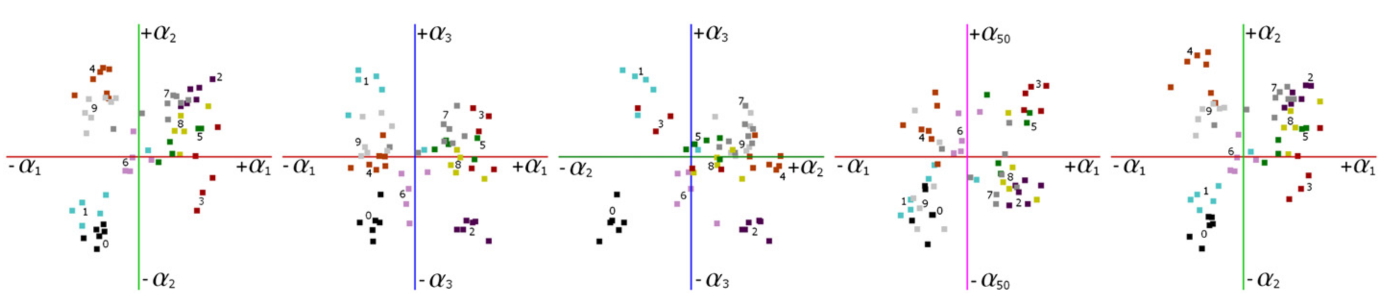

Fig. 9. Identity clustering in coefficient space. Each coefficient aids the classification of subjects (shown in color). The last graph shows the projection onto α_1 and α_2 after normalizing the length of all vectors $\boldsymbol{\alpha}$.

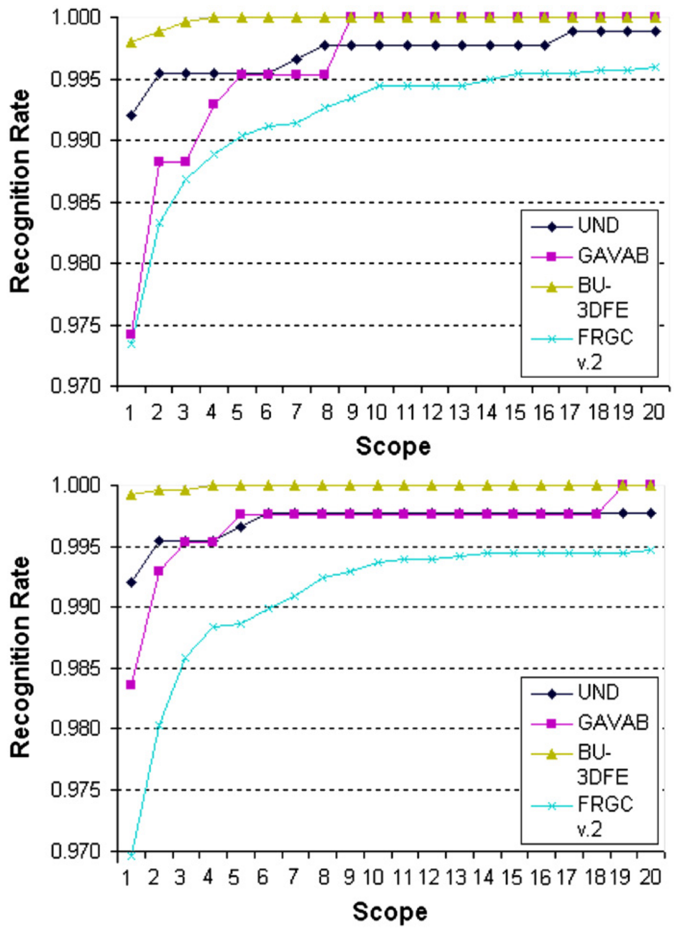

Fig. 10. The cumulative match characteristic curves for the multiple component (MC) results (top) and mulitple minima (MM) results (bottom) on the four datasets.

7. Conclusion

We presented a complete 3D expression invariant face recognition system. Starting from pose normalized face scans, we proposed an easy to implement method to semi-automatically build identity and expression models from the BU-3DFE dataset. With the presented model fitting algorithm, we can automatically establish full correspondence with new scan data and bootstrap the identity and expression models. Statistical face models provide strong shape priors that allow for the robust handling of noise and holes. Results show that our method can be effectively used for landmark extraction, bootstrapping, face completion, and face matching, which is an advantage over other methods.

The method that we presented, coarsely fits the identity model in combination with each of the expression models and keeps the overall best fit. Because separate models are used for the identity and expression deformations, an expression can be neutralized and the separate identity coefficients used for the (expression invariant) face matching. Three identity coefficient vectors were acquired for the face matching, one based on the face as a single component, one for multiple face components, and one for multiple local minima. Compared to the literature, all our coefficient vectors perform very well on the publicly available datasets. Our system effectively recognizes faces with expressions from several data sets, and is also very time-efficient: After the model is fitted to each scan in at most 17 s (linear to the number of scans), our face matching (quadratic to the number of scans) requires only the comparison of either 80 or 320 float values in our experiments. Therefore, our method can be very well applied to authentication scenarios (e.g. airport check-ins) as well as face retrieval scenarios (e.g. searching criminal records).

References

[1] Amberg B, Knothe R, Vetter T. Expression invariant face recognition with a morphable model. In: Automatic face and gesture recognition, 2008.
[2] Besl PJ, McKay ND. A method for registration of 3D shapes. PAMI 1992;14(2):239–56.
[3] Binghamton University 3D Facial Expression Database. BU-3DFE.
[4] Blanz V, Scherbaum K, Seidel H-P. Fitting a morphable model to 3D scans of faces. In: ICCV, 2007. p. 1–8.
[5] Chang KI, Bowyer KW, Flynn PJ. An evaluation of multimodal 2D+3D face biometrics. PAMI 2005;27(4):619–24.
[6] Cook J, Chandran V, Fookes C. 3D face recognition using log-gabor templates. In: BMVC, 2006. p. 83–93.
[7] Cootes T, Edwards G, Taylor C. Active appearance models. PAMI 2001;23(6): 681–685.
[8] Faltemier T, Bowyer K, Flynn P. A region ensemble for 3-D face recognition. Transactions on Information Forensics and Security 2008;1(3):62–73.
[9] Kakadiaris IA, Passalis G, Toderici G, Murtuza MN, Lu Y, Karampatziakis N, Theoharis T. Three-dimensional face recognition in the presence of facial expressions: an annotated deformable model approach. PAMI 2007;29(4): 640–649.
[10] Lay DC. Linear algebra and its applications. Addison-Wesley; 1997.
[11] Lu X, Jain A. Deformation modeling for robust 3D face matching. PAMI 2008;30(8):1346–56.
[12] Mian AS, Bennamoun M, Owens R. An efficient multimodal 2D-3D hybrid approach to automatic face recognition. PAMI 2007;29(11):1927–43.
[13] Moreno A, Sanchez A. GavabDB: a 3D face database. In: Workshop on biometrics on the internet COST275, Vigo, 2004. p. 77–85.
[14] Mpiperis I, Malassiotis S, Strintzis MG. Bilinear models for 3-D face and facial expression recognition. Transactions on Information Forensics and Security 2008;3(3):498–511.
[15] Phillips PJ, Flynn PJ, Scruggs T, Bowyer KW, Worek W. Preliminary face recognition grand challenge results. In: Automatic face and gesture recognition, 2006. p. 15–24.
[16] Samir C, Srivastava A, Daoudi M, Klassen E. An intrinsic framework for analysis of facial surfaces. IJCV 2009;82(1):80–95.
[17] ter Haar FB, Daoudi M, Veltkamp RC. SHape retrieval contest 2008: 3D face scans. In: Shape modeling and applications (SMI), 2008. p. 225–6.
[18] ter Haar FB, Veltkamp RC. 3D face model fitting for recognition. In: European conference on computer vision (ECCV), 2008. p. 652–64.
[19] Turk M, Pentland A. Face recognition using eigenfaces. In: CVPR, 1991. p. 586–91.
[20] University of South Florida, Sudeep Sarkar. USF HumanID 3D face database.
[21] Yin L, Wei X, Sun Y, Wang J, Rosato MJ. A 3D facial expression database for facial behavior research. In: Automatic face and gesture recognition, 2006. p. 211–6.