Fast and Accurate 3D Face Recognition

Using Registration to an Intrinsic Coordinate System and Fusion of Multiple Region Classifiers

Luuk Spreeuwers

Abstract In this paper we present a new robust approach for 3D face registration to an intrinsic coordinate system of the face. The intrinsic coordinate system is defined by the vertical symmetry plane through the nose, the tip of the nose and the slope of the bridge of the nose. In addition, we propose a 3D face classifier based on the fusion of many dependent region classifiers for overlapping face regions. The region classifiers use PCA-LDA for feature extraction and the likelihood ratio as a matching score. Fusion is realised using straightforward majority voting for the identification scenario. For verification, a voting approach is used as well and the decision is defined by comparing the number of votes to a threshold. Using the proposed registration method combined with a classifier consisting of 60 fused region classifiers we obtain a 99.0% identification rate on the all vs first identification test of the FRGC v2 data. A verification rate of 94.6% at FAR $= 0.1$% was obtained for the all vs all verification test on the FRGC v2 data using fusion of 120 region classifiers. The first is the highest reported performance and the second is in the top-5 of best performing systems on these tests. In addition, our approach is much faster than other methods, taking only 2.5 seconds per image for registration and less than 0.1 ms per comparison. Because we apply feature extraction using PCA and LDA, the resulting template size is also very small: 6 kB for 60 region classifiers.

Keywords 3D Face recognition · Registration · Fusion · Region classifiers · FRGC

L. Spreeuwers (✉)
Chair of Signals and Systems, Department of EEMCS, University of Twente, Twente, The Netherlands
e-mail: l.j.spreeuwers@utwente.nl

1 Introduction

3D face recognition has made much progress during the last decade. Both in the area of 3D face acquisition as well as in 3D face matching significant steps were made. Currently, a wide range of sensors for 3D face acquisition is available, mostly based on laser scanning and structured light techniques. Many 3D face recognition approaches use the distance between the aligned facial surfaces as a measure of how well faces match. To align the 3D facial shapes, nearly all state-of-the-art 3D face recognition methods minimise the distance between two face shapes or between a face shape and an average face model. This process of aligning facial shapes to a common coordinate system is called registration. In contrast to registration to a second or an average face shape, we present an approach that registers 3D facial shapes to an intrinsic coordinate system of the face, defined by 3D landmark structures. For classification we use the fusion of many regional likelihood ratio based classifiers and PCA-LDA to extract compact feature vectors. Registration to an intrinsic coordinate system has received little attention since the early days of 3D face recognition due to lack of success. In this paper, we show, however, that excellent results can be obtained if the registration is sufficiently robust and accurate. Below we briefly outline the basics and advantages and disadvantages of the different approaches. A popular method to align two faces is the Iterative Closest Point (ICP) algorithm, Besl and McKay (1992). In this approach, two 3D point clouds, representing the surfaces of two different faces, are registered to each other by minimising the distance between the surfaces in an iterative process. The distance between the surfaces is calculated by finding the closest point in the second point cloud for each of the points in the first point cloud and taking the average of all these distances. The distance between the surfaces is minimised by rotating and translating one of the point clouds

relative to the other. The resulting distance measure is then used for face matching. Many of the top ranking papers on 3D face recognition of the last 5 years are based on ICP-like approaches: Faltemier et al. (2008a), Kakadiaris et al. (2007), Maurer et al. (2005), Mian et al. (2007), Queirolo et al. (2010). Queirolo et al. (2010) actually do not use ICP, but Simulating Annealing to obtain a closest fit between two point clouds. The ICP approach and Queirolo's approach, however, have several major disadvantages. Since the point clouds (or other surface representations) are used in the matching process directly, the only way to store the templates is to store the whole point cloud. Firstly, this requires much more space than normally is reserved for biometric templates (a point cloud of 50.000 vertices requires in the order of 600 kB). Secondly, it prevents the use of privacy protecting techniques aimed at the impossibility to be able to reconstruct the original biometric data based on the template. A third disadvantage is the fact that ICP is relatively slow, generally taking several seconds for registration and calculation of the distance measure. This is not necessarily a problem in the verification scenario where only two images must be compared, but it is a problem in the identification scenario where a probe image is compared to a gallery of many images. Therefore the ICP approach is not very fit for identification, as is also pointed out in Faltemier et al. (2008a) and Queirolo et al. (2010), who incidentally report the highest 3D face identification rates.

The approach we propose in this paper, does not register two point clouds to each other, but transforms each point cloud to an intrinsic coordinate system of the face. This reference coordinate system is based on the vertical symmetry plane of the face and the tip and orientation of the nose. The point cloud is then resampled into a range image from which features are extracted using PCA and LDA. The features form a template that is far more compact than a complete point cloud. The likelihood ratio is used as a similarity measure. Like in many other approaches, see e.g. Faltemier et al. (2008a) and Queirolo et al. (2010), we divide the facial surface into parts that are more or less stable under variation of facial expressions. We found that using multiple overlapping regions and combining them with a simple decision level fusion approach using voting, gives excellent robustness against variations in facial expression.

The proposed approach has some major advantages over ICP-like approaches. Firstly, since we do not register two point clouds to each other for each match, but use an independent registration and store templates consisting of extracted features, in the identification scenario, where one image is compared to many images in a list, we save many registrations. If the list contains N entries, for the ICP-like approaches, N registrations must be performed for each probe image. In our case only a single registration is required, because all gallery probes are pre-registered and only the templates are stored. The face matching using the PCA LDA likelihood approach on two templates is extremely fast and allows for many thousands of comparisons per second. Secondly, because the coordinate system of the face is fixed, this could be standardised. Thirdly, unlike to the point clouds, to the templates we store, biometric template protection techniques can be applied, see e.g. Buhan et al. (2010), Kelkboom et al. (2009, 2010), Chen et al. (2009). This means that the "encrypted" templates cannot be traced back to the original 3D data (or templates) and the matching takes place in the protected domain. Privacy protection of biometric data is an ever increasing concern, so this is a very useful property of our approach. Finally, our approach provides excellent recognition results, besting the highest published identification results and ranking between the highest verification scores. The performance was evaluated using the Face Recognition Grand Challenge (FRGC) benchmarking for 3D face recognition (Phillips et al. 2005). In this benchmark a challenging database consisting of 4007 images of 466 subjects with varying facial expression is used. Summarising, we present a 3D face recognition approach that is superior both in speed and recognition performance relative to other methods and has the additional advantage of being better fit for biometric template protection.

This paper is organised as follows. Section 2 presents an overview of related work. In Sect. 3 the registration method is described in detail. Section 4 describes the PCA-LDA-likelihood ratio classifier. In Sect. 5 the region classifiers are defined and the used decision level fusion approaches are explained for both the identification and the verification cases. Section 6 contains experiments and results and a description of the used 3D facial data. Finally, Sect. 7 gives conclusions.

2 Related Work

This section on related work consists of two parts. The first part addresses related work on 3D face registration. The second part concentrates on 3D face recognition, i.e. the classification or comparison of 3D face images or extracted features. In practise, the two are often tightly interwoven, like in e.g. the ICP approach.

2.1 3D Face Registration

Registration basically means transforming shapes in such a way that they can be compared. For 2D face recognition, e.g. it is common to locate a number of landmarks (e.g. eyes, nose, mouth) in each face and rotate, translate and scale these landmarks in such a way that they are projected to fixed, predefined positions. The same geometric transformation is then applied to the facial image. The facial image is thus transformed to an *intrinsic coordinate system*. Once the images are represented in this intrinsic coordinate system, they can be compared, because corresponding features
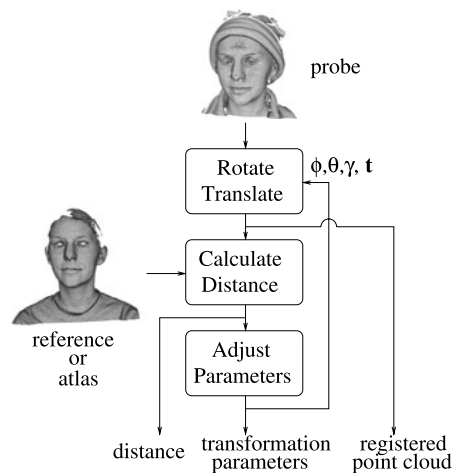
Fig. 1 Iterative registration of one 3D point cloud to a reference point cloud

are more or less in the same positions in the different facial images.

Basically three different approaches to 3D face registration can be distinguished:

– One-to-all registration (register one face to another)
– Registration to a face model or atlas
– Registration to an intrinsic coordinate system using geometric properties of the face like landmarks

Apart from this division in three classes, we can also distinguish rigid and non-rigid registration. The former only performs rotation and translation (and possibly scaling) of the point clouds. The latter also allows for (small) deformations of the point cloud to realise an optimal registration. Non-rigid registration can be useful in handling facial expressions. Using non-rigid registration, e.g. a smiling mouth can be fitted to a neutral mouth etc. which is impossible for rigid registration.

The first approach: one-to-all registration (see Fig. 1) registers two surfaces or point clouds to each other using an iterative procedure. One of the point clouds is the reference (from a gallery) while the other is the probe. The aim of this registration approach is to find rotation and translation parameters that will transform the probe point cloud to lie as close as possible to the reference point cloud. To this end, a distance measure must be defined between the two point clouds. Examples of such distance measures are the Mean Square Error (MSE) between the surfaces and the Surface Interpenetration Measure (SIM), see Silva et al. (2005), Queirolo et al. (2010). Based on the distance between the point clouds (or the change in distance due to a change in the registration parameters) the registration parameters $(\theta, \phi, \gamma, t)$ are updated and the probe is transformed again etc. This process continues for a number of iterations until convergence is reached. As a result, the registration parameters, the transformed probe and the resid-

ual distance between the two point clouds become available for further processing. The Iterative Closest Point (ICP) approach is the most popular method for this optimisation process of aligning one point cloud to another. Generally, a reasonably good initial estimate of the registration parameters $(\theta, \phi, \gamma, t)$ is required to obtain convergence. Usually landmarks like the tip of the nose and sometimes the vertical symmetry plane are used to obtain this initial estimate. Examples of one-to-all registration are Maurer et al. (2005), Mian et al. (2007), Queirolo et al. (2010), Faltemier et al. (2008b). All of these address only rigid registration. As pointed out in Sect. 1, one-to-all registration has the disadvantage that a probe must be registered to all images in the gallery. Because the iterative registration procedure generally is quite time-consuming, this makes application to an identification scenario (one-to-many) impractical. For a verification scenario (one-to-one), only a single registration is required, so a somewhat slower registration is entirely acceptable.

The second approach: registration to a model or atlas basically operates in the same way, however, the probe image is not registered to a gallery image, but to a model or atlas (see Fig. 1). The model or atlas is learnt from a training set. Examples of this approach are Kakadiaris et al. (2007), Gokberk et al. (2006), Salah et al. (2007), Boehnen et al. (2009), Alyüz et al. (2009). In Kakadiaris et al. (2007) and Gokberk et al. (2006) also non-rigid registration is explored. In all these articles the Average Face Model (AFM) is built from training examples. A significant advantage relative to the one-to-all approach described above, is that each image has to be registered only once. This means images in the gallery can be pre-registered and application in an identification scenario becomes possible. A disadvantage is that probes may be less accurately registered to an average face model than to an image of the same subject.

The third approach: registration to an intrinsic coordinate system using e.g. landmarks, requires the accurate localisation of 3D landmarks on the face. The set of 3D landmarks is mapped on the corresponding 3D landmarks in the intrinsic coordinate system. The resulting transformation is then also applied to the complete point cloud of the face, resulting in the registered point cloud (see Fig. 2).

A problem is that most 3D landmarks are not stable under facial expressions and/or can be covered by hair or occluded by other parts of the face. Landmark based registration is discussed in some depth in Papatheodorou and Rueckert (2007). Registration to an intrinsic coordinate system has the same advantages as registration to an atlas or model: each image has to be registered only once. An added advantage is that the intrinsic coordinate system can be precisely defined and standardised. Because atlases and AFM's are obtained using training sets, basing a standard on these models is hardly possible. Tang et al. (2008) present a registration method to an intrinsic coordinate system based on the

Fig. 2 Registration using 3D landmarks on the face
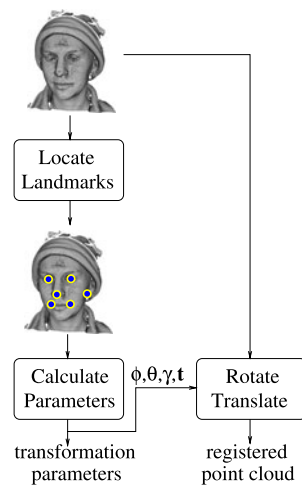
vertical symmetry plane of the face, the tip of the nose and the slope of the nose bridge. These could be called *landmark structures* in the image as opposed to landmarks, which only mark positions. The advantage of using the symmetry plane and the nose tip and bridge is that these features are relatively stable under facial expression variations, while they still completely define a 3D intrinsic coordinate system (see also Fig. 4). Our approach, as presented in this paper, is based on the same features: the vertical symmetry plane, the location of the tip of the nose and the slope of the bridge of the nose (see Fig. 3). However, we take a robust approach to determine these which, together with a more advanced 3D face classifier, results in far better recognition rates (see Sect. 6). Furthermore, we present more results on a far larger database and compare our results with the state of the art, which Tang et al. do not.

It is interesting that most of the best performing approaches to 3D face recognition are based on one-to-all registration and registration to an atlas or model, mostly using ICP, while on the other hand for 2D face recognition landmark based methods are more common. In Boom et al. (2007) and Spreeuwers et al. (2007) we proposed approaches to 2D one-to-all registration and registration to an AFM registration and showed significant advantages over landmark based approaches. Ironically, here we present a landmark (structures) based approach to 3D face recognition and show significant advantages over 3D one-to-all registration and registration to an AFM.

2.2 3D Face Recognition

A recent overview on 3D face recognition until 2006 is presented in Bowyer et al. (2006). Other reviews are presented in Papatheodorou and Rueckert (2007), Scheenstra et al. (2005). More recent work was covered in Faltemier et al. (2008a), Queirolo et al. (2010), Boehnen et al. (2009), Alyüz et al. (2009). Since these give an extensive overview of work
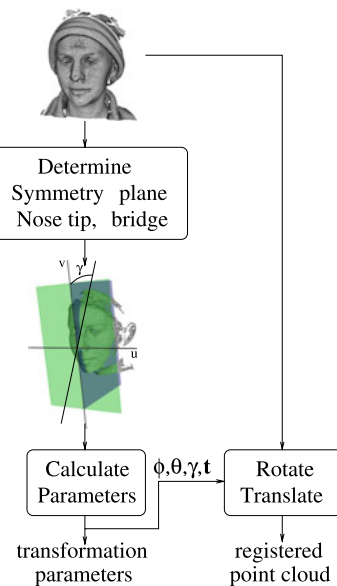

Fig. 3 Registration using vertical symmetry plane, nose tip and the slope of the bridge of the nose

on 3D face recognition, in this section only a brief summary is presented and the reader is referred to the above papers for more details.

Early work on 3D face recognition started around 1989 using profile and minimum distance between surfaces approaches (Cartoux et al. 1989) and e.g. application of PCA to range images (Achermann et al. 1997; Hesher et al. 2003). One of the problems was that in the beginning only small datasets were available and there was no unified approach to comparing performance of the different 3D face recognition methods.

In 2004, the Face Recognition Grand Challenge (FRGC) data (Phillips et al. 2005) was released containing in total 4950 images of 466 persons and the definition of a number of experiments for evaluation, among which a number of verification experiments and identification experiments, normally using 4007 of the 4950 images. The FRGC dataset also contains many images with various expressions. Unfortunately, the FRGC database contains a number of images with serious motion artifacts, acquisition errors and extreme expressions, which might be rejected for classification in actual situations.

As described in the previous section on 3D face registration, ICP can be used to align 3D point clouds. Apart from the aligned point clouds ICP also produces a measure for the distance between the facial surfaces if they are aligned. This measure can be used as a matching criterion, because the distance between aligned 3D point clouds of two different individuals will be larger than between two different aligned point clouds of a single individual. The use of the iterative closest point (ICP) approach started around 2003 (Medioni and Waupotitsch 2003) and because it was very successful

it has dominated the world of 3D face recognition since. Because ICP only works properly if the two point clouds are already quite close to each other, generally a form of pre-registration is performed and often the data are cleaned somewhat: noise is suppressed and spikes are removed. Improvements of the ICP approach using several regions in the face that were more or less sensitive to expressions and modified distance measures were published in e.g. Maurer et al. (2005), Mian et al. (2007), Queirolo et al. (2010), Faltemier et al. (2008b). A major drawback of the ICP approach to 3D face comparison, is that it is a slow method, generally taking several seconds to minutes per comparison. For the verification scenario, where only two images have to be compared, this may still be acceptable, but for the identification scenario, where a single probe must be compared to all gallery images, it is not a practical solution. As described in Sect. 2.1 another approach is to register to an average face model (AFM) using ICP and then extract features which are used for the classification. In this case, ICP has to be performed only once and more compact templates of the faces can be stored for the gallery images. This approach with registration to an AFM is used in Kakadiaris et al. (2007), Gokberk et al. (2006), Alyüz et al. (2009), Papatheodorou and Rueckert (2005).

In recent work (Mian et al. 2007; Kakadiaris et al. 2007; Gokberk et al. 2006; Alyüz et al. 2009; Faltemier et al. 2008b; Maurer et al. 2005; Queirolo et al. 2010), generally performance comparison to the state-of-the-art is done using the FRGC database (often in addition to other databases). Two of the most challenging tests that are most cited in publications are an all vs all verification test, resulting in a score matrix of 4007×4007 and a closed set identification test using a gallery consisting of the first images of all 466 subjects and the rest of the 4007 images as probes. For the former the recognition rate at a false accept rate of 0.1% is reported, while for the latter the rank-1 recognition rate is reported. On the all vs. all verification test, currently the best performance ranges from 93.2% (Faltemier et al. 2008b) to 97% (Kakadiaris et al. 2007). For the closed set identification test, the best rank-1 results reported were 98.4% (Queirolo et al. 2010). Our approach using rigid registration to an intrinsic coordinate system and multiple region PCA-LDA likelihood ratio classifiers yields excellent results with a verification rate of 94.6% and a rank-1 score of 99.0% while offering a significant advantage in processing speed.

3 3D Face Registration Method

3.1 Introduction

As explained in Sect. 2.1, our registration method does not map one point cloud on another, but transforms each point
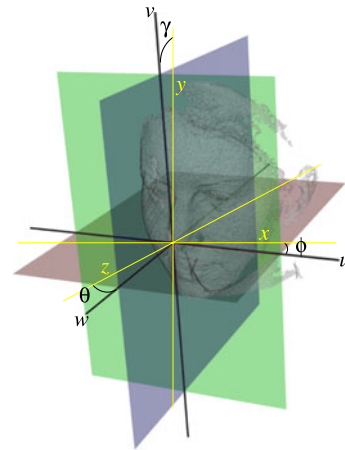

Fig. 4 The intrinsic coordinate system with u-, v- and w-axis of the 3D face is defined by its origin in the tip of the nose and 3 rotation angles: ϕ around the z-axis, θ around the y-axis and γ around the x-axis

cloud to an intrinsic coordinate system. In 2D face registration, generally landmarks, like the centres of the eyes, nose tip and mouth are used to determine a transformation to an intrinsic coordinate system. In the 3D data, often only a single stable landmark can be distinguished: the tip of the nose. At the centres of the eyes and the mouth, often there are holes in the 3D data, making accurate localisation of these landmarks very difficult. Also these landmarks may move due to facial expressions. Therefore, we used two different geometric properties of facial data: the vertical symmetry plane of the face and the slope of the bridge of the nose. Both geometrical properties are stable under variation of facial expressions (Tang et al. 2008). To define an intrinsic coordinate system, three angles and an origin must be determined. The symmetry plane defines two angles (θ, ϕ, see Figs. 4 and 8). The nose tip defines the origin and the angle of the nose bridge defines the third angle (γ, see Figs. 4, 11 and 14). The intrinsic coordinate system of a 3D face is shown in Fig. 4. The world coordinate system is spanned by the vectors x, y and z. The intrinsic coordinate system is spanned by the vectors u, v and w. The v-axis is chosen such that the angle with the nose bridge is $\frac{\pi}{6}$ rad. This will generally place faces in a frontal position.

As mentioned in Sect. 2.1, a 3D face registration method based on similar geometric properties was presented by Tang et al. (2008). However, the verification results they present on the FRGC v1 data are far inferior to the results we obtained as will be shown in the experiments in Sect. 6.

Our registration method operates on the rough 3D point cloud and consists of the following main steps:

1. Determine a region of interest containing the face
2. Determine the vertical symmetry plane of the face through the nose
3. Determine the nose tip and the slope of the nose bridge

Fig. 5 Two samples from the FRGC v2 data (04529d93.abs and 04529d101.abs) of the same subject represented as a surface

Fig. 6 Fitting a cylinder piece to two points with associated normals. *Left*: finding the axis and centroid C of the cylinder piece. *Right*: the fitted cylinder piece

4. Transform the point cloud to a coordinate system defined by the symmetry plane, nose tip and angle of the nose bridge
5. Construct a range image by projecting the point cloud to a plane perpendicular to the symmetry plane
6. Perform hole filling and spike removal

The resulting range image can be readily used for face comparison with a variety of face recognition methods. We use the likelihood ratio classifier (Bazen and Veldhuis 2004; Veldhuis et al. 2006), which is described in Sect. 4. Furthermore, we fuse the results of multiple classifiers of overlapping regions of the face. The regions and fusion is described in Sect. 5.

Because there is much variation in the 3D images due to pose, expression, facial hair etc. we designed a robust approach to the steps of the registration method. This basically means that some of the steps are performed twice: once applying a very robust approach with a large search space for the parameters, but with lower accuracy and once with a narrow search space for the parameters but aimed at high accuracy. Each step will be explained in detail below.

3.2 Region of Interest

The full 3D scans may contain more than just the face. An example from the FRGC v2 data set (Phillips et al. 2005) is shown in Fig. 5. Because other body parts may disturb the determination of the symmetry plane of the face, first a Region of Interest (ROI) around the face is determined.

The region of interest is determined by first mapping the 3D point cloud to a grid consisting of cells with size $20 \times 20 \times 20$ mm. For each cell the average 3D coordinates are determined and the surface normal is determined using eigenvector/eigenvalue analysis. Only those cells are kept with a sufficient number of 3D points and a largest eigenvalue that is clearly larger than the other two eigenvalues. The latter signals that the cell represents a reasonably flat surface with a clear normal.

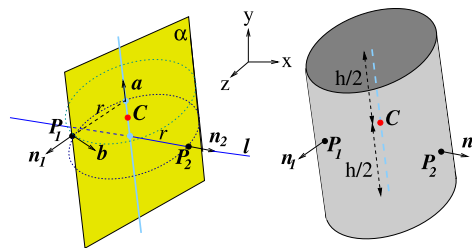Next a RANSAC (RANdom SAmple Consensus; Fischler and Bolles 1981) is used to fit a cylinder piece to the 3D facial data. RANSAC an iterative method to estimate parameters of a mathematical model from data which contains outliers. The mathematical model in our case is the cylinder piece and the outlayers are 3D points on the shoulders, torso etc. The "inliers" are the points on the face that are modelled reasonably well by the cylinder. The basic idea of RANSAC is to use a small random subset of points from the data to hypothesise the mathematical model and to calculate the consensus of the hypothesis by counting the number of points in the dataset that can be explained by the hypothesis. The process of hypothesising is repeated a number of times and the hypothesis with the maximum consensus is selected as the best fit of the mathematical model to the data. Advantages of the RANSAC approach are its robustness against outliers and its speed.

In our case, two cells can be used to define a cylinder piece using the averages of the 3D coordinates of the points in the cells and the normals. This is illustrated in Fig. 6.

The direction a of the axis of the cylinder piece is perpendicular to both normals n_1 and n_2. The intersection of a plane α through P_1 with normal $b = a \times n_1$ and the line through P_2 with direction n_2 is a point on the axis of the cylinder piece. The radius of the cylinder piece is given by the distance of P_1 and P_2 to the axis. Finally, the extent of the cylinder piece is determined by calculating the centroid C between the projections of P_1 and P_2 and cutting off the cylinder below and above half of the average face height h. The average face height was set to 200 mm.

For the RANSAC algorithm, we consider all cell pairs for fitting cylinder pieces with distance between the cells in the x-direction dc_x of $[dc_x^{\min}, dc_x^{\max}]$ and in the y-direction dc_y of less than $dc_y^{\max}$ (see Fig. 6 for definition of the axes). We chose $dc_x^{\min} = 50$ mm, $dc_x^{\max} = 100$ mm and $dc_y^{\max} = 50$ mm. The consensus C_{cyl} is calculated by counting the number of cells with distance d less than $d^{\max}$ (here: 20 mm) from the cylinder piece and normal less than $\alpha^{\max}$ (here: $\frac{\pi}{4}$ rad) deviating from the normal at the corresponding position on the cylinder.

$$C_{\text{cyl}}(i, j) = \sum_k \begin{cases} 1, & \text{if } d(k) \leq d^{\max} \wedge \alpha(k) \leq \alpha^{\max} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$
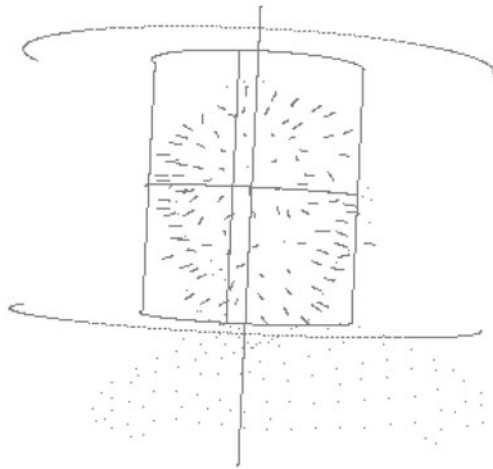
Fig. 7 ROI determined by fitting a cylinder piece to the point cloud of Fig. 5 using a RANSAC method. For points in the ROI the normals are shown as well

**Table 1** Parameter settings used in determination of the ROI

Description	Symbol	Value
Distances between pairs of points used to hypothesise cylinders	$dc_x^{\min}$	50 mm
	$dc_x^{\max}$	100 mm
	$dc_y^{\max}$	50 mm
Thresholds for contributing points to consensus	$d^{\max}$	20 mm
	$\alpha^{\max}$	$\frac{\pi}{4}$ rad
Distance to cylinder for points in reduced point set		75 mm

Where $C_{\text{cyl}}(i, j)$ is the consensus for the cylinder fit through the cells i and j, $d(k)$ is the distance of cell k to the cylinder and $\alpha(k)$ is the angle between the normals of cell k and normal on the closest point on the cylinder. The cylinder piece with the maximum consensus is chosen as the best fit. An example of a fitted cylinder piece is shown in Fig. 7. This approach of extraction of the face region appeared very reliable and did not fail a single time on a total of approximately 10 000 3D images.

All points with a distance larger than 75 mm to the cylinder piece are discarded from the point cloud. We will call the remaining point cloud the *reduced point cloud*.

Table 1 summarises the parameter settings for ROI extraction. Pairs of points used to hypothesise cylinders should more or less lie in the same horizontal plane (we are looking for cylinders with a vertical axis), hence the threshold $c_y^{\max}$. The distance between the pairs of points should not be too small or too large, because this results in inaccurate estimates of the parameters of the cylinder. The choices for the thresholds relate directly to the average size of the face and are not very critical.

Fig. 8 The symmetry plane is defined by 3 parameters: θ, ϕ and d_x
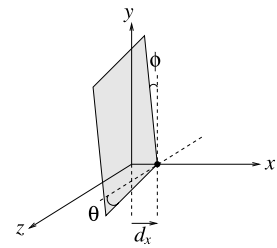
3.3 Symmetry Plane

The next step is finding the vertical symmetry plane of the face through the nose. The determination of the vertical symmetry plane takes place in two stages: first a rough estimate of the parameters of the symmetry plane and next a refinement of the parameters.

#### 3.3.1 Rough Symmetry Plane Estimation

First, a range image is created from the reduced point cloud by projecting them to the xy plane. A grid is defined on the xy plane consisting of square pixels of 5×5 mm. The projection of the centre of gravity of the reduced point cloud defines the origin of the grid. The value of a pixel is determined by calculating the average distance to the xy plane of the points that project to the pixel (i.e. the average of their z-coordinates). The result is a low resolution range image, which is shown in Fig. 10 on the left.

The symmetry plane is defined by 3 parameters as shown in Fig. 8: the rotation θ around the y-axis, the rotation ϕ around the z-axis and the x coordinate of the intersection of the symmetry plane with the x-axis: d_x. Note that the angle ϕ in both Figs. 8 and 4 refers to the rotation around the z-axis.

To find the parameters of the symmetry axis, for θ and ϕ in a range of $[-\frac{\pi}{4}, \frac{\pi}{4}]$, new range images are generated for which the projection plane is rotated such that it is perpendicular to the symmetry plane. The step sizes for θ and ϕ were set to $\frac{\pi}{40}$ rad. New range images only have to be generated from the point cloud for each value of θ. The range images for different values of ϕ for a fixed value of θ are obtained by in-plane rotation of the range image.

The new range images are mirrored in the y-axis and shifted along the original range image with distances d_x in a range of $[-\frac{3}{4}w, \frac{3}{4}w]$ with a step size of 5 mm, where w is the width of the range image. For each displacement d_x, the z-coordinates of the pixels at the same grid positions (i, j) are compared and the differences $d_z(i, j)$ for pixels that differ less than a threshold $d_z^{\min}$ are accumulated into a sum S. This sum S is a measure for the symmetry: a low S means a good match, a high S means a bad match. The threshold is used to decide if the pixels are outliers. Outlier pixels have very large differences in z-coordinates and would, therefore,

Fig. 9 Nose template used for rough nose fitting. *Darker pixels* means nearer to the observer, *brighter pixels* means further away
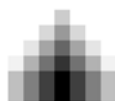
have a large impact on the sum S. This is the reason why they should not contribute to S. The symmetry measure S also depends on the number of pixels that contributed to the sum (i.e. those with $d_z < d_z^{\min}$). To make the measure independent for the number of pixels that contributed, we divide by the number of contributing pixels. Because few contributing pixels generally means a bad overlap, we punish this by dividing the sum through the number of contributing pixels once more. The resulting expression for the symmetry measure S thus becomes:

$$S(\theta, \phi, d_x) = \frac{\sum_{i,j} \begin{cases} 0, & \text{if } d_z(i,j) > d_z^{\min} \\ d_z(i,j), & \text{otherwise} \end{cases}}{\left(\sum_{i,j} \begin{cases} 0, & \text{if } d_z(i,j) > d_z^{\min} \\ 1, & \text{otherwise} \end{cases} \right)^2} \quad (2)$$

Where $d_z(i, j)$ is the absolute difference of the z-coordinates of two pixels at the same grid position (i, j) of the two range images and $d_z^{\min}$ the threshold used to decide if the pixels are outliers. In all experiments, we set $d_z^{\min} = 10$ mm. All local minima in the 3 dimensional parameter space (θ, ϕ, d_x) are recorded as potential symmetry plane candidates. The candidates for the symmetry plane are sorted in a list with increasing S.

For all candidate symmetry planes a nose model is fitted to the area around the symmetry plane on the facial surface using a simple 3D nose model as a template and Normalised Cross Correlation (NCC) as a matching criterion (see e.g. van der Heijden and Spreeuwers 2007). The nose template is shown in Fig. 9.

For each symmetry plane, the projection plane is tilted around the x-axis with an angle γ and the best position of the nose around the symmetry plane is selected. The search range in the y-direction is across the full height of the face and in the x-direction $\pm$ 15 mm from the symmetry plane. The step size in x- and y-directions is 5 mm. The range for the head tilt γ is $[-\frac{\pi}{5}, \frac{\pi}{5}]$ and the step size is $\frac{\pi}{40}$ rad.

We now select the symmetry plane with low S while at the same time a good nose fit. A good nose fit is in our case defined as a NCC of 0.6 or larger (NCC has a range of $[-1, 1]$ with 1 the best match). If there are more candidate symmetry planes with a good nose fit, the one with the best symmetry (lowest S) is selected. If there is no good nose fit, the symmetry plane candidate with the best nose fit is selected. The threshold for the NCC was found experimentally and is not very critical. The main purpose is to discard false symmetry planes, e.g. vertical planes through the eyes.

Apart from a first estimate of the symmetry plane, we now also have a first estimate of the position of the nose and
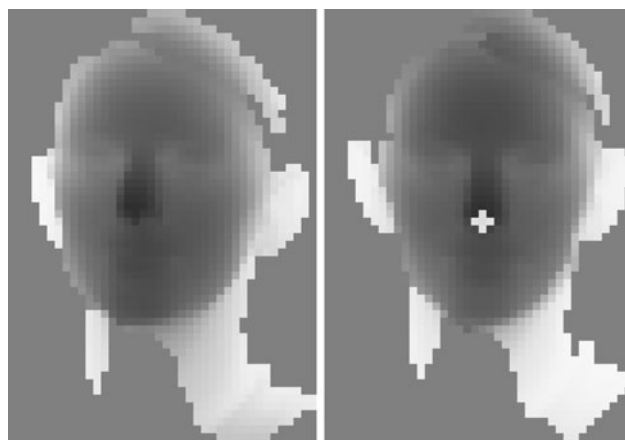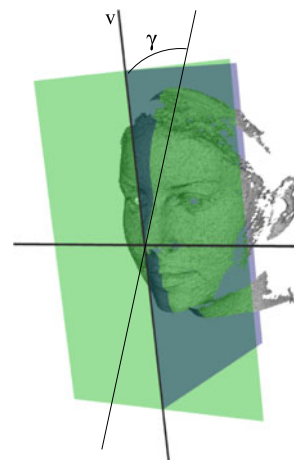

Fig. 10 Rough symmetry plane detection. *Left*: low resolution range image of original data in ROI; *Right*: rotated to frontal. The rough estimate of the nose tip is marked with a *cross*

Fig. 11 The projection plane is perpendicular to the symmetry plane, has an angle γ with the nose bridge and has its origin in the tip of the nose

the tilt of the face (γ), so basically we have a first estimate of the intrinsic coordinate system.

The parameters are used to transform the point cloud to the intrinsic coordinate system and again a low resolution range image is created as described before. The result for the image in Fig. 5 is shown in Fig. 10. Darker pixels are closer to the observer and brighter pixels are further away.

Figure 11 shows the symmetry plane and the projection plane with the origin in the tip of the nose.

This first estimate of the intrinsic coordinate system parameters appeared very reliable. The next step is a refinement of the estimation of the parameters of the symmetry plane and the nose tip and the slope of the nose bridge.

Table 2 summarises the parameter choices for the rough symmetry plane determination. The ranges for θ, ϕ and γ determine the maximum rotations the registration method can handle.

Table 2 Parameter settings for the rough symmetry plane detection process

Description	Symbol	Value
Pixel size range image		5 mm
Search range + step ϕ and θ		$[-\frac{\pi}{4}, \frac{\pi}{4}]$, $\frac{\pi}{40}$ rad
Search range + step d_x		$[-\frac{3}{4}w, \frac{3}{4}w]$, 5 mm
Threshold to exclude points for symmetry calculation	$d_z^{\min}$	10 mm
Search range + step γ		$[-\frac{\pi}{5}, \frac{\pi}{5}]$, $\frac{\pi}{40}$ rad
Search range + step nose x-pos		$[-15, 15]$, 5 mm
Threshold on NCC nose fit		0.6

3.3.2 Refinement of Symmetry Plane Estimation

For the refinement of the estimation of the symmetry plane, the point cloud is first rotated and translated to frontal view using the parameters found in the rough symmetry plane estimation, so all parameter estimation is relative to the already found rotations and translations. For the refinement of the estimation of the symmetry plane, the same symmetry measure from (2) is used. However, now a higher resolution range image with a grid size of 1 mm is used and a circular ROI with a radius of 110 mm around the tip of the nose (see Fig. 12 on the left). We used an exhaustive search strategy in the θ direction in two stages: first in a range of $[-\frac{\pi}{10}, \frac{\pi}{10}]$ with a step size of $\frac{\pi}{100}$ rad and next around the found optimum θ_1^{opt} in the range $[\theta_1^{opt} - \frac{\pi}{50}, \theta_1^{opt} + \frac{\pi}{50}]$ with a step size of $\frac{\pi}{1000}$ rad. For each value of θ, the point cloud is mirrored in the symmetry plane and projected to the projection plane perpendicular to the symmetry plane. The resulting range image is then rotated around the z-axis over an angle of ϕ and shifted in the x-direction over a distance d_x and compared to the original range image. The differences of the z-coordinates of the projected points and the pixels of the range image are again accumulated using (2). To find the optimal ϕ and d_x for each value of θ, we applied a one dimensional parabolic fit optimisation approach as described in Brent (1973), Press et al. (1988). The search ranges were $[-\frac{\pi}{10}, \frac{\pi}{10}]$ for ϕ and $[-10, 10]$ mm for d_x. The parabolic fit method iteratively fits a parabola through 3 points and substitutes the worst point by the maximum of the parabola. First the optimal value for d_x was determined for $\phi = 0$ and then this d_x value was used in the optimisation of ϕ, which in turn is then used in a second optimisation of d_x etc. and after that in a third iteration. The number of iterations for each individual parameter was set to a maximum of 10 and the optimisation was stopped if the difference of ϕ resp. d_x relative to the values in the previous iteration was less than $\frac{\pi}{1000}$ rad resp. 0.1 mm.

The circular ROI used as input to the fine symmetry plane estimation is shown on the left in Fig. 12. The result after
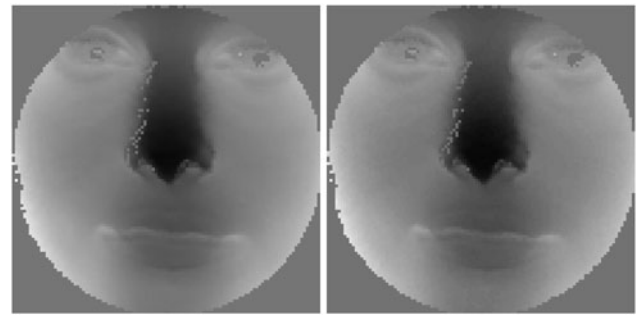

Fig. 12 Fine symmetry plane detection. *Left*: high resolution range image of circular ROI around the nose; *Right*: rotated to frontal

Table 3 Parameter settings for the refinement of the symmetry plane estimation

Description	Symbol	Value
Pixel size range image		1 mm
Radius range image		110 mm
Range and step 1st search θ		$[-\frac{\pi}{10}, \frac{\pi}{10}]$, $\frac{\pi}{100}$ rad
Range and step 2nd search θ		$[-\frac{\pi}{50}, \frac{\pi}{50}]$, $\frac{\pi}{1000}$ rad
Range and resolution search ϕ		$[-\frac{\pi}{10}, \frac{\pi}{10}]$, $\frac{\pi}{1000}$ rad
Range and resolution search d_x		$[-10, 10]$, 0.1 mm
Max # iterations 1D search		10
Max # iterations 2D search		3

adjustment using fine symmetry plane estimation is shown on the right in Fig. 12. Note there is only a minor adjustment to the rough symmetry estimation. The holes on the right side of the nose (left in the images) occur because these parts are invisible in the original 3D recording of Fig. 5.

The next step in the registration procedure is accurate estimation of the tip of the nose and the slope of the nose bridge. This will be detailed in the next section.

Table 3 shows the parameter settings for the refinement of the symmetry plane estimation.

3.4 Nose Tip and Slope of Nose Bridge

In order to locate the nose tip and determine the slope of the nose bridge, a rough estimate of the tilt angle γ of the face is required. A first estimate of the tilt angle was already obtained in the rough nose detection process in the symmetry plane estimation. However, it turned out that sometimes this estimate was insufficiently reliable, because it relies on fitting a crude local nose model to a very low resolution (5 mm) range image.

Therefore, a second more accurate and reliable estimate of γ is determined by fitting a cylinder to the circular ROI of the face, thus using higher resolution (1 mm) and more global data. Basically this means finding the 'up' axis of the face. The cylinder in this case has a fixed radius $r =$
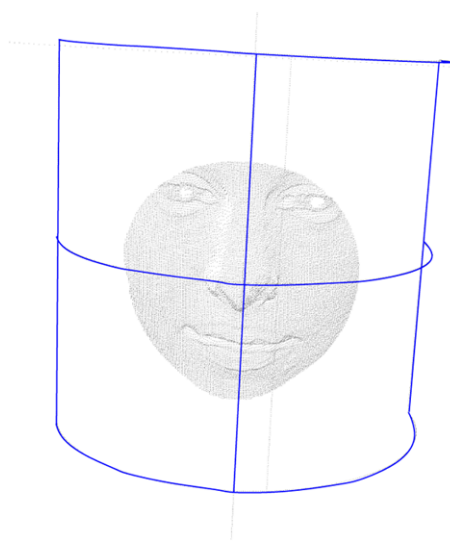
Fig. 13 A cylinder fitted to the circular ROI surface provides a first estimate of the "up"-axis
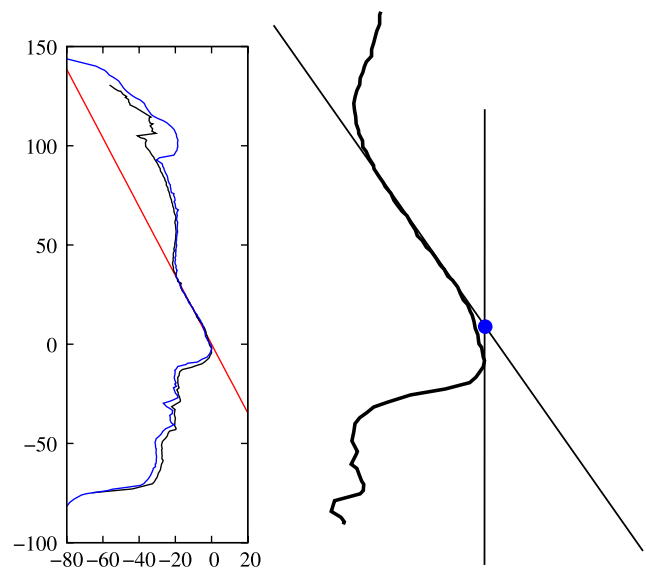

Fig. 14 *Left*: line fitted to nose bridge of the profile of the face. The tip of the nose is at coordinates $(0, 0)$. Two profiles of the same person are shown on top of each other. *Right*: definition of the tip of the nose as *intersection of two lines*

100 mm and the axis of the cylinder lies in the symmetry plane with an angle γ to the vertical y-axis. The angle γ of the axis is varied between $-\frac{\pi}{2}$ rad and $\frac{\pi}{2}$ rad with a step of $\frac{\pi}{100}$ rad. The γ that gives the cylinder with the highest consensus according to (1) is selected as an initial estimate for γ. The result of fitting a cylinder to the circular ROI is shown in Fig. 13.

Next the projection plane for the range images is adjusted for the new γ and a profile of the face is extracted by projecting all points of the point cloud with a distance less than 5 mm from the symmetry plane on the symmetry plane and record their v and w coordinates (see coordinate system in Fig. 4).

First outliers are removed from the profile. The profile is resampled with a point distance of 1 mm in the v direction, recording both the maximum as well as the average in the w-direction for each position. Outliers are defined as points with a w-coordinate deviating more than 5 mm from the average. Next the first estimate of the tip of the nose is found by detecting the point with the maximum w-coordinate. Around the tip of the nose, now straight lines are fitted to the profile. The line fitting is done again using the RANSAC (Fischler and Bolles 1981) approach. All combinations of two points around the tip of the nose that have a distance to each other of at least 10 mm are used to construct straight lines. The consensus of a line with the profile is calculated by counting the points above the tip of the nose that have a distance in the w-direction of less or equal to $d_l^{\max}$ to the line:

$$C_1(i, j) = \sum_k \begin{cases} 1, & \text{if } d_1(k) \leq d_l^{\max} \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

Where $C_1(i, j)$ is the consensus of the line through points i and j on the profile and $d(k)$ is the distance in the w-direction of point k to the line. The line with the highest consensus is selected as the best fit. Because the nose bridge is the longest more or less straight line piece around the tip of the nose, the found line lies on the nose bridge. As mentioned before, the RANSAC approach is very robust against outliers and generally results in an accurate estimate of the best fitting line. The angle γ defining the tilt of the head is now defined as the angle of the found line on the nose bridge. The profile is rotated such that $\gamma = \frac{\pi}{6}$ rad. This places the face in an upright position, resulting in a frontal view. Finally the tip of the nose is found as the intersection of a line parallel to the v-axis through the point on the profile with the maximum w-coordinate and the line on the nose bridge. It turned out that choosing this point as the tip of the nose is slightly more stable than the point with the highest w-coordinate or the point with the highest curvature. The result of the line fitting to the nose bridge is shown in Fig. 14. To illustrate how well the alignment of two faces works, two profiles of different 3D images of the same person are shown.

At this point, all parameters needed for registration of a facial point cloud to the intrinsic coordinate system defined in Fig. 4, have been determined. For further processing using face classifiers, some post processing steps are required, which are described in the subsequent section.

Table 4 summarises the parameter settings for the estimation of the position of the nose tip and slope of the nose bridge. The radius of the cylinder is derived from the average size of the head. We observed that the registration accuracy is not very sensitive to these parameters, which is supported

Table 4 Parameter settings for the nose tip and slope estimation

Description	Symbol	Value
Radius cylinder	r	100 mm
Search range + step γ		$[-\frac{\pi}{2}, \frac{\pi}{2}]$, $\frac{\pi}{100}$ rad
Max distance to cylinder	$d^{\max}$	20 mm
Max deviation from normal	$\alpha^{\max}$	$\frac{\pi}{4}$ rad
Max distance to symmetry plane		5 mm
Resample density profile		1 mm
Outlayer threshold		5 mm
Min dist points for line hypothesis		10 mm
Max dist point to line for consensus	$d_l^{\max}$	5 mm

Fig. 15 High resolution range image. For the *black areas*, no depth information is available in the original point cloud

by the fact that the same parameter values resulted in correct registrations for about 10 000 3D faces.

3.5 Range Image

The first of the post processing steps is the generation of a high resolution range image. This may not be necessary for all types of 3D face recognition methods, but the PCA-LDA-likelihood ratio approach that we chose requires an input vector of fixed length. Therefore, a high resolution range image is constructed by projection of the original point cloud to the projection plane defined by the found parameters. In principle a higher resolution may give better recognition, because details are better represented. For each pixel of the grid of the range image, the average of the w-coordinates of the points projected on the pixel is determined. The number of contributing points is stored in a counter flag f for each pixel as well. A simple filter for removing occluded points from the point cloud is also applied. These are points that lie more than several millimetres behind other points that project on the same pixel in the grid of the range image. The resulting range image is shown in Fig. 15.

Due to resampling, imperfect scanning and the fact that some areas in the face may not have been visible during scanning, holes occur in the range image. Furthermore, errors in the scanning process may produce spikes. In order to further process the range images, the holes must be filled and the spikes must be removed.

Fig. 16 Spikes near the eyes and on the forehead in a 3D face surface
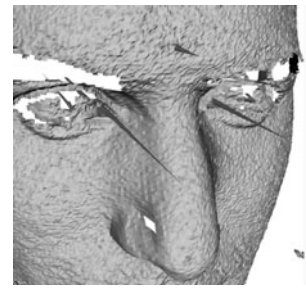
3.6 Spike Removal, Hole Filling and ROI

Spikes occur in the data due to scanning errors. These errors may be caused by specular reflections in e.g. the pupils of the eyes. Smaller spikes can occur anywhere in the data. Figure 16 shows an example of spikes in the eye.

Spike removal is performed by low-pass filtering the range image and discarding all points from the point cloud with a w-coordinate that deviates more than $d_{\mathrm{sr}}^{\max}$ from the average w-value of the corresponding pixel of the grid of the range image. We chose $d_{\mathrm{sr}}^{\max} = 5$ mm. The low pass filtering step is a special kind of filtering, because to some pixel of the grid, no points are projected, while to other pixels more or fewer points of the point cloud are projected. The low pass filtering takes the number of points that project on a pixel into account. Cells with a larger count are considered more reliable and given a higher weight in the averaging process. The low pass filtering now proceeds as follows: First, the average contributing point count $\overline{f}$ per pixel is determined for the range image. Next, for each pixel i, a new w-value $w_{\mathrm{sr}}(i)$ and count $f_{\mathrm{sr}}(i)$ are determined by adding together the average w-value of the pixel and the pixels in a square neighbourhood $N(i)$, weighted with their respective counts $f(j)$ dividing by the total count of the neighbourhood:

$$f_{\mathrm{sr}}(i) = \sum_{j \in N(i)} f(j) \tag{4}$$

$$w_{\mathrm{sr}}(i) = \frac{1}{f_{\mathrm{sr}}(i)} \sum_{j \in N(i)} w(j) f(j) \tag{5}$$

The size of the neighbourhood of each pixel is chosen such that the new count $f_{\mathrm{sr}}(i)$ is larger or equal to a fixed multiplier M_{sr} times the average count $\overline{f}$. In the left image of Fig. 17 the spikes of Fig. 16 are visible as dark spots (closer to the observer). The resulting filtered range image for a multiplier $M_{\mathrm{sr}} = 25$ is shown in the middle in Fig. 17 and on the right the result of the spike removal is shown.

Holes can be distinguished in small holes, large holes and missing face parts. Small holes are caused by minor scanning failures, the high resolution resampling process or the spike removal process. Large holes are caused by scanner failures and occlusion. A typical example of large holes caused by scanner failure are the pupils of the eyes (see
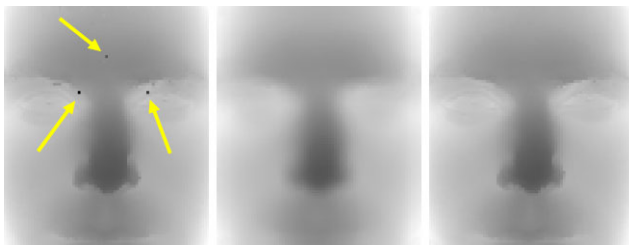
Fig. 17 *Left*: spikes in the range image; *Middle*: filtered range image for spike removal; *Right*: after spike removal

Fig. 18 Result after hole filling (*left*) and after selection of an elliptical ROI (*right*). The latter is the final result of the post processing

Table 5 Parameter settings for spike removal and hole filling

Description	Symbol	Value
Max deviation to low pass filtered surface for spike removal	d_{sr}^{max}	5 mm
Multiplier defining neighbourhood for spike removal	M_{sr}	25
Multiplier defining neighbourhood for hole filling	M_{hf}	0.25

Fig. 5). An example of large holes caused by occlusion are the sides of the nose, which can be occluded if the face is rotated around the y-axis. Missing facial parts can be caused by scanner failure and by large rotations of the face around the x-axis and/or the y-axis. If e.g. a person is looking down, the part between the upper lip and the nose may not be visible in the 3D scan. After rotation to frontal pose, this causes a hole.

Small holes are filled using interpolation. This interpolation works similar to the low pass filtering used for the spike removal with the exception that the w-values and the counts of the neighbourhood are weighted with the reciprocal of the distance to the centre of the neighbourhood:

$$f_{hf}(i) = \sum_{j \in N(i)} \frac{f(j)}{r(i,j)} \tag{6}$$

$$w_{hf}(i) = \frac{1}{f_{hf}(i)} \sum_{j \in N(i)} \frac{w(j)f(j)}{r(i,j)} \tag{7}$$

Where $r(i,j)$ is the distance between pixels i and j in the grid of the range image. In this case the multiplier $M_{hf} = 0.25$, or if $M_{hf}\overline{f} < 1$, M_{hf} is chosen such that $M_{hf}\overline{f} = 1$.

Large holes and missing parts are filled using the symmetry of the face. Large holes are detected by testing if a pixel i and all its immediate neighbours j have counts $f(i)$ and $f(j)$ that are less than $M_{hf}\overline{f}$. If for a pixel i in a big hole the pixel i_m on the position mirrored in the symmetry plane has a count larger than $M_{hf}\overline{f}$, then $w(i_m)$ and $f(i_m)$ are copied to pixel i.

The order of processing holes is that first the big holes are filled and then the remaining small holes. If the big holes cannot be filled using symmetry, because holes occur on both sides of the symmetry axis, the big holes will still be filled using the approach for small holes.

An example of the result of the hole filling is shown in Fig. 18 on the left.

The final step of the post processing is cutting out an elliptical region of interest (ROI), keeping only parts of the face that are visible in all images. Choosing a larger ROI may result in including parts of the background for smaller faces. The final range image is shown in Fig. 18 on the right.
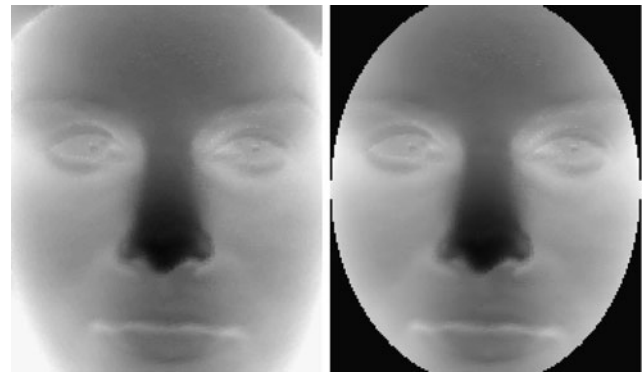
Although the simple approaches to spike removal and hole filling perform well in most of the cases (evaluated using visual inspection and supported by the excellent 3D face recognition results reported in Sect. 6), more advanced approaches, like hole filling using a PCA model (see Colombo et al. 2006) may yield even better results.

Table 5 shows the parameter settings used in the spike removal en hole filling process. Spikes deviating more than 5 mm from the average surface are removed. The values of the multipliers result for spikes consisting of a single pixel in a neighbourhood of 5×5 and for holes of a single pixel in a neighbourhood of 3×3. For larger holes/spikes, the neighbourhoods are extended to include a sufficient number of 3D points to make a reasonable prediction of the local depth value.

3.7 Alternative Registration Approach

A disadvantage of determining the tilt of the head by the slope of the nose bridge is that for some people if they show severe facial expressions the tip of the nose is pushed upwards. In addition, in some scans the tip of the nose cannot be determined accurately either because it is missing (see Fig. 25) or because inaccuracies occur due to e.g. specular reflections. We, therefore, investigated a second approach to determining the tilt of the face (the angle γ) and the origin of the intrinsic coordinate system. In this case, the origin is defined not at the tip of the nose, but at the point just below
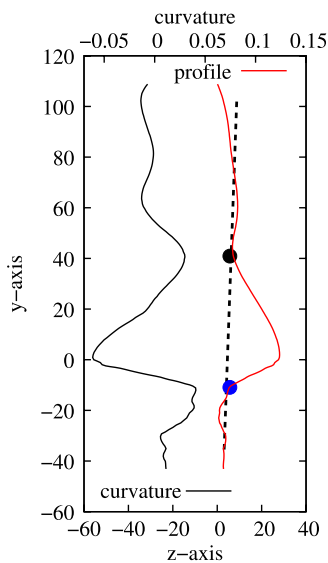
Fig. 19 Determining the dent in the nose bridge (*upper dot*) and the *point below* the tip of the nose (*lower dot*) from curvature. Both points are a local maximum of the curvature. The tilt of the face is determined by *the line through the two points*

Fig. 20 Profiles of two images of the same subject registered by the normal registration (*left*) and the alternative registration (*right*). The normal registration incorrectly localises the tip of the nose, because the shape of the nose is different. The alternative registration is not dependent on the shape of the tip

Fig. 21 Range images resulting from the normal registration. The incorrect localisation of the tip of the nose causes a vertical shift in one of the images

the nose. Both of these points can easily be determined using the curvature of the (smoothed) profile. This is illustrated in Fig. 19 which shows the smoothed profile of a face and the corresponding curvature. The tip of the nose has a large negative curvature. The dent above the nose bridge (black dot) and the point just below the tip of the nose (blue dot) both have large positive curvatures and are the first strong local maxima near the tip of the nose. The tilt of the face is determined by first locating the dent at the top of the nose and fitting a line through these two points (dashed line).

In Fig. 20 a case is illustrated for which the alternative registration approach works better than the original. The figure shows the profiles of two images of the same subject for the registration based on the slope of the nose bridge and nose tip ("normal registration") on the left and on the dent above the nose tip and the point just below the nose tip ("alternative registration") on the right. Note that for the normal registration the nose tip is at the origin $(0, 0)$, while for the alternative registration the point just below the nose is at the origin.

Because of small motion artifacts at the tip of the nose, the shape of the tip of the nose is deformed leading to incorrect localisation of the tip of the nose in one of the images which causes a vertical shift in the range image, see Fig. 21. The incorrect localisation of the tip of the nose is because it is defined as the intersection of two lines (see Fig. 14). Furthermore, there is compression of the nose area caused by the facial expression in the face on the right, making the slope of the nose bridge a less accurate measure of the tilt of the head. Note that the eyes and nose tip are not at the same vertical position in Fig. 21. The alternative registration does
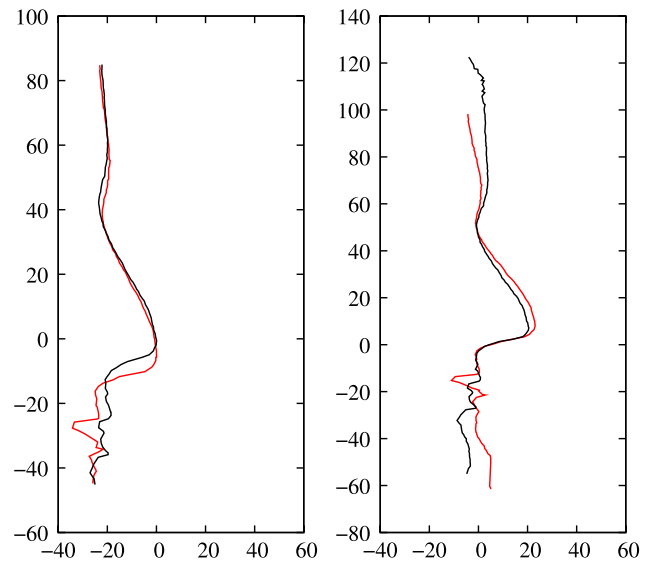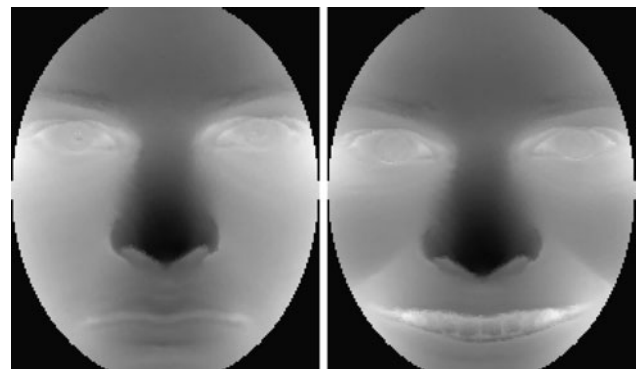
find the correct points in both images and correctly registers both images. The range images of the alternative registration are shown in Fig. 22. Note that the nose tip is better aligned now, but because of the compression of the nose area caused by the facial expression, the eyes are still not aligned.

Because only a small part of the complete registration changes for this alternative approach, a range image with the alternative registration can easily be generated in addition to the original range image at very little cost, i.e. it takes hardly more time to generate two range images instead of a single range image. The alternative registration appeared slightly less robust than the "normal" registration. However, because it makes different mistakes, it makes sense to fuse
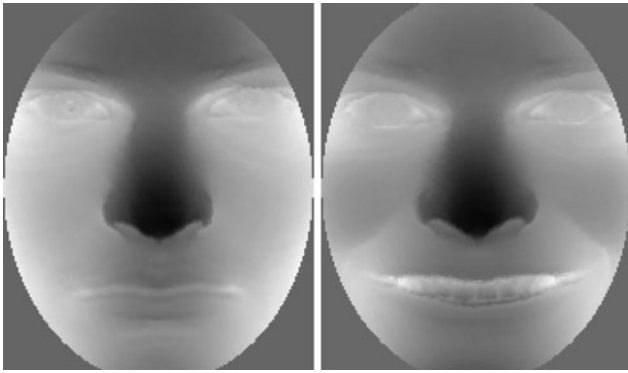
Fig. 22 Range images resulting from the alternative registration. The faces are better aligned this time

classifiers trained on images registered with the two different registration approaches.

4 PCA-LDA-Likelihood Ratio Classifier

For comparison of the 3D range images, we use a classifier based on the likelihood ratio as described in Bazen and Veldhuis (2004), Veldhuis et al. (2006), Beumer et al. (2006). The likelihood ratio is defined as:

$$L(\mathbf{x}) = \frac{p(\mathbf{x}|c)}{p(\mathbf{x})} \tag{8}$$

Where $p(\mathbf{x}|c)$ is the conditional probability on a feature vector $\mathbf{x}$ for class c and $p(\mathbf{x})$ is the unconditional probability on feature vector $\mathbf{x}$ for any class. The classes here refer to the identities of the subjects. If we assume that $p(\mathbf{x}|c)$ and $p(\mathbf{x})$ are normally distributed, then:

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{m}{2}} |\Sigma_T|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\mu_T)^T \Sigma_T^{-1} (\mathbf{x}-\mu_T)} \tag{9}$$

and:

$$p(\mathbf{x}|c) = \frac{1}{(2\pi)^{\frac{m}{2}} |\Sigma_W|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\mu_c)^T \Sigma_W^{-1} (\mathbf{x}-\mu_c)} \tag{10}$$

Where m is the dimension of the feature vector. μ_T and μ_c are the mean feature vectors of the total distribution (of all classes) and the within class distribution (for a single class). Σ_T and Σ_W are the covariance matrices of the total distribution resp. the within class distribution. μ_T, Σ_T and Σ_W are estimated from training data. Because generally only few samples are available per class, we assume that the within class variation Σ_W is the same for all classes. In this way the data of all classes can be used to estimate Σ_W by subtracting the class mean.

To compare two 3D range images, we first vectorise the images. Next, one of the images is selected as probe $\mathbf{x}_p$ and

the other as reference $\mathbf{x}_r$. We want to find out if the probe and the reference are of the same class, i.e. are recordings of the same subject. Since we have only one reference vector available, the best estimate of the class mean is the reference vector itself, so we set $\mu_c = \mathbf{x}_r$. The likelihood can then be calculated using (8). If the likelihood is above a certain threshold, the probe is accepted as a recording of the same subject as the reference, otherwise it is rejected. Prior to classification, the feature vectors are transformed to a lower dimensional subspace by a $d \times m$ transformation matrix $\mathbf{M}$ that simultaneously diagonalises the within class and the total covariance matrices, such that the latter becomes the identity matrix.

The transformation matrix $\mathbf{M}$ is found by PCA followed by LDA (Veldhuis et al. 2006). The expression for the likelihood ratio can now be simplified by applying the transformation $\mathbf{M}$ and taking the natural logarithm:

$$l(\mathbf{y}) = \log \frac{p(\mathbf{y}|c)}{p(\mathbf{y})} = -\frac{1}{2}(\mathbf{y}-\nu_c)^T \Lambda^{-1} (\mathbf{y}-\nu_c)$$
$$+ \frac{1}{2}(\mathbf{y}-\nu_T)^T (\mathbf{y}-\nu_T) - \frac{1}{2}\log|\Lambda| \tag{11}$$

Where $\mathbf{y} = \mathbf{M}\mathbf{x}$, $\nu_c = \mathbf{M}\mu_c$, $\nu_T = \mathbf{M}\mu_T$ and $\Lambda = \mathbf{M}^T \Sigma_W \mathbf{M}$ a diagonal matrix. The transformation matrix $\mathbf{M}$ depends on the number of retained PCA components p and the number of retained LDA components d. The dimensionality of the transformed feature vectors $\mathbf{y}$ is d. One of the interesting results of this research is that only very few components are needed for a good classification. As is shown in Sect. 6.3, as few as 12 numbers suffice ($d = 12$) to obtain a recognition rate of around 80% for a FAR of 0.1%. This means that discriminating 3D range maps of faces requires very little information and very compact feature vectors can be used as templates.

Because the estimate of the class mean vector ν_c is based on a single reference vector, the estimate is not very accurate. Bazen and Veldhuis (2004) argue that in this case all elements of the within class covariance matrix are twice as large as for the case with known class mean vectors. We use the proposed correction to the within class covariance matrix, resulting in an acceptance region $2^{d/2}$ times as large.

5 Fusing Multiple Regions

5.1 Region Classifiers

One of the main deficiencies of the PCA-LDA-based classifier described in the previous section is its limited capability to handle local variations in the faces, caused by e.g. expressions or acquisition errors like missing data, motion

deformation etc. In principle these can be learnt from example data, however, only if sufficient examples of each type of variation are available. Normally, this is not the case. One way to handle local variations is to divide the face into a number of regions, perform recognition on the separate regions and fuse the results. This approach was used in several recent publications, including ICP based approaches using local ICP, see e.g. Faltemier et al. (2008a), Queirolo et al. (2010), Boehnen et al. (2009), Alyüz et al. (2009). Generally, the regions are chosen disjunct in order to obtain independent recognition results. A problem with smaller regions is, however, that the recognition rates are very low. Therefore, we investigated the fusion of many relatively large overlapping regions. We defined a set of 30 overlapping regions which are shown in Fig. 23 where the white area is included and the black area is excluded. The regions were chosen in such a way that for different types of local variation they would allow stable features for comparison. Examples of such regions are those that leave out the upper or the lower part of the face because of variation in hair, caps etc. or variation in expression of the mouth. Other examples are leaving out areas covered by glasses and the left or right side of the face, which are less visible for large rotations around the vertical y-axis.

We started by combining a few overlapping regions, but as it became clear that adding more still improved recognition results, we added more regions until the 30 regions shown in Fig. 23 resulted. After this point adding more regions did not seem to result in significant improvements anymore, as can be observed in the experiments presented in Sect. 6.4. However, more careful research into the definition of the regions and the combination of the right regions may still give some performance improvement.

From now on we will call the classifiers for a certain region *region classifiers*. The next step is the fusion of the results of the region classifiers into a single score or decision. Of course the region classifiers for the smaller regions will perform worse than those of the larger regions, but they may still contribute to the fused score if the small region happens to be one of the few stable regions in the image (i.e. sometimes, due to acquisition errors, large occlusions by hats or hair or extreme expressions, only a small part of the face, e.g. the nose is still unchanged relative to the neutral face). In the subsequent sections, the fusion methods used for the verification and identification scenarios are explained.

5.2 Fusion

There are many ways to fuse the results of a pool of classifiers. In Ross et al. (2006a, 2006b), 5 levels of fusion are distinguished:

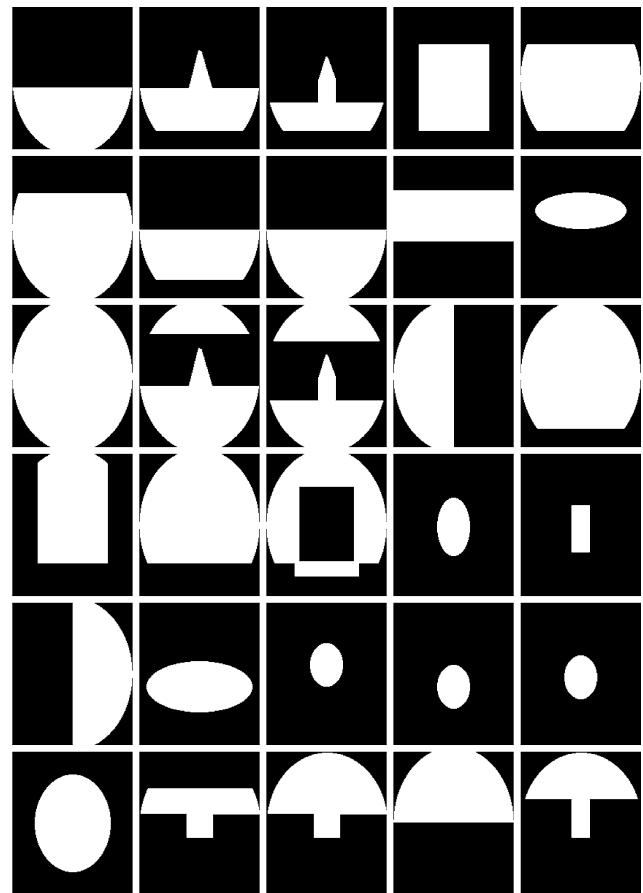1. Sensor level fusion—fusion of raw data from different sensors before feature extraction

Fig. 23 Regions used for different classifiers. Parts excluded by regions include upper, lower parts, mouth region, hair region, glasses etc. Some regions only use a small area around the nose. Note that most regions overlap and the corresponding classifiers are, therefore, not independent

2. Feature level fusion—fusion of extracted features extracted from different sensors, feature extraction methods or different recordings of the same subject
3. Rank level fusion—combination of sorted lists of identities in decreasing order (only for identification)
4. Decision level fusion—combination of decisions of the different classifiers, e.g. AND and OR rules and majority vote
5. Score level fusion—combination of the scores of the different classifiers e.g. the (weighted) sum and product of likelihoods

Since we only use a single 3D sensor, sensor level fusion is not applicable in our case. Feature level fusion can in principle be applied in our case, but because all features of all region classifiers are extracted from the same image using the same feature extraction technique (PCA-LDA), it is questionable if this will result in any performance improvement. The other 3 fusion approaches are all applicable to our approach and indeed we performed a number of experiments with different fusion techniques like the optimal OR

decision fusion (Tao and Veldhuis 2007, 2009; Tao et al. 2007). In the end we opted for one of the most common approaches to fusion: majority voting. Majority voting is a form of decision level fusion, where the identity is assigned on which the majority of the classifiers agree. Majority voting very well fits the idea of using multiple region classifiers that each represent more or less stable regions for different expressions or facial occlusions. For neutral faces, generally all region classifiers will present the correct decision. For faces with expressions, some of the region classifiers (e.g. the full face region classifier) may present the wrong decision, but still many others will present the correct decision. A further characteristic of the region classifiers we use is that they are dependent, because the regions used for feature extraction overlap. Support for using majority voting for the fusion of many dependent classifiers is provided in Kuncheva et al. (2003). Applying simple majority voting fusion to the region classifiers already gave extremely good results as is presented in Sect. 6. In this paper we, therefore, did not explore the benefits of the different fusion approaches in depth. However, in Ross et al. (2006a) and Kuncheva et al. (2003) several approaches are described (weighted majority voting, Dempster-Shafer Theory of Evidence, selection of the best combination of classifiers etc.) that will likely further improve the results. Another promising fusion approach combining optimal decision OR fusion and the sum rule score level fusion was presented in Tao and Veldhuis (2008). In future research, we will investigate other fusion strategies in more depth.

5.2.1 Identification—Closed Set

Application of majority voting fusion is straightforward for the closed set identification scenario. In this case, it is guaranteed that identity of the probe image matches one of the identities of the gallery images. Each region classifier compares the probe image to all images in the gallery and selects the one with the highest score. This results in one vote for the identity corresponding to the selected gallery image. The identity of the subject in the gallery that gathers most votes is the winner and presented as the output of the fusion.

5.2.2 Identification—Open Set

In case of an open set identification scenario, it is not guaranteed that the identity of the probe image is represented in the gallery. In this case we need a threshold on the minimum number of votes. If the number of votes is below this threshold, the probe image is not recognised and rejected. An example of this scenario is access control for e.g. buildings where entrance must be denied to all people not present in the gallery.
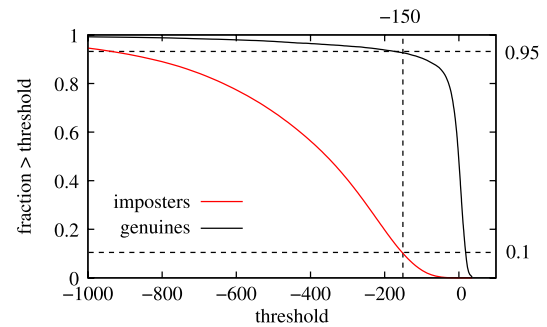

Fig. 24 The tippet plot of a classifier shows the fraction of imposter scores and genuine scores that are larger than the threshold as a function of the threshold. The *dashed lines* show that at a FAR of 10% the threshold is −150 and the VR is 95%

5.2.3 Verification

In the verification scenario, the identity of a subject must be verified against a claimed identity. In face recognition, this normally means that a facial recording must be compared to an image on some kind of identification document. In principle, this scenario corresponds to an open set scenario with a gallery consisting of a single image. A typical example is border control using the photograph on a passport. In this case a decision is made by comparing the score of the classifier to a threshold. This threshold is chosen to match the requirements of the application. Requirements can be formulated in terms of e.g. maximum verification rate (VR) at a predefined false acceptance rate (FAR). A requirement often used in verification experiments is maximum VR at FAR = 0.1%, i.e. if 1 out of 1000 imposter claims is accepted as a genuine claim.

We implemented the majority voting fusion for the verification scenario, by first determining the decision thresholds for all region classifiers using a separate dataset for a fixed FAR. For each pair of images in the dataset, the matching score is determined. For an imposter claim this results in an imposter score and for a genuine claim into a genuine score. If we plot the fraction of imposter scores larger than the threshold (that is the FAR) as function of the threshold, we can determine the required threshold for a certain required FAR. By also plotting the fraction of genuine scores larger than the threshold, we also obtain the VR. This plot is sometimes referred to as the Tippet plot, see e.g. Gonzalez-Rodriguez et al. (2002). An example is shown in Fig. 24, where the VR at FAR = 10% is equal to 95% at a threshold of −150.

To determine the fused decision for the comparison of a probe to a reference image, the scores S_i for each region classifier i are compared to the threshold T_i of the region classifier and the decisions are accumulated:

$$V = \sum_i^{\text{all regions}} \begin{cases} 1, & \text{if } S_i > T_i \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

The number of votes is then regarded as a new score and is compared to a another threshold T_v to reach a decision:

$$D = \begin{cases} 1, & \text{if } V_i > T_v \\ 0, & \text{otherwise} \end{cases} \qquad (13)$$

The threshold T_v must be determined using a second dataset and again is tuned for a specific FAR or FRR which is not necessarily the same as the one used in obtaining the thresholds T_i of the individual region classifiers. We call the FAR that is used to obtain the first set of thresholds T_i the *projected FAR* FAR_p. The optimal setting for FAR_p can be different from the FAR required for the fused classifier.

6 Experiments and Results

In this section a number of experiments is presented. The first set of experiments (Sect. 6.2) reports results of the registration method. The second set of experiments (Sect. 6.3) concern the selection of the parameters of the classifier. The third set of experiments (Sect. 6.4) shows results of region classifiers and fusion. The fourth experiment (Sect. 6.5) shows a comparison to Tang's work. The last set of experiments concerns a comparison with the state of the art of 3D face classification methods both in processing speed (Sect. 6.6) as well as on performance (Sect. 6.7). For all comparisons the FRGC v2 3D data set was used (Phillips et al. 2005). For training, other data sets were used: the Bosphorus (Savran et al. 2008) data and a part of the 3Dface data (3DFace 2009). The different data sets are detailed in Sect. 6.1.

6.1 3D Face Databases

In our experiments we used 3 databases with range images:

1. FRGC database
2. Bosphorus database
3. 3DFace database

The FRGC database was used for evaluation, while the other two databases were used for training of the region classifiers. In the subsequent sections follows a brief description of the three databases.

6.1.1 FRGC Data

In our experiments we use the FRGC v2 database (Phillips et al. 2005) for evaluation. This database was released in 2004 and consists of 4007 images of 466 different subjects. The number of images per subject varies from 1 to 22. The 3D scans were acquired using a Minolta Vivid 910 laser scanner and the 3D data are represented as a grid of 480×640 3D points. The effective resolution on the face surface is around

04746d44   04814d22   04349d412

Fig. 25 Examples of images in the FRGC data base with motion artifacts, a missing nose and puffed cheeks

0.6 mm between points in x- and y-directions. All recorded faces are close to the frontal pose. Facial expressions vary from neutral through mild to extreme expressions with puffy cheeks. Some of the images have artifacts due to e.g. motion during acquisition and improper distance to the scanner, see Fig. 25. One of the subjects has a wrong identification number as reported in Queirolo et al. (2010) (subject 04643 is the same as subject 04783). For fair comparison, we did not make any changes to the data or signature sets.

In addition to the FRGC v2 database, we used the FRGC v1 database consisting of 943 images of 275 of the same subjects as the FRGC v2 data for adjusting thresholds for fusion for the verification experiment.

6.1.2 Bosphorus Database

The Bosphorus database (Savran et al. 2008) became available in 2008 and consists of 3396 recordings of 81 subjects with 31–53 samples per subject. The 3D images were recorded using an Inspeck Mega Capturor II 3D, which is a commercial structured-light based 3D digitiser device. The resolution of the images is generally somewhat lower than those of the FRGC v2 database at around 0.8 mm between points on the face surface, because the images are subsampled. There are, however, fewer acquisition artifacts like motion or spikes. It contains images under several different poses with up to 90 degree side views, different expressions and partly occluded faces by hands, glasses etc. We used the Bosphorus data only for training and selected all frontal images without occlusions but with expressions. This resulted in a set of 2733 images of 81 subjects.

6.1.3 3DFace Database

The 3DFace dataset was acquired for an EU FP6 research project: 3DFace (3DFace 2009). This dataset is not public. It was acquired using the viSense scanner, based on structured light and developed within the 3DFace project. The images are available as a grid of 480×640 3D points. The resolution of these faces is around 0.5 mm between points at the face surface. It consists of images with different poses, with and without glasses, occlusion (caps) and various expressions.
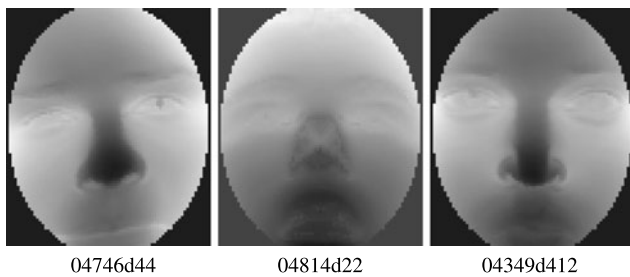
04746d44   04814d22   04349d412

Fig. 26 Registration results of images in the FRGC data base with motion artifacts, a missing nose and puffed cheeks

We used the 3DFace dataset only for training purposes and selected images without glasses, resulting in a set of 673 images of 64 subjects.

6.2 Performance of Registration

For this research a total of 8356 images were registered using the new registration method. Registration failed for very few of the images, most notably those with missing noses (two images of the FRGC dataset) and extreme expressions. But even images with severe expressions, occlusions by glasses, hats, caps etc. are handled quite well. Some expressions, liked puffed cheeks etc. deform the face in such a way that although the faces are registered correctly to the intrinsic coordinate system, still comparison with other images of the same subject may fail. Results for the images with artifacts of Fig. 25 are shown in Fig. 26. The left image shows the effect of motion artifacts: the nose bridge is still vertical but the rest of the face is distorted. The image in the middle shows the effect of a missing nose: here the tilt of the face cannot be estimated accurately. The image on the right shows the result on a face with puffed cheeks.

The registration method was originally developed for the 3DFace data, which means high resolution and high quality data from a 3D scanner based on structured light. The 3DFace dataset also contains images with glasses and with large variation in pose and, hence also large variation in pose is handled well too (up to ±45 deg in all directions). An example is shown in Fig. 27.

The range image in Fig. 27 also shows that the hole filling using interpolation and mirroring is not always working perfectly and needs some improvements. This, together with extensive experiments on 3D face recognition for large variations in pose, expressions and occlusions are subjects of our future research.

The registration was implemented in C++ on a standard Linux PC as a single threaded program. Registration takes on average 2.5 seconds for images from the FRGC v2 database and 3DFace database. Both contain images with 50 000 to 120 000 points. Registration of images from the Bosphorus database is faster with around 1.2 seconds, because they contain fewer points (around 40 000).
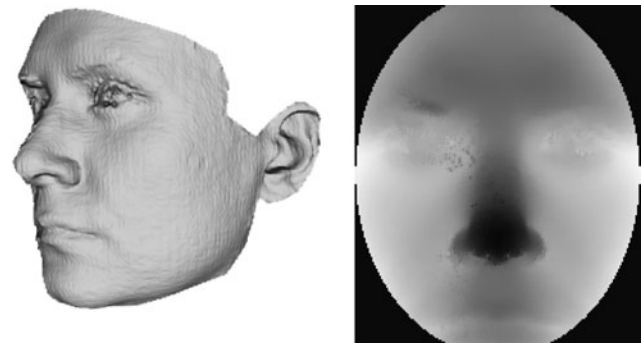

Fig. 27 Sample bs000_YR_R45_0_3D from the Bosphorus data with a rotation of 45° and the resulting range image

It hardly makes sense to present 8000 examples of correctly registered faces. A better means of evaluation of the quality of the registration is by applying it for face recognition as will be done in the subsequent sections.

6.3 Selection of Parameters

The PCA LDA likelihood ratio based classifier has two important parameters that must be chosen: the number of principle components for the PCA step and the number of components in the LDA step. From our experience in 2D face recognition, we expected optimal number of around 200 PCA components and 100 LDA components. However, for 3D face recognition, the optimal number of required LDA components turned out to be much lower. We performed two experiments to determine optimal values for the number of PCA and LDA components. We selected a single classifier region (top right in Fig. 23) and trained the classifier using the Bosphorus frontal data. In the first experiment the number of LDA components was varied from 3 to 90 while the number of PCA components was fixed to 100. In the second experiment the number of PCA components was varied from 25 to 500 while the number of LDA components was fixed to 25. The resulting classifiers were evaluated using the verification test on the complete FRGC v2 target set of 4007 images using an all vs all test. Figures 28 and 29 show the results of the experiments.

As can be observed in Fig. 28, which shows the EER (Equal Error Rate) and the FRR (False Rejection Rate) at a FAR (False Accept Rate) of 0.1% as a function of the number of LDA components, if the number of LDA components is above a certain value, the performance of the classifier remains nearly constant. For the EER this value is 10, resulting in an EER of approx. 2.2% and for the FRR at FAR = 0.1% it is around 25 resulting in a FRR of approx. 12%. The conclusion we can derive from this is that although the human face seems a complicated 3D surface, only 10–25 floating point numbers suffice to discriminate between faces of different subjects. This results in very compact feature vectors.
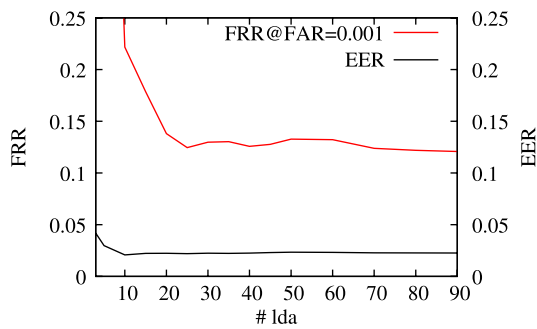
Fig. 28 FRR@FAR = 0.001 and EER of a single classifier as function of the number of LDA components

Fig. 30 FRR@FAR = 0.001 and EER of a single classifier as function of the resolution of the range image
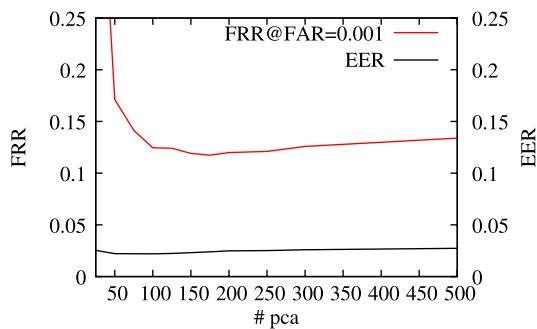

Fig. 29 FRR@FAR = 0.001 and EER of a single classifier as function of the number of PCA components

If we choose 25 components, each classifiers needs features with a length of 25 floating point numbers, which results in a total of $25 \times 30 = 750$ numbers or 3000 bytes for fusion of 30 region classifiers. Compared to the space required for a complete point cloud (around 600 kB) this means a huge compression.

Figure 29 shows the EER and the FRR at a FAR of 0.1% as a function of the number of PCA components for a fixed number of 25 LDA components. As can be observed, the optimal number of components for minimum FRR (around 12%) at FAR = 0.1% is around 175 and remains nearly constant in a range of 100–250. The EER has a minimum of 2.2% at 75 PCA components and remains nearly constant between 50 and 150 components. Above 150 components the EER gradually increases to 2.8% at 500 components.

Based on the above experiments, we chose to use 100 PCA components and 25 LDA components. For convenience, we chose the same number of PCA and LDA components for all region classifiers, although we observed different optima for different classifiers. For 150 PCA components and 40 LDA components and an in-plane resolution of 1.5 mm, for example, the best region classifier resulted in a rank-1 score of 96.1% and a FRR of 88.3% at FAR = 0.1%. For the smaller regions, however, large numbers of PCA and LDA components leads to overtraining (see e.g. Jain et al. 2000) and, hence, worse results. Optimisation of these parameters for the individual region classifier is, however, a time
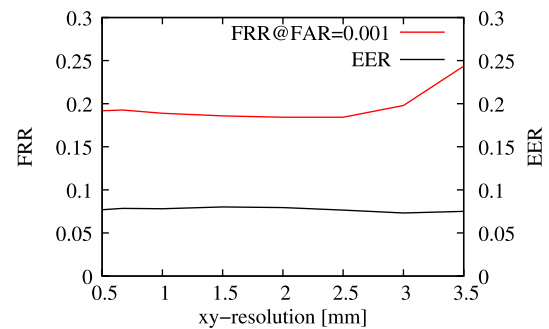
consuming process and, therefore, we made a compromise. In future research, we will further optimise these parameters and the regions which will likely result in further improvement of the results.

Another parameter that influences the results of the classifiers is the resolution of the range image generated by the registration module. A very high resolution in x- and y-directions may seem preferable, because all details can be represented. However, choosing a very high resolution may degrade classifier results, because of the curse of dimensionality (Jain et al. 2000) and imperfect interpolation for filling of holes. Choosing a very low resolution, on the other hand, will speed up processing, but will result in lower recognition rates. In Boom et al. (2006) we showed that for 2D face recognition using a PCA-LDA likelihood ratio classifier the recognition performance is relatively insensitive to image resolution. In Fig. 30 we can observe that the same is true for 3D face recognition. The figure shows the FRR at FAR = 0.1% and the EER for a single region classifier as a function of the resolution in x- and y-directions of the range image. Both FRR at FAR = 0.1% as well as the EER remain nearly constant at around 19% resp. 7.5% for a large range of resolutions. For most of our experiments we chose a resolution of 1.5 mm/pixel, resulting in range images of 75×87 pixels.

Again, for convenience we chose the same resolution for all region classifiers. In future work we will investigate the effect of the resolution on the individual region classifiers as well.

6.4 Classifier Performance and Fusion

We performed two evaluation experiments on the FRGC v2 data: a closed set identification experiment and an all vs all verification experiment.

6.4.1 Identification

For the identification experiment, the 4007 images of the FRGC are split into a gallery and a probe set. The gallery set
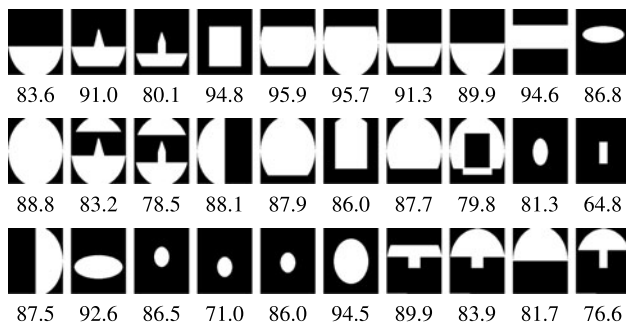
Fig. 31 Performance of individual region classifiers on FRGC v2 rank-1 identification experiment. The best rank-1 score for a single region classifier is 95.9% (number 5 in *the top row*)

Fig. 32 Fusion of region classifiers using majority voting using 1 and 2 training sets. The final result using 2×30 region classifiers from 2 training sets is 99.0%

consists of the first image of each subject in the data base, resulting in a set of 466 images. Most of these first images are neutral images, but not all of them. The remaining 3541 images are used as a probe set.

Figure 31 shows the rank-1 scores for 30 region classifiers trained on the frontal Bosphorus data set. The maximum rank-1 score obtained for a single region classifier was 95.9%. Similar results were obtained for the region classifiers trained on the 3DFace data set.

As described in Sect. 5, straightforward majority voting is used to fuse the rank-1 scores of the individual region classifiers. Figure 32 shows the result of adding the region classifiers one by one and taking the majority vote. The first curve shows the combination of 30 region classifiers trained on the Bosphorus frontal data set, which results in a final rank-1 score of 97.9% for 30 regions. The second curve combines region classifiers trained on the Bosphorus frontal data set and the 3DFace data set. The fusion of two times 30 region classifiers results in a rank-1 score of 99.0%. The region classifiers trained on the 3DFace data set used the alternative registration described in Sect. 3.7. Because the two registration approaches make different errors and the two datasets contain different variations of 3D facial shapes, fusion results in further improvement of the performance. Note that we used a complete separation of training and evaluation data. The training data are even acquired using a completely different process: using structured light as opposed to laser scanning for the FRGC v2 data base.

The drop in performance around the 9th region suggests that better region selection is possible. This is one of the subjects of our future research.

6.4.2 Verification

For the verification scenario, we use the all vs all experiment defined in the FRGC protocol. Since there are 4007 images in the FRGC v2 database, this results in 4007×4006 comparisons (images are not compared to themselves). For each region classifier, the VR at FAR = 0.1% is determined using a Tippet plot (see Sect. 5.2.3). Figure 33 shows this VR
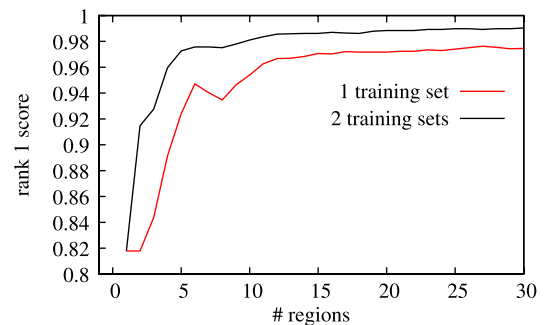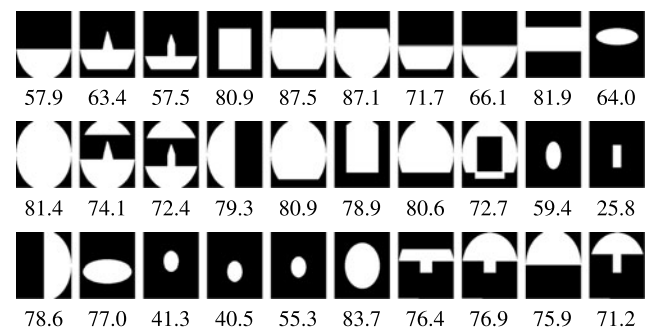

Fig. 33 Performance of individual region classifiers on FRGC v2 all vs all verification experiment. The best performance of a single region classifier is VR = 87.5% @ FAR = 0.1% (*5th on top row*)

for all the region classifiers trained on the Bosphorus database. The best performance of a single region classifier is VR = 87.5% or FRR = 12.5% at FAR = 0.1%. Similar results are obtained using the 3DFace dataset for training and the alternative registration of Sect. 3.7.

For fusion, we use the approach as described in Sect. 5.2.3. We used the FRGC v1 data to determine the thresholds for the region classifiers for a given projected FAR_p. Figure 34 shows the FRR at FAR = 0.1% for fused classifiers of 30, 2×30, 3×30 and 4×30 regions as a function of the projected FAR_p. The minimum FRR of 5.4% is reached at $FAR_p = 0.00023$ for the 4×30 region classifier (fusion of 30 region classifiers trained on Bosphorus data and 30 region classifiers trained on the 3DFace dataset using both the standard as well as the alternative registration).

We can also observe that like for the other parameters, there is a reasonably large range where the performance is more or less constant, i.e. the exact choice of FAR_p is not very critical.

As an extra indicator of performance, the EER is shown as well in Fig. 35. Note that unlike the FRR the EER does not show a clear minimum for the chosen range of FAR_p. The best performance for the 4×30 regions is an EER of 1.2%.

Fig. 34 Performance of fused region classifiers on FRGC v2 all vs all verification experiment as a function of the projected FAR_p. The best performance for the 4×30 regions is a FRR of 5.4% or a VR of 94.6%
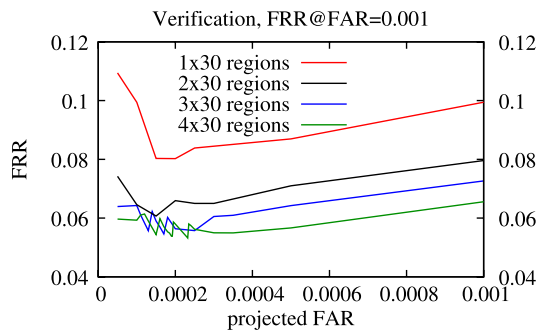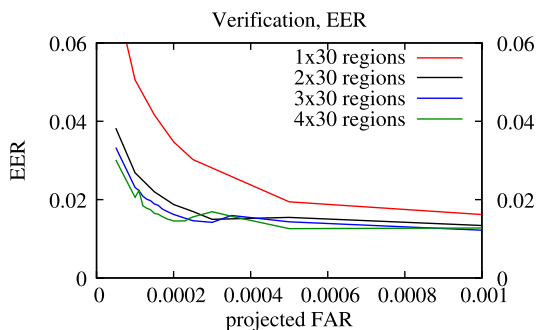

Fig. 35 Performance of fused region classifiers on FRGC v2 all vs all verification experiment as a function of the projected FAR_p. The best performance for the 4×30 regions is an EER of 1.2%

As mentioned in Sect. 5.2.3 further performance improvements may be expected by more in depth research into the optimal combination of region classifiers. This will be part of our future research.

6.5 Comparison to Tang

Although the main objective of the paper of Tang et al. is to show that using the nose area to define the symmetry plane (NSP) results in a more accurate estimate of the symmetry plane than by using the full face (FSP) and not to perform an extensive comparison of the recognition results to the state of the art, it still is interesting to compare our results with theirs, because our registration method is based on the same landmark structures as Tang's.

While Tang's and our method use the same landmarking structures, the approaches to estimate the corresponding parameters differ in several aspects. Whereas Tang et al. base the estimation of the symmetry on a small area around the nose, we determine the symmetry plane using a large area around the nose, which, in our opinion, gives a more reliable result. Another difference is the definition of the tip of the nose, which we defined as the intersection of two lines rather than the point with maximum curvature. Furthermore, our method uses a course-to-fine robust approach in each

Table 6 Comparison of our approach to Tang's. Reported numbers are EER's on an all vs all experiment on the FRGC v1 data

Method	EER
Tang-manual	6.1%
Tang-FSP	7.1%
Tang-NSP	5.5%
Spreeuwers 1	0.7%
Spreeuwers 2×30	0.3%

stage, whereas Tang's method doesn't. Finally, our 3D face classifier is far more advanced than Tang's. Tang et al. use average distance between profiles of the face as a distance measure for face comparison.

Tang et al. perform a single verification experiment on the FRGC v1 3D data set, which consists of a part of the data of the FRGC v2 data set (the Fall2003range data). This data set consists of 943 images of 275 subjects and is regarded as a relatively easy 3D data set. The experiment Tang et al. performed was one-to-one verification for all data, resulting in a 943×943 score matrix. Tang et al. only report equal error rates (EER) where, generally, false reject rates (FRR) at a false accept rate (FAR) of 0.1% are reported. In order to compare our results to Tang's, we trained the classifiers using the Bosphorus and the 3Dface databases. Two score matrices were evaluated: one for a single classifier (Spreeuwers 1: trained on the Bosphorus data base using the region at the top right of Fig. 23) and a full classifier using vote-fusion (Spreeuwers 2×30), where the classifiers were trained on both the Bosphorus and the 3Dface databases and the threshold for the vote-fusion was tuned using the Spring2004range and the Fall2004range data from the FRGC v2 data. The results are presented in Table 6.

Note that the classifiers we used for this test were not specifically optimised for EER. We can also conclude that unlike Tang's full face symmetry (FSP) approach, our robust approach to symmetry plane estimation does not suffer from inaccuracy and performs even far better than Tang's nose symmetry (NSP) approach.

6.6 Comparison to State of the Art: Processing Speed

Processing speed is one of the major advantages of our approach. Table 7 shows for a number of top-ranking 3D face recognition methods the times required for registration and comparison of two images together with the processors that were used to perform the calculations as reported in the publications. Of course, processing times are difficult to compare, because they not only depend on the used processors (which were comparable for all methods), but also on implementation, used programming language, amount and speed of storage etc. However, since our method is an order of

Table 7 Computation times for registration and number of comparisons per second of 3D recognition methods. Boehnen et al. do not specify the registration time, but it is likely to take several seconds as it involves ICP to an average model

Method	Processor	Prep. [sec]	Comp. per sec.
Queirolo et al. (2010)	3.4 GHz P4	–	0.25
Faltemier et al. (2008a)	2.4 GHz P4	7.5	0.36
Al-Osaimi et al. (2009)	Core 2 Quad	4	10
Kakadiaris et al. (2007)	3.0 GHz P4	15	1 000
Boehnen et al. (2009)	2.2 GHz T7500	?	12 500
Alyüz et al. (2009)	2.7 GHz Core i7	131	20 000
Spreeuwers 1	2.8 GHz P4	2.5	670 000
Spreeuwers 30	2.8 GHz P4	2.5	22 300
Spreeuwers 2 × 30	2.8 GHz P4	2.5	11 150
Spreeuwers 4 × 30	2.8 GHz P4	2.5	5 575

Table 8 Estimated times for identification of a single probe using a gallery of 466 subjects of the FRGC v2 data. Boehnen et al. did not report registration times which are likely to take several seconds

Method	Total time [sec]	rank-1 score
Queirolo et al. (2010)	1864	98.4%
Faltemier et al. (2008a)	1312	97.2%
Al-Osaimi et al. (2009)	50.6	96.5%
Kakadiaris et al. (2007)	15.5	97.0%
Boehnen et al. (2009)	0.03 + reg	95.5%
Alyüz et al. (2009)	131	97.5%
Spreeuwers 1	2.5	95.9%
Spreeuwers 30	2.5	98.0%
Spreeuwers 2 × 30	2.5	99.0%

10–1000 times faster than the competing methods and was run on rather modest hardware, we can safely state that it is one of the fastest methods around. Also, the method could as easily profit from parallel processing and more modern processors as any other method (it currently runs in a single thread on an old 2.8 MHz Pentium 4).

Queirolo et al. (2010) report an average time for comparison of two faces of 4 seconds. Faltemier et al. (2008a) require 7.5 seconds for data preprocessing and rough registration. Face matching of 28 regions then takes 2.4 seconds. Al-Osaimi et al. (2009) extract a cropped pose-corrected 3D facial range image which is used as a template and takes 4 seconds. Face comparison is based on ICP and takes 100 ms per face. Kakadiaris et al. (2007) perform registration to a spin model which takes 15 seconds and use extracted features for matching which is very fast at 1 000 matches per second. Boehnen et al. do not report the time for prealignment for their signature search, but because it involves an ICP step for registration to an average model, it is likely to take several seconds. Their matching for 8 regions on the signatures is very fast with 100 000 matches per second per region or 12 500 full matches per second. Alyüz et al. need a total of 131 seconds for preprocessing and registration using regional ICP. Matching is again very fast at 20 000 matches per second.

In Table 7 four of our classifiers are shown: "Spreeuwers 1" is the best performing single region classifier, "Spreeuwers 30" is a fusion of 30 region classifiers trained using the Bosphorus database and "Spreeuwers 2 × 30" is a fusion of 30 region classifiers trained on the Bosphorus database and 30 region classifiers trained on the 3DFace dataset with alternative registration. The "Spreeuwers 4 × 30" classifier consists of 4 sets of 30 region classifiers trained on the 3DFace and the Bosphorus datasets using both types of registration. Using registration to an intrinsic coordinate system

as an independent step in the recognition chain instead of a one-to-all registration approach results in a significant advantage to ICP-like approaches. As is shown in Table 7, our registration method also compares favourably to registration to AFM (average face models) methods, taking on average around 2.5 seconds on point clouds of 50 000–100 000 points like those in the FRGC v2 data set. Comparison itself, consisting of only a few matrix multiplications on the reduced set of features is very fast with 670 000 regions per second or 22 300 and 11 150 per second for the 30 and 2 × 30 region classifiers. Even the fusion of 120 region classifiers (4 × 30) is still very fast with 5575 comparisons per second.

As argued before, methods applying an one-to-all registration approach are impractical for the identification scenario. Table 8 shows for top-ranking methods the times required for identification of a single probe image to a gallery of 466 subjects from the FRGC v2 data. Identification takes in this case 1 registration/preprocessing of the probe image and 466 comparisons. Table 8 also shows the maximum rank-1 performance that was reported for the different methods. Our approach gives the highest rank-1 performance and is more than 700 times faster than the second best performing approach of Queirolo et al. (2010). The fastest competing approach is by Boehnen et al. (2009), which probably is about as fast as our method (registration to an average model using ICP takes probably several seconds), but its rank-1 performance is far worse.

In Table 9 we estimated the time required to calculate the complete 4007×4007 score matrix for the all vs all experiment on the FRGC v2 data. We assume that the matrix is symmetric and we do not compare images to the image self. This means that a total of 4007 registrations/preprocessing are required and $4007 \times 4006/2 = 8\,026\,021$ comparisons. Clearly, our method has a huge advantage in computation speed. Boehnen et al. (2009) did not report verification results and are therefore not included in the table.

Table 9 Estimated times to calculate the complete 4007×4007 matrix of the all vs all verification experiment on the FRGC v2 data. Boehnen et al. did not report verification results

Method	Total time [hours]
Queirolo et al. (2010)	8900
Faltemier et al. (2008a)	6250
Al-Osaimi et al. (2009)	227
Kakadiaris et al. (2007)	19
Alyüz et al. (2009)	146
Spreeuwers 1	2.8
Spreeuwers 30	2.9
Spreeuwers 2×30	3.0
Spreeuwers 4×30	3.2

6.7 Comparison to State of the Art: Performance

In this section we compare the performance of our approach using registration to an intrinsic coordinate system and fusion of multiple PCA-LDA likelihood region classifiers to the best performing 3D face recognition methods that published results on the FRGC v2 database. Two experiments were selected. The first experiment is an identification experiment using a gallery of all 466 first images of the dataset and the rest of the 4007 images as a probe set. For this experiment rank-1 results are reported. The second experiment is a verification experiment where each of the 4007 images is compared to all others. The verification rate (VR) at FAR $= 0.1\%$ is reported for 3 different masks of the data: mask I (within semester recordings), mask II (within year recordings) and mask III (between semester recordings). In some publications separate results are published with only neutral faces in the gallery or as reference and probe images. We only show results on the most difficult case where all faces with expressions are included and even some of the faces in the gallery of the identification experiment show expressions.

For these experiments we used registration to a range image with a resolution of 1.5 mm, resulting in range images of 75×87 pixels. For the region classifiers the number of PCA and LDA components was set to 100 resp. 25. For the identification experiment we fused 30 region classifiers trained on the Bosphorus data and 30 region classifiers trained on the 3DFace data with alternative registration resulting in a total of 60 region classifiers. For the verification experiment we used 30 region classifiers for both registration methods and datasets, resulting in a total of 120 region classifiers. Straightforward majority voting was used as before and for the verification the vote fusion method as described in Sect. 5 with projected $\text{FAR}_p = 0.00023$. The thresholds for the region classifiers were determined using the FRGC v1 data.

Table 10 Comparison of rank-1 score on FRGC v2 data to top performing 3D face recognition methods

Method	rank-1 score
Kakadiaris et al. (2007)	97.0%
Faltemier et al. (2008a)	97.2%
Alyüz et al. (2009)	97.5%
Queirolo et al. (2010)	98.4%
Spreeuwers	99.0%

Table 11 Comparison of verification rates at FAR $= 0.1\%$ on FRGC v2 data to top performing 3D face recognition methods

Method	Verification rate VR @ FAR $= 0.1\%$			
	Mask I	Mask II	Mask III	All vs all
Maurer et al. (2005)				86.5%
Kakadiaris et al. (2007)	97.2%	97.1%	97.0%	
Faltemier et al. (2008a)			94.8%	93.2%
Alyüz et al. (2009)	85.8%	86.0%	86.1%	
Al-Osaimi et al. (2009)	94.6%	94.1%	94.1%	
Queirolo et al. (2010)			96.6%	96.5%
Spreeuwers	94.6%	94.6%	94.6%	94.6%

The results of the identification experiment are shown in Table 10. For this experiment our method obtains the best performance. As shown in Sect. 6.6 in addition our method is much faster than the other methods.

The results of the verification experiments are shown in Table 11. As can be observed, our method does not reach the highest verification rate, but still respectable scores are obtained that puts our method between the top contenders. As mentioned before, we plan to investigate more advanced fusion strategies and optimisation of the regions and region classifiers and expect further improvement of the verification scores (and identification scores as well).

Finally, Fig. 36 shows the ROC for the verification experiment for the three masks. Note that since the scores are actually counts of the number of votes, the range of the score is [0..120], because there are 120 region classifiers. This means that the ROC consists of 121 points rather than a continuous line and hence the staircase shaped curves.

7 Conclusions

We present a new fully automatic registration approach for 3D face recognition which registers 3D point clouds to an intrinsic coordinate system defined by the vertical symmetry plane through the nose, the slope of the nose and the tip of the nose. A robust approach to the registration is used where first on low resolution an exhaustive search strategy is used
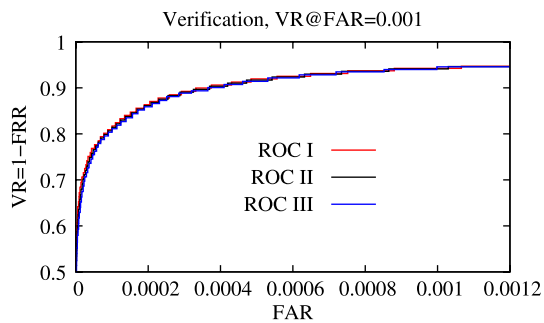
Fig. 36 ROC curves for the verification experiments for the 3 masks defined in the FRGC. Verification rates at FAR = 0.1% are 94.6%

to obtain rough estimates of the parameters and in a second stage for a smaller search range accurate estimation of the registration parameters is performed. Post processing includes resampling to a range image, spike removal and hole filling. The registration method is robust, accurate and fast: a C++ implementation takes an average of 2.5 seconds on a standard PC for images from the FRGC database. An alternative registration method was also developed based on the dent above the nose bridge and the point just below the tip of the nose instead of the tip of the nose and the slope of the nose bridge. In some cases this alternative approach is more robust against artifacts around the tip of the nose and severe facial expressions. Computation of this additional registration adds only negligible extra processing time to the average of 2.5 seconds.

A 3DFace classifier was developed by fusion of many region classifiers which are trained on specific regions of the face that are supposed to remain stable under variation of expression and occlusions by e.g. glasses, hair, caps etc. For the region classifiers we used PCA-LDA likelihood ratio classifiers. For the identification scenario the fusion of the classifiers is straightforward majority voting. For the verification scenario all region classifiers are first tuned for a specific projected FAR using a separate fusion training set. The resulting thresholds are then applied to the region classifiers and fusion takes again place by counting votes of the region classifiers. The resulting vote is regarded as a similarity score and the performance of the fused classifier. Because the classification only consists of a number of matrix multiplications, it is very fast at 670 000 comparisons per second for a single region classifier.

A system consisting of the described registration methods and 2 or 4 sets of 30 region classifiers was evaluated using the FRGC v2 database and the results were compared to the best performing methods. The region classifiers were trained using independent datasets: the Bosphorus dataset and the 3DFace dataset. The FRGC v1 dataset was used to find the thresholds for the region classifiers. Two tests of the FRGC were performed: a closed set identification test using the first image of all 466 subjects of the FRGC v2 dataset

and the remaining images as a probe set and an all vs all verification test. The performance of our method compares very favourably to the top ranking methods in the world: the identification rate is 99% for this test, while the previously best reported identification rate was 98.4%. In addition, our method is more than 700 times faster than this method. For the verification experiment our method reaches an all vs all verification rate of 94.6%, which is not the highest score reported (97%), but is certainly one of the best scores ever obtained. Although less significant for the verification scenario, again the computation speed is much higher than for the competition: the complete calculation of the full 4007×4007 score matrix for this experiment takes only 3.2 hours including registration.

Because registration is to an intrinsic coordinate system of the face, each image only has to be registered once, which is a huge advantage over one-to-one registration methods like ICP. Furthermore, feature extraction from the range images allows for a compact template: for the classifier based on 60 region classifiers with 25 LDA components, only 60×25 floating point numbers have to be stored or 6 kB. For 120 region classifiers the required storage for a template increases to a still very acceptable 12 kB. Unlike 3D point clouds, these features can be used for privacy preserving template protection schemes as well.

There are many possibilities for further improvement of the proposed method and extension to other fields. The registration method still leaves room for further improvement, e.g. using more advanced hole filling and alternative approaches to nose tip localisation and determination of the slope of the nose bridge, e.g. using more advanced robust estimation techniques, like MLESAC (see Torr and Zisserman 2000) instead of RANSAC. Currently, for convenience, the parameters for all of the region classifiers are chosen the same. First experiments show however, that different region classifiers perform better or worse for different resolutions of the range image and numbers of PCA and LDA components. For fusion we used very simple methods based on majority voting. In literature, more advanced approaches to automatically select the best combination of region classifiers and calculate weights for the voting process are known, which would likely result in improved performance. In future work we will investigate these improvements to the method in depth. Finally, the proposed registration method works well for variations not represented in the FRGC v2 data, like large variation in pose, occlusions by caps or hats, glasses and hands. In some other datasets like the 3DFace dataset and the Bosphorus dataset, these variation are present. In future research we intend to further investigate 3D face recognition for large variations of pose, expression and various occlusions. We plan to make the 3D face registration available as open source software so other 3D face comparison methods can be tested using this registration.

References

3DFace (2009). 3D face project web page. http://www.3dface.org/home/welcome.html.

Achermann, B., Jiang, X., & Bunke, H. (1997). Face recognition using range data. In *Proceedings of the international conference on virtual systems and multimedia* (pp. 129–136). Geneva: IEEE Press.

Al-Osaimi, F., Bennamoun, M., & Mian, A. (2009). An expression deformation approach to non-rigid 3D face recognition. *International Journal of Computer Vision*, *81*(3), 302–316. doi:10.1007/s11263-008-0174-0.

Alyüz, N., Gökberk, B., & Akarun, L. (2009). Regional registration and curvature descriptors for expression resistant 3D face recognition. In *Proceedings of the IEEE 17th signal processing and communications applications conference (SIU-2009)* (pp. 544–547).

Bazen, A. M., & Veldhuis, R. N. J. (2004). Likelihood ratio-based biometric verification. *IEEE Transactions on Circuits and Systems for Video Technology*, *14*(1), 86–94.

Besl, P. J., & McKay, N. D. (1992). A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *14*(2), 239–256. doi:10.1109/34.121791.

Beumer, G. M., Tao, Q., Bazen, AM, & Veldhuis, R. N. J. (2006). A landmark paper in face recognition. In *Proceedings of the 7th international conference on automatic face and gesture recognition (FGR 2006)* (pp. 73–78). Los Alamitos: IEEE Computer Society Press.

Boehnen, C., Peters, T., & Flynn, P. J. (2009). 3d signatures for fast 3D face recognition. In *Proceedings of the third international conference on advances in biometrics (ICB '09)* (pp. 12–21). Berlin: Springer.

Boom, B. J., Beumer, G. M., Spreeuwers, L. J., & Veldhuis, R. N. J. (2006). The effect of image resolution on the performance of a face recognition system. In *Proceedings of the 9th international conference on control, automation, robotics and vision (ICARCV)*, Singapore, Malaysia (pp. 409–414).

Boom, B. J., Spreeuwers, L. J., & Veldhuis, R. N. J. (2007). Automatic face alignment by maximizing similarity score. In A. Fred & A. K. Jain (Eds.), *Proceedings of the 7th international workshop on pattern recognition in information systems (Madeira, Portugal, 2007)* (pp. 221–230). Madeira: INSTICC Press. Biosignals.

Bowyer, K. W., Chang, K., & Flynn, P. (2006). A survey of approaches and challenges in 3D and multi-modal 3d + 2d face recognition. *Computer Vision and Image Understanding*, *101*(1), 1–15. doi:10.1016/j.cviu.2005.05.005.

Brent, R. P. (1973). *Algorithms for minimization without derivatives*. Englewood Cliffs: Prentice Hall, Chap. 5.

Buhan, I. R., Doumen, J. M., Hartel, P. H., Tang, Q., & Veldhuis, R. N. J. (2010). Embedding renewable cryptographic keys into continuous noisy data. *International Journal of Information Security*, *9*(3), 193–208.

Cartoux, J. Y., Lapreste, J. T., & Richetin, M. (1989). Face authentification or recognition by profile extraction from range images. In *Proceedings of the workshop on interpretation of 3D scenes* (pp. 194–199).

Chen, C., Veldhuis, R. N. J., Kevenaar, TAM, & Akkermans, A. H. M. (2009). Biometric quantization through detection rate optimized bit allocation. *EURASIP Journal on Advances in Signal Processing*, *2009*, 784–834.

Colombo, A., Cusano, C., & Schettini, R. (2006). Detection and restoration of occlusions for 3D face recognition. In *IEEE international conference on multimedia and expo* (pp. 1541–1544). doi:10.1109/ICME.2006.262837.

Faltemier, T., Bowyer, K., & Flynn, P. (2008a). A region ensemble for 3-d face recognition. *IEEE Transactions on Information Forensics and Security*, *3*(1), 62–73. doi:10.1109/TIFS.2007.916287.

Faltemier, T. C., Bowyer, K. W., & Flynn, P. J. (2008b). Using multi-instance enrollment to improve performance of 3D face recognition. *Computer Vision and Image Understanding*, *112*(2), 114–125. doi:10.1016/j.cviu.2008.01.004.

Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, *24*(6), 381–395. http://doi.acm.org/10.1145/358669.358692.

Gokberk, B., Irfanoglu, M. O., & Akarun, L. (2006). 3D shape-based face representation and feature extraction for face recognition. *Image and Vision Computing*, *24*(8), 857–869. doi:10.1016/j.imavis.2006.02.009.

Gonzalez-Rodriguez, J., Fiérrez-Aguilar, J., Ortega-Garcia, J., & Lucena-Molina, J. J. (2002). Biometric identification in forensic cases according to the bayesian approach. In *Proceedings of the international ECCV 2002 workshop Copenhagen on biometric authentication* (pp. 177–185). London: Springer.

van der Heijden, F., & Spreeuwers, L. J. (2007). Image processing. In *Multimedia retrieval* (pp. 163–165). Berlin: Springer.

Hesher, C., Srivastava, A., & Erlebacher, G. (2003). A novel technique for face recognition using range imaging. In *7th international symposium on signal processing and its applications* (Vol. 2, pp. 201–204). doi:10.1109/ISSPA.2003.1224850.

Jain, A. K., Duin, R. P. W., & Mao, J. (2000). Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(1), 4–37. doi:10.1109/34.824819.

Kakadiaris, I. A., Passalis, G., Toderici, G., Murtuza, M. N., Lu, Y., Karampatziakis, N., & Theoharis, T. (2007). Three-dimensional face recognition in the presence of facial expressions: an annotated deformable model approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*(4), 640–649. doi:10.1109/TPAMI.2007.1017.

Kelkboom, E. J. C., Zhou, X., Breebaart, J., Veldhuis, R. N. J., & Busch, C. (2009). Multi-algorithm fusion with template protection. In *IEEE 3rd international conference on biometrics: theory, applications, and systems, 2009 (BTAS '09)* (pp. 1–8). Washington: IEEE Computer Society Press.

Kelkboom, E. J. C., Garcia-Molina, G., Breebaart, J., Veldhuis, R. N. J., Kevenaar, T. A. M., & Jonker, W. (2010). Binary biometrics: an analytic framework to estimate the performance curves under gaussian assumption. *IEEE Transactions on Systems, Man and Cybernetics. Part A. Systems and Humans*, *40*(3), 555–571.

Kuncheva, L. I., Whitaker, C. J., & Duin, R. P. W. (2003). Limits on the majority vote accuracy in classifier fusion. *Pattern Analysis and Applications*, *6*, 22–31.

Maurer, T., Guigonis, D., Maslov, I., Pesenti, B., Tsaregorodtsev, A., West, D., & Medioni, G. (2005). Performance of geometrix activeid™ 3d face recognition engine on the frgc data. In *Proceedings of the 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)—workshops* (p. 154). Washington: IEEE Computer Society. doi:10.1109/CVPR.2005.581.

Medioni, G., & Waupotitsch, R. (2003). Face modeling and recognition in 3-d. In *Proceedings of the IEEE international workshop on analysis and modeling of faces and gestures (AMFG '03)* (p. 232). Washington: IEEE Computer Society.

Mian, A., Bennamoun, M., & Owens, R. (2007). An efficient multimodal 2D–3D hybrid approach to automatic face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*(11), 1927–1943. doi:10.1109/TPAMI.2007.1105.

Papatheodorou, T., & Rueckert, D. (2005). Evaluation of 3D face recognition using registration and pca. In T. Kanade, A. K. Jain, & N. K. Ratha (Eds.), *Lecture notes in computer science: Vol. 3546. Proceedings of the 5th international conference on audio- and video-based biometric person authentication (AVBPA 2005)* (pp. 997–1009). Hilton Rye Town: Springer.

Papatheodorou, T., & Rueckert, D. (2007). *3D face recognition*. Vienna: I-Tech Education and Publishing.

Phillips, P. J., Flynn, P. J., Scruggs, T., Bowyer, K. W., Chang, J., Hoffman, K., Marques, J., Min, J., & Worek, W. (2005). Overview of the face recognition grand challenge. In *Proceedings of the 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)* (pp. 947–954). Washington: IEEE Computer Society. doi:10.1109/CVPR.2005.268.

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1988). *Numerical recipes in C. The art of scientific computing*. Cambridge: Press Syndicate of the University of Cambridge. Chap. 10.

Queirolo, C. C., Silva, L., Bellon, O. R., & Segundo, M. P. (2010). 3D face recognition using simulated annealing and the surface interpenetration measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*, 206–219. http://doi.ieeecomputersociety.org/10.1109/TPAMI.2009.14.

Ross, A. A., Jain, A. K., & Nandakumar, K. (2006a). Levels of fusion in biometrics. In *International series on biometrics* (Vol. 6, pp. 59–90). Berlin: Springer. Chap 4. doi:10.1007/0-387-33123-9.

Ross, A. A., Jain, A. K., & Nandakumar, K. (2006b). Score level fusion. In *International series on biometrics* (Vol. 6, pp. 91–142). Berlin: Springer. Chap. 5. doi:10.1007/0-387-33123-9.

Salah, A. A., Alyüz, N., & Akarun, L. (2007). Alternative face models for 3D face registration. In *Proceedings of the SPIE conference on vision geometry XV*, San Jose, CA, USA (Vol. 6499). Berlin: Springer. doi:10.1117/12.705860.

Savran, A., Alyüz, N., Dibeklioğlu, H., Çeliktutan, O., Gökberk, B., Sankur, B., & Akarun, L. (2008). Bosphorus database for 3D face analysis. In B. Schouten, N. C. Juul, A. Drygajlo, & M. Tistarelli (Eds.), *Lecture notes in computer science: Vol.* 5372. *Biometrics and identity management* (pp. 47–56). Berlin: Springer. Chap 6. doi:10.1007/978-3-540-89991-4_6.

Scheenstra, A., Ruifrok, A., & Veltkamp, R. C. (2005). A survey of 3D face recognition methods. In *Lecture notes in computer science* (pp. 891–899). Berlin: Springer.

Silva, L., Bellon, O. R. P., & Boyer, K. L. (2005). Precision range image registration using a robust surface interpenetration measure and enhanced genetic algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*, 762–776. doi:10.1109/TPAMI.2005.108.

Spreeuwers, L. J., Boom, B. J., & Veldhuis, R. N. J. (2007). Better than best: matching score based face registration. In R. N. J. Veldhuis & H. S. Cronie (Eds.), *Proceedings of the 28th symposium on information theory in the Benelux*, Enschede, The Netherlands (pp. 125–132). Eindhoven: Werkgemeenschap voor Informatie- en Communicatietheorie.

Tang, X., Chen, J., & Moon, Y. (2008). Accurate 3D face registration based on the symmetry plane analysis on nose regions. In *Proceedings of the 16th European signal processing conference (EUSIPCO 2008)*, Lausanne, Switzerland.

Tao, Q., & Veldhuis, R. N. J. (2007). Optimal decision fusion for a face verification system. In L. Seong-Whan & Z. L. Stan (Eds.), *Image processing, computer vision, pattern recognition. Proceedings of the 2nd international conference on biometrics*. Seoul, Korea (pp. 958–967). Berlin: Springer.

Tao, Q., & Veldhuis, R. N. J. (2008). Hybrid fusion for biometrics: Combining score-level and decision-level fusion. In *Proceedings of the IEEE computer society conference on computer vision and pattern recognition, workshop on biometrics* (pp. 1–6). Anchorage: IEEE Computer Society Press.

Tao, Q., & Veldhuis, R. N. J. (2009). Threshold-optimized decision-level fusion and its application to biometrics. *Pattern Recognition*, *42*(5), 823–836.

Tao, Q., van Rootseler, R. T. A., Veldhuis, R. N. J., Gehlen, S., & Weber, F. (2007). Optimal decision fusion and its application on 3D face recognition. In A. Bromme, C. Busch, & D. Huhnlein (Eds.), *Proceedings of the special interest group on biometrics and electronic signatures*, Darmstadt, Germany, Gesellschaft fur Informatik e.V. (pp. 15–24). Berlin: GI-Edition.

Torr, P. H. S., & Zisserman, A. (2000). Mlesac: a new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, *78*, 138–156. doi:10.1006/cviu.1999.0832.

Veldhuis, R. N. J., Bazen, A. M., Booij, W. D. T., & Hendrikse, A. J. (2006). Hand-geometry recognition based on contour landmarks. In M. Spiliopoulou, R. Kruse, C. Borgelt, A. Nurnberger, & W. Gaul (Eds.), *Studies in classification, data analysis, and knowledge organisation. Proceedings of the 29th annual conference of the Gesellschaft fur Klassifikation e.V.*, Magdeburg, Germany (pp. 646–653). Berlin: Springer.