Normalisation of 3D Face Data

Chris McCool, George Mamic, Clinton Fookes and Sridha Sridharan
Image and Video Research Laboratory
Queensland University of Technology,
2 George Street, Brisbane, Australia, 4001, GPO 2434
{c.mccool, g.mamic, c.fookes, s.sridharan}@qut.edu.au

Abstract

In this paper a novel appraoch to normalising 3D face data is proposed. The aim of this normalisation process is to retain as much of the raw data as possible and so it is imperative that no distortion occurs to this data, for instance no median or mean filtering is applied. The normalisation is a multi-stage approach and is applied to the Face Recognition Grand Challenge v1.0 database. Experiments show that this normalisation process results in 3D face surfaces which are significantly more discriminable compared to those surfaces used for the FRGC experiments, when comparing surfaces using the Manhattan distance the Equaul Error Rate improves from 21.65% *for the FRGC normalisation procedure to* 13.81% *for the proposed method.*

I. Introduction

Research into face recognition has blossomed in the past decade because the face is a socially accepted biometric that is used to identify people for drivers licences and passports. However, traditional face recognition methods which use intesity images (2D data) have several limitations. The 2002 Face Recognition Vendor Test (FRVT) [1] found that the performance of state of the art face recognition systems degraded significantly in the presence of both pose and illumination variations. To overcome these limitations researchers have examined alternate methods of capturing face data.

The Face Recognition Grand Challenge (FRGC) [2] examined the use of 3D face data to perform face recognition. The 3D face data is considered to be robust to both pose and illumination variations and so can alleviate the limitations of the 2D face data. Robustness to pose variation is obtained because the full set of 3D points are available and so both in-plane and out-of-plane rotations can be normalised; in-plane rotations are rotations that occur within the image plane of the camera (caused by tilting the head from side to side) and out-of-plane rotations are rotations about the image plane of the camera (tilting the head up or down and turning the head to a profile view). The 3D data is robust to illumination variation because the underlying 3D strucutre is being captured. The largest available 3D face database is the FRGC database and there have been several methods proposed to normalise this data.

In [3], Chang et al. normalised the FRGC 3D face data by performing a set of rotations along with median filtering and linear interpolation; this is very similar to method used to perform normalisation for the FRGC baseline experiments [2]. The first step in their normalisation process is to rotate the surface using three landmark points: left eye, right eye and center of the lower lip. The angle between these points on the 3D face are compared to predefined reference points so that the angle between the X- and then Y-axes are normalised. The angle around the Z-axis is then normalised by calculating the angle between the two eyes. A fourth landmark point, the nose, is then used to define the largest range value in the depth image. Invalid data is removed by applying hole and spike removal along with median filter and linear interpolation of valid points. Chang et al. recently outlined another normalisation for the FRGC 3D face data which performs normalisation by first generating a binary mask of valid and invalid data points.

In [4] a normalisation process is described which utilises a three step process to remove invalid data points. The first step considers the range image as a binary image and isolated regions are removed using a morphological opening operator. The variance of the remaining range values are then calculated and those points which lie outside a threshold are removed. The underlying 3D data is then examined by removing those points which have an angle to the optical axis which is greater than a threshold value.

There are several issues with these two normalisation methods. In [3] median filtering is applied to remove invalid data, this results in a smoothed 3D face surface which alters both invalid data points and valid data points. The alternate approach in [4] uses empirically derived threshold values (for variance and angle to the optical axis) that are applied to every image in the database, this means that the normalisation approach will need to be altered for each 3D face data. Also, both methods identify invalid data using the depth or Z values and so do not consider the full 3D face surface.

This paper examines a novel method for performing normalisation of 3D face data using the full 3D data. The aim of this procedure is to retain as much of the raw 3D data as is possible and identifies invalid data using only the statistics from the 3D face surface being normalised, no database specific threhold values are derived. The final output of this procedure consists of the raw 3D face data points which are considered to be valid, no smoothing functions are applied to this data.

II. 3D Data Normalisation

The 3D data for the FRGC database was captured using a Konica Minolta Vivid 900/910 [5]. This sensor captures the 3D using a laser light which takes approximately 2.5 seconds to capture a 3D image of size 640×480, an example 3D face capture is provided in Figure 1. An issue that has been noted by several researchers is that the 3D face data captured using the Minolta Vivid 900 often results in erroneous data from around the eye, eyebrow and nostril regions.

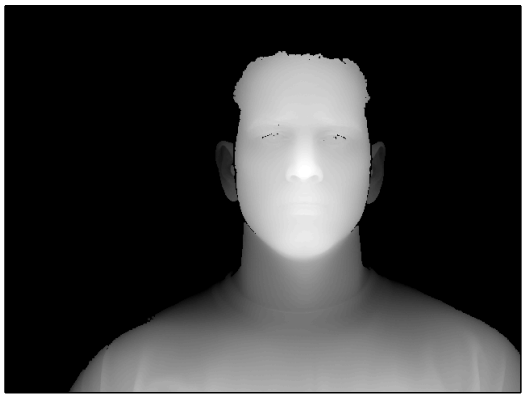


Fig. 1. An example 3D face capture using the Minolta Vivid 900.

Chang et al. [3] noted that regions such as the eyes and eyebrows can have missing or invalid data. They perform median filtering and linear interpolation to fill in the missing points and remove the invalid data, however, this appears to be done only on the range data (using the Z data and not the X, Y and Z data). An example of the erroneous data around the eye is shown in Figure 2.

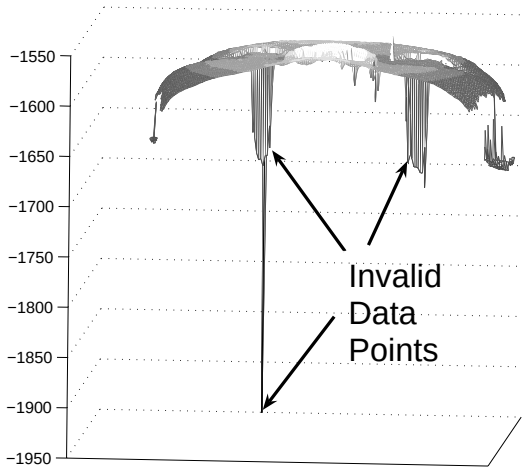


Fig. 2. An image of the raw 3D face surface, highlighted are the invalid data points in the eye region. It can also be seen that there are invalid points at the near the cheek region.

III. Proposed Method

The aim of this normalisation procedure is to retain as much of the valid 3D data as possible. This valid 3D data is then registered so that comparisons between to 3D face images can be easily performed. The three steps of this procedure are:

1) cropping out a region of interest (ROI),
2) removing invaid data (noise removal), and
3) rotating and registering the data.

A flowchart of this procedure is provided in Figure 3.

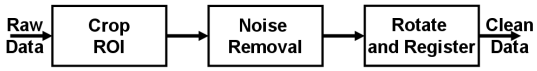


Fig. 3. A flowdiagram of the normalisation procedure.

A. ROI Cropping

The first step in this normalisation procedure is to extract the 3D face from the background. To do this two landmark points, the left eye and the right eye, are used to extract the face and exclude background as well as the shoulder, torso and neck region.

The square ROI of the 3D face is extracted using the left eye position (x_l, y_l, z_l) and right eye position (x_r, y_r, z_r). These eye positions are used to define the eyewidth,

$$eyewidth = \sqrt{(x_l - x_r)^2 + (y_l - y_r)^2 + (z_l - z_r)^2}, \quad (1)$$

which is then used to define a mask of dimensions $1\frac{1}{2}eyewidth \times 1\frac{1}{2}eyewidth$, an example of this mask region is shown in Figure 4. This is a similar to process to the one used in previous work by this author in [6].

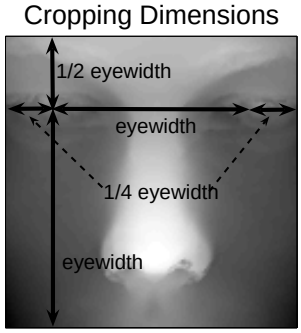


Fig. 4. An example of the cropping dimensions used to extract the ROI for the 3D face.

The cropped region is assumed to contain only 3D face data or the face surface. This assumption is used in the following noise removal procedure to gather statistics about the face surface to identify invalid data points.

B. Noise Removal

The 3D data that has been captured often contains erroneous points, particularly around the eye and eyebrow region. The noise removal procedure aims to remove all the erroneous points so that only valid 3D data points remain.

The noise removal procedure is conducted in three steps and each step uses the statistics of the entire 3D surface (and not just the range values) to identify erroneous points. The first two steps identify erroneous points based on a distance transform and a face shell. The final step removes the remaining erroenous points by examining the Gaussian surface curvatures, a flow diagram showing these steps is provided in Figure 5.

1) Distance Transform: This step aims to simultaneously cull oultying points and refine the region of interest (ROI). To achieve this the mean of the face surface is calculated and those points which lie outside a statistically reasonable boundary are removed. The mean and standard deviation of the cropped face region by considering each set of coordinates X,Y and Z separately and so a mean and standard deviation value is calculated for each grid

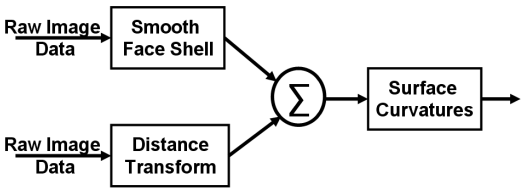


Fig. 5. A flow diagram showing the processes used to remove erroneous points.

such that $\mu_{crop} = (\mu_{X,crop}, \mu_{Y,crop}, \mu_{Z,crop}$ and $\sigma_{crop} = (\sigma_{X,crop}, sigma_{Y,crop}, \sigma_{Z,crop})$.

The statistics from each grid, μ_{crop} and σ_{crop}, are used to remove invalid data from the 3D face surface. A point on any grid which is four standard deviations from the mean is considered to be invalid and so is removed, this process is outline in Algorithm 1.

Algorithm 1 Distance Transform

Require: Cropped X, Y and Z data
1: μ_{crop} = Mean(X, Y, Z)
2: σ_{crop} = StdDev(X, Y, Z)
3: **if** $abs((X, Y, Z) - \mu_{crop}) > 4 \times \sigma_{crop}$ **then**
4: discard 3D point
5: **end if**
6: **return** Valid X, Y and Z data

2) Face Shell: A smooth face shell is calculated by applying a Gaussian filter (of size 4×4). This face shell, F_{shell}, is then used to calculate the distance at every point to the raw data, resulting in a distance map D_{shell}. The mean and standard deviation of D_{shell}, μ_{shell} and σ_{shell}, are then used to identify invalid data points by discarding those points which lie four standard deviations from the mean, this process is described in algorithm 2.

Algorithm 2 Face Shell

Require: Valid data for X, Y and Z data
1: F_{shell} = GaussianSmooth(X,Y,Z)
2: $D_{shell} = F_{shell} - (X, Y, Z)$
3: μ_{shell} = Mean(D_{shell})
4: σ_{shell} = StdDev(D_{shell})
5: **if** $abs(D_{shell} - \mu_{shell}) > 4 \times \sigma_{shell}$ **then**
6: discard 3D point
7: **end if**
8: **return** Valid X, Y and Z data

3) Gaussian Surface Curvatures: The Gaussian surface curvature is an intrinsic measure of curvature. This measure can be calculated using the coefficients of the first and second fundamental form of the surface such that,

$$\kappa = \frac{eg - f^2}{EG - F^2}. \quad (2)$$

The coefficients of the first fundamental form (E, F and G) can be calculated by taking the dot product of the partial derivatives,

$$\begin{aligned} E = &\quad X_u.X_u, \\ F = &\quad X_u.X_v, \text{and} \\ G = &\quad X_v.X_v, \end{aligned} \quad (3)$$

where X_u is the partial differentation in the horizontal direction and X_v is the partial differentiation in the vertical direction.

The coefficients of the second fundamental form (e, f and g) can be calculated using the dot product of the second partial derivatives with the unit normal,

$$\begin{aligned} e = &\quad X_{uu}.N, \\ f = &\quad X_{uv}.N, \text{and} \\ g = &\quad X_{vv}.N, \end{aligned} \quad (4)$$

where X_{uu} is the second partial derivative in the horizontal direction, X_{uv} is the partial derivative in the horizontal and vertical direction and X_{vv} is the second partial derivative in the vertical direction.

The Gaussian surface curvature κ is used to identify erroneous points by using the full 3D information available; simultaneously incorporating information from the X-, Y- and Z-grids. The erroneous points are highlighted by identifying those points which lie more than four standard deviations from the mean surface curvature. This step culls the majority of obviously erroneous points, however, there are often still points such as those around the eyes which are erroneous but not detected using this procedure.

The erroneous points which are not detected using this procedure are those which are surronded by erroneous points. Around the eye regions there are often points which lie in clusters well outside the obvious face boundary, or shell. Because these points lie as a cluster the use of a surface curvature will not detect these points. Therefore, the next two steps in this procedure aim to overcome this limitation with surface curvatures.

C. Rotation and Registration

This process aims to bring two 3D face images into alignment so that they can be compared using distnace measures, such as the Euclidean distance. The 3D face image is registered to a template image by first bringing the two landmark eye points, left and right eye, into alignment such that the points lie on the same y-line as well as the same (x, y) plane.

To ensure that the left and right eye y-line as well as the same (x, y) plane two rotations are performed. The first rotation about the Y-axis ensures that the eyes lie upon the same (x, y) plane and the second rotation about the Z-axis ensures that the eyes lie along the same y-line. Performing these rotation coarsely aligns the 3D face data to the template data, to obtain more accurate registration the Iterative Closest Point (ICP) algorithm is employed [7].

The ICP algorithm is used to perform the final registration between the input 3D face image and the template image. The template image is chosen to be the first image in the database and so all images are registered to this. Before performing ICP the mean of the input data is removed, to ensure there are minimal translations between the 3D face and the template image.

The normalisation processes yields raw 3D face data, however, there are often different regions of the face that are considered to be erroneous. For example in Figure 6 (a) most of the face region is retained whereas in Figure 6 (b) the chin and part of the cheek region has been removed. To overcome this issue the upper face region is extracted and the missing points are interpolated, the upper part of the face is chosen as previous work by these authors [8] showed that the upper face contains the most information. The final face region is interpolated onto a common grid, using the template image, and consists of 108×108 pixels, as shown in Figure 7.

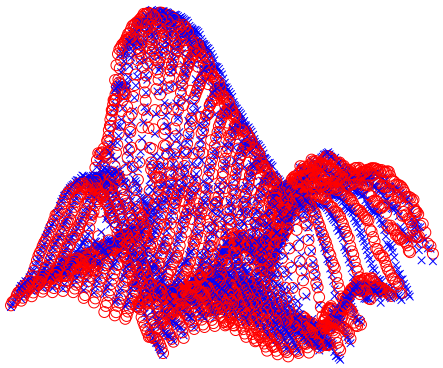
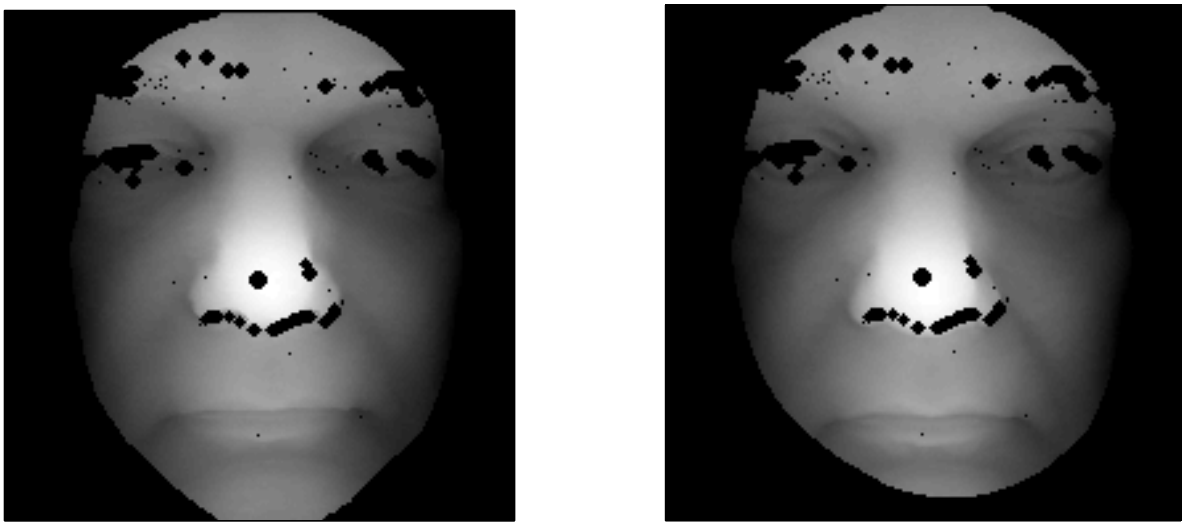


Fig. 7. An example output from the normalisation and registration procedure, the two face images are of the same ID.

IV. Analysis and Results

To analyse the performance of the proposed normlisation procedure a baseline normalisation method was



(a) (b)

Fig. 6. In (a) a normalised face image is shown where most of the face surface has been retained, whereas in (b) the chin and parts of the cheek region have been removed.

chosen. The basline method chosen is the normalisation process used to obtain baseline results for the FRGC, which is very similar to the normalisation process proposed in [3]. This method process produces full face images of size 150×130 and so a cropped region, of size 108×108 pixels, was extracted so that there could be a direct comparison between the baseline method and proposed method, an example of the full face and cropped face images is provided in Figure 8.

The normalised images should well represent each 3D face surface. Therefore the 3D surface of the same ID should be close to one another whereas the 3D surface for different IDs should lie further away. To measure this performance the Manhattan distance,

$$d_{L1} = \sum_{x=1}^{M} \sum_{y=1}^{N} |F_1(x,y) - F_2(x,y)|, \qquad (5)$$

between each 3D face surface ($F(x,y)$) is calculated; each face surface is of size $N = M = 108$. Using these distances a Detection Error Tradeoff (DET) plot was generated, DET plots are often used to examine verification performance [9].

The DET plot for the baseline and proposed normalisation procedures, Figure 9 show that proposed normalisation procedure outperforms the baseline system. This is a significant increase in discriminability with the Equal Error Rate (EER) improving from 21.65% for the baseline to 13.81% for the proposed method. This increase in performance was attributed to improved data normalisation and data registration, however, further analysis showed that there are several images for which the data is not registered.

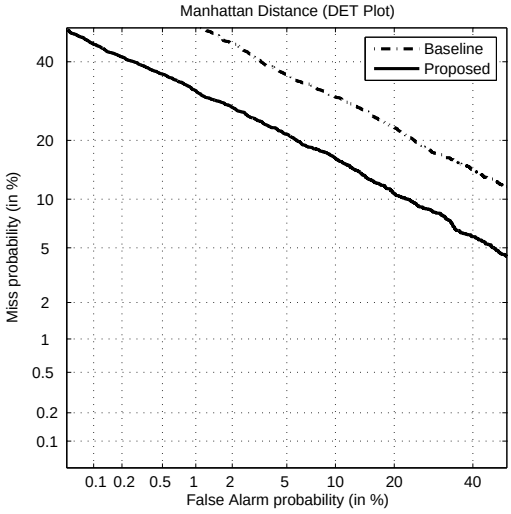
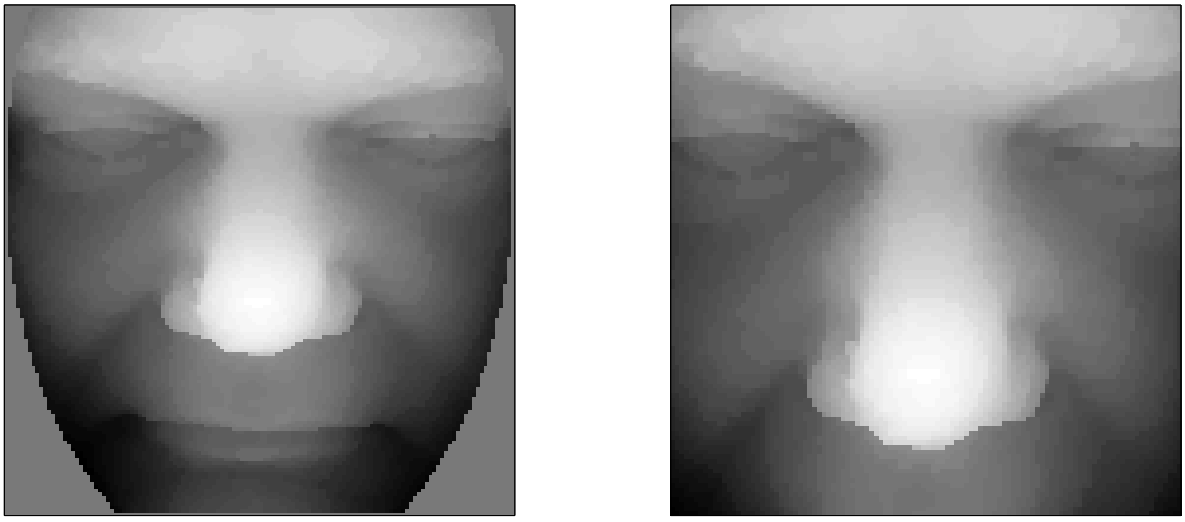


Fig. 9. A DET plot derived by using the Manahattan distance of all the data using the normalisation procedure used to derive baseline FRGC 3D results (Baseline) compared the proposed normalisation procedure (Proposed).

The ICP algorithm is used to register the 3D face



(a) (b)

Fig. 8. An example of full face image and cropped face image obtained by applying the normalisation technique used to obtain initial FRGC 3D baseline results. In (a) the full face region is supplied and in (b) the cropped face region which is used for these experiments is provided.

surface with a template image. There are several instances when the ICP algorithm results in images which are not registered, an example of inaccurate registration can be seen in Figure 10. This failure to register two images is a significant problem because in order to compare two images a consistent basis of comparison is required. One method for overcoming these registration errors is to have a template image for each ID and so perform a client specific registration.

V. Conclusions and Future Work

The normalisation procedure outlined in this work provides more accurate 3D face surfaces than the normalisation procedure used for the FRGC experiments. The proposed normalisation procedure uses the statistics of the available 3D face surface to identify invalid points and performs registrations to a template image. This means that this process can be readily applied to any 3D face database to generate a 3D face surface consisting of only valid data points.

Analysis of this normalisation procedure indicates that there are several images which are not registered. This is considered to be a significant problem and as such future work would examine methods to derive client specific templates, or several template images to improve the robustness of the registration technique.

Acknowledgements

This project was supported by the Australian Government Department of the Prime Minister and Cabinet.

References

[1] P. J. Phillips, P. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and M. Bone, "Face recognition vendor test 2002: Overview and summary," *IEEE Interational Workshop on Analysis and Modeling of Faces and Gestures*, p. 44, 2003.

[2] J. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," *Proceedings of IEEE Conference of Computer Vision and Pattern Recognition*, vol. 1, pp. 947–954, 2005.

[3] K. I. Chang, K. W. Bowyer, and P. J. Flynn, "Face recognition using 2d and 3d facial data," *Workshop in Multimodal User Authentication*, pp. 25–32, 2003.

[4] K. Chang, K. Bowyer, and P. Flynn, "Multiple nose region matching for 3d face recognition under varying facial expression," *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 28, pp. 1695–1700, 2006.

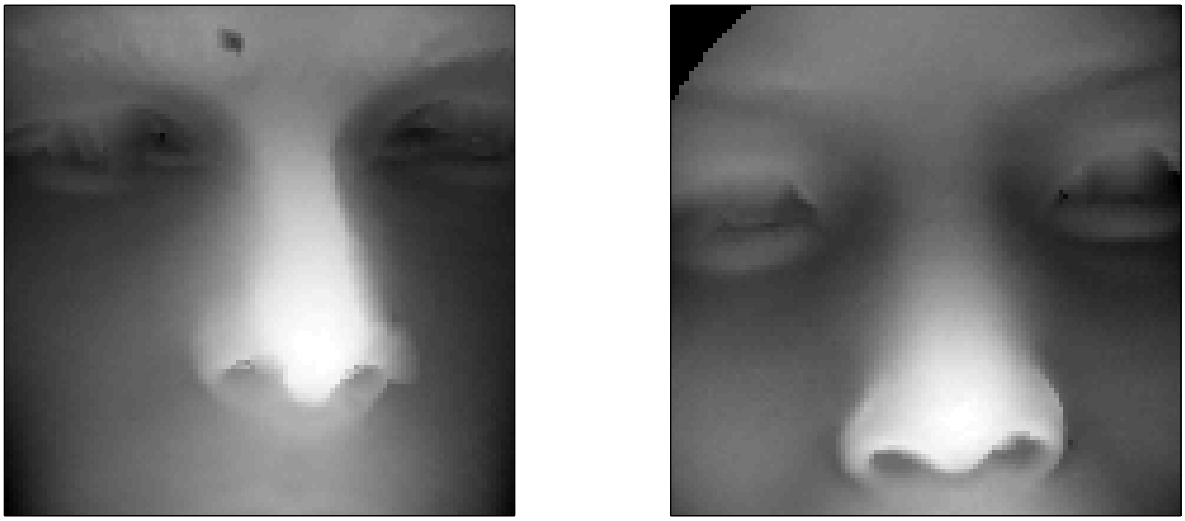
[5] Minolta, "Konica minolta vivid 910 website," *http://se.konicaminolta.us/products/3d_scanners/vivid_910/index.html*, 2006.

[6] C. McCool, V. Chandran, and S. Sridharan, "2d-3d hybrid face recognition based on pca and feature modelling," *Proceedings of the 2nd Workshop on Multimodal User Authentication*, 2006.

[7] P. Besl and H. McKay, "A method for registration of 3-d shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, pp. 239–256, 1992.

[8] G. Mamic, C. Fookes, and S. Sridharan, "Human face reconstruction using bayesian deformable models," *IEEE International Conference on Video and Signal Based Surveillance*, p. 59, 2006.

[9] A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki, "The DET curve in assessment of detection task performance," in *Eurospeech*, vol. 4, 1997, pp. 1895–1898.



(a) (b)

Fig. 10. Two images of the second ID within the FRGC database. It can be seen that there is a significant translation from the image in (a) to the image in (b).