On Discriminative Properties of TPS Warping Parameters for 3D Face Recognition

Nesli Erdogmus, Jean-Luc Dugelay
Multimedia Communications Department, EURECOM
2229, Route de Cretes, BP 193
F-06560, Sophia-Antipolis, France
{nesli.erdogmus, jean-luc.dugelay}@eurecom.fr

***Abstract*—** Due to the advances in the acquisition systems, three-dimensional (3D) facial shape information has been increasingly used for human face recognition. In order to be able to compare different facial surfaces, various registration approaches have been proposed, including Thin Plate Spline (TPS) based algorithms. In this paper, instead of adopting the TPS for registration purposes, we analyze the discriminative properties of the parameters obtained by deforming a generic face model onto target faces using TPS. The warping parameters (WP) that describe the non-global and non-linear transformations and represent the deviations from the common geometric structure are given to the classifier for face recognition. The descriptiveness of those vectors is analyzed on the FRGC database where total of 4569 3D face models are utilized. In spite of its low complexity compared to other proposed approaches, this method yields promising accuracy rates.

I. INTRODUCTION

On account of the recent advances in 3D acquisition tools, the decrease in their prices and the higher computational power of computers, 3D facial shape information has become a practical modality in face recognition systems. For most of the 3D face recognition methods which use local similarity metrics, the success rates are highly dependent on the registration step, where dense correspondences are established among faces, in order to be able to make a comparison.

The Iterative Closest Point (ICP) is the most frequently employed registration algorithm [1]-[4] for the alignment of the 3D facial surfaces. By regarding them as free-form curved surfaces, ICP minimizes the distance between two point clouds while preserving the rigidity. However, initialization before starting the iterations is a crucial problem, since the ICP algorithm converges to a local minimum monotonically.

A surface matching framework that takes into account both rigid and non-rigid variations is proposed in [5], where ICP is extended by applying a Thin Plate Spline (TPS) warping algorithm, which can establish registration in non-rigidly deformed surface patches. After originally proposed in [6] by Bookstein, this elegant mathematical framework has been extensively utilized to define the deformation of 3D surfaces [7]-[9].

In [3], the aforementioned two registration techniques are compared through 3D face recognition via some of the most commonly used features such as point clouds, facial profiles, curvature-based features, depth maps and surface normals. The experiment results indicate that warping the faces in 3D causes some losses in the discriminative shape information. The deformation of the probe models, even though it is minimal, is proven to be disadvantageous to the recognition performance.

As more and more control points are used for the TPS warping, the amount of deformation increases and the face becomes more and more similar to the target surface.

By exploiting this fact, in this paper we propose to utilize warping parameters as a biometric signature for 3D face recognition. Instead of minimally deforming the faces to a common target, a common (i.e. generic) face is strongly deformed to fit target faces. In this case, there exists a tradeoff between the number of control points (and hence the accurate approximation of the facial surface) and the alignment between faces. If all points on the generic model are used as landmarks, it completely deforms into the face. However, in this case the registration data is lost.

A. Related Work

In [10], Ju et al., proposes to fit a generic human animation model to the 3D images of real people in order to achieve highly realistic human animation models. Correspondences are established for segmented surface conformations in two major steps: global mapping and local deformation. Based on this work, Mao et al. [11] later develop a similar technique to construct dense correspondences for 3D facial analysis in which a generic model is mapped onto a 3D surface of a face model as opposed to forming correspondences between the raw 3D scans directly. As is the case in [10], global mapping is achieved by Radial Basis Functions (RBF) based on manually selected landmark pairs.

On the other hand, for the local deformation, as an extension to the previous work, a similarity measure is proposed for finding the most similar point on the raw scanned data for each vertex of the generic model. The similarity measure is calculated by using the surface distance, normal and curvature. Afterward, the generic model is deformed based on the selected reliable point pairs. The test results are reported to give better results than [10].

However, in a more recent work [12] that presents a large scale evaluation of the face registration method proposed in [11], it is shown that the curvature shape index does not help to find the matching of compatible features, but the matching of compatible shapes, and hence degrades the quality of the fit. It is also stated that, given the variability of the face features, it is difficult to discriminate outliers from a large face variation. For this reason, better results are achieved by removing the reliable point selection from the algorithm.

B. Proposed System

In this paper, we propose a system which employs parameters that are extracted from a mesh warping algorithm based on the findings in [12], as biometric signatures. A generic head model, with holes for the eyes and an open mouth, (shown in Figure 1) is strongly deformed to fit the facial models in the database, using the TPS method. The approach is similar to but much simpler than the work presented in [16], in the sense that an annotated model of the human face is fitted to each individual data. However, our aim is to analyze the discriminative properties of the warping parameters obtained during fitting (which have mostly been overlooked as superfluous sediment in registration applications), rather than utilizing the output of the fitting process.

When a generic model is deformed to fit an individual face in the database, an approximate representation of the facial surface is obtained. If we think of the generic model as the common geometric structure of all faces, these parameters represent the deviations from the conventional. Therefore, they may be claimed to possess concentrated information about the facial shape.

In the next section, the proposed feature extraction method based on TPS warping is explained in detail. After that, an extensive analysis on the descriptiveness of the TPS parameters is presented with the experimental setup and the achieved results.

II. Feature Extraction

The generic model is a symmetrical head model which does not give an impression of any gender. Based on the manually labeled landmarks provided by Szepycki et al. [14] on both target faces and the generic model (Figure 1), firstly an alignment is applied. Then the local deformations are replicated using the TPS method.

A. Thin-Plate Splines

In this section, a short explanation on the mathematical background of TPS is given. As the name indicates, this method is based on a physical analogy to how a thin sheet of metal bends under a force exerted on the constraint points. TPS method was made popular by Fred L. Bookstein in 1989 in the context of biomedical image analysis [6].

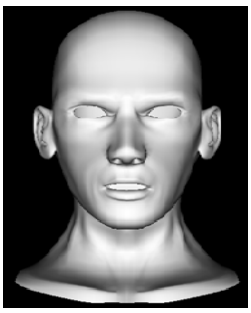
For the 3D surfaces S and T, and a set of corresponding points on each surface, P_i and M_i respectively, the TPS algorithm computes an interpolation function f(x,y) to compute T', which approximates T by warping S:

$$T' = \left\{ \begin{matrix} (x', y', z') \mid \forall (x, y, z) \\ x' = x, y' = y, z' = z + f(x, y) \end{matrix} \right\} \quad (1)$$

$$f(x, y) = a_1 + a_x x + a_y y + \sum w_i U(|P_i - (x, y)|) \quad (2)$$

with U(.), the kernel function, expressed as:

$$U(r) = r^2 ln r^2, \; r = \sqrt{x^2 + y^2} \quad (3)$$



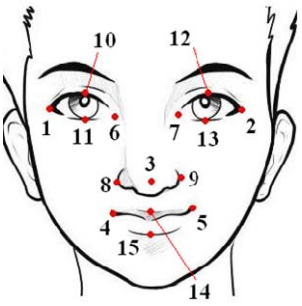


Figure 1: The generic head model and the manually annotated landmarks

Given in (2), the interpolation function f(x,y) includes the warping coefficients: w_i, $i \in \{1,2,...n\}$ to be utilized. It consists of two distinct parts. The affine part ($a_1+a_xx+a_vy$) which accounts for the affine transformation necessary for the surface to match the constraint points and a warping part ($\Sigma w_i U(|P_i-(x,y)|)$).

B. Preprocessing

The purpose of the preprocessing is to crop the face region, to eliminate the spikes and the holes and to smooth the facial surface.

In order to crop the face region, robust nose tip detection is applied based on the method suggested in [17] and the facial surface is cropped by a sphere with radius 80mm, centered 10mm away from the nose tip in +z direction.

Subsequently, spikes are removed by thresholding and a hole filling procedure is applied. Finally, a bilateral smoothing filter is employed to remove white noise while preserving the edges.

C. Rescaling and Alignment

Before warping the generic model, a linear transformation is computed in a least square sense, based on two sets of landmarks [19]. The best fit mapping is calculated by minimizing the squared distance (LSS) between the point sets. Then, the obtained transformation that includes rotation, translation and isotropic scaling is applied onto the generic model, aligning it with the subject's face.

Due to the difficulties and accuracy problems in automatic extraction of facial interest points, less number of landmarks to be used is always more desirable. An in-depth analysis is presented in the experiments section, demonstrating the trade-off between number of points used and the recognition rate. An additional ICP alignment becomes necessary if the number of points is too few. It is shown that ICP stabilizes the recognition rate independent of the number of points used.

D. Warping

After the generic model is rescaled and aligned, it is ready to be warped to compensate for the remaining local non-rigid transformations.

In the related works [10, 11, 12] on which this study is based, establishing the dense correspondence is achieved in two stages. In the first stage, called global mapping, two sets of landmarks that consist of small number of points are perfectly aligned using TPS

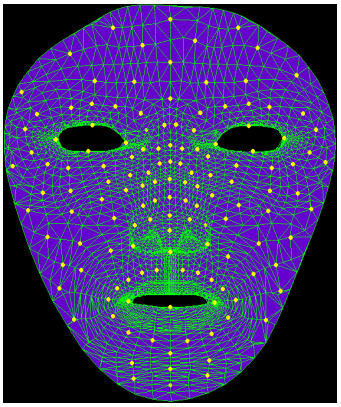


Figure 3: 140 arbitrary points regularly marked on the generic model

interpolation, namely the generic model is warped coarsely. In the second stage, for each vertex on the generic face, the most "similar" vertex on the target scan is found within a search radius of 10mm. However, different from these cited works, our aim is to represent facial surfaces with the generic model and a set of parameters. Hence, a two-step warping, in which a different intermediate state is achieved for each facial surface after global warping stage, is not suitable for our case. Warping results after the first step will not be correlated and hence will not be comparable. For this reason, we proceed with a single-step TPS warping.

Similarity between vertices on generic (G) and test (T) models is determined by the weighted sum of different metrics (4), where D is the distance between them, N is the angle between surface normals at those points and C is the difference between curvature shape indices.

$$S = -\alpha D - \beta\left(\frac{\arccos(N_G \cdot N_T)}{\pi}\right) - \gamma(C_G - C_T) \quad (4)$$

The vertex pairs that are more "similar" than a threshold are used in the energy minimization process that warps the generic model to the target face.

However, after being evaluated on the FRGC database, the curvature shape index is proven to have a detrimental effect on the results [12]. Taking into account also that the surface normals on the relatively noisy part of the face like eyes and nose are less reliable, we return back to the initially proposed matching criteria, which was based solely on distance (α=1, β=0, γ=0).

Additionally, in [12], the similarity (in our case distance) limitation is also removed from the algorithm since better fits can be obtained without it.

To construct point pairs, 140 points are manually selected on the generic face (Figure 2) and they are coupled with the closest vertices in the target surface after alignment. Using those point pairs, TPS warping is applied to the generic head model. The function f(x,y) given in (2) performs interpolation in z direction. Since the generic face is rescaled and aligned beforehand, the affine part is ignored and the weights w_i are taken into account. When we transpose the formula for the other two directions, as a result we obtain the warping vector: $[(w_{1x},w_{1y},w_{1z}), (w_{2x},w_{2y},w_{2z})\ldots (w_{140x},w_{140y},w_{140z})]$. It is a dense representation of facial surfaces, since the bulk of the data is stored in the generic model.

In this study, we investigate the performance of these warping parameters as a feature vector to be used in a biometric system.

E. Distance Metrics

Feature vectors for each face are structured as a 140x3 matrix, in which for each control point there is a 1x3 warping vector. In order to measure the distance between facial surfaces, the cosine of the angle and the Euclidean distance between the two warping vectors are calculated. This results in two 140x1 distance vectors for the compared face pair.

Two distance vectors are calculated for each face pair in the gallery. By measuring the central tendency of these vectors, two distance values are assigned. In order to avoid the sensitivity to the outliers, (e.g. mean value can be easily corrupted with a small number of extreme values.) trimmed mean approach is adopted during the experiments. Additionally, for each control point, N closest persons are assessed and for each person in the gallery, a weight is computed based on its occurrence count in those possible solution sets. The two distance values are fused by addition after being normalized to [0, 1] range.

The proposed feature extraction scheme is presented and the warping process is illustrated on a sample, in

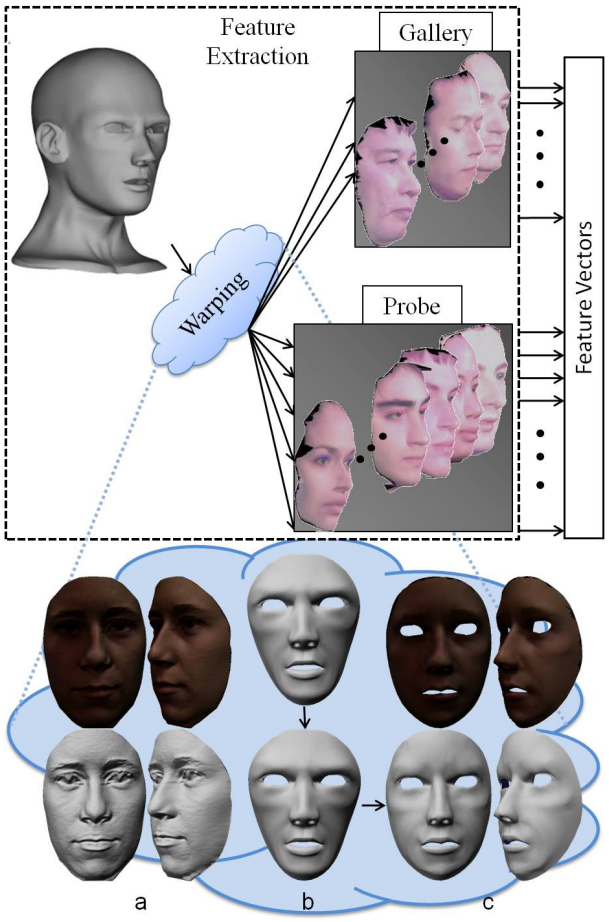


Figure 2: The proposed feature extraction scheme and an illustration on a sample model: (a) The target model with and without texture (b) generic model before and after alignment (c) generic model after warping with and without texture

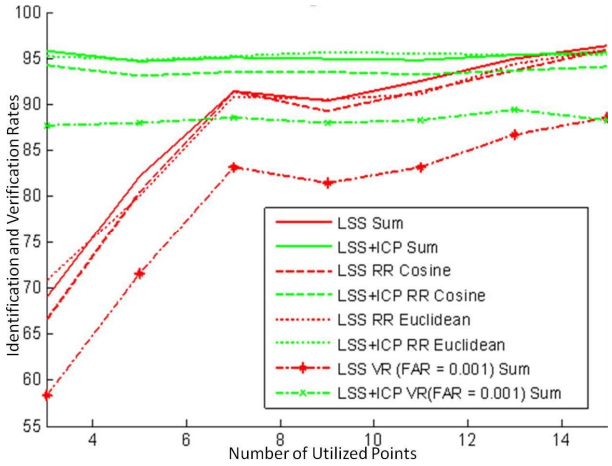


Figure 5: The identification and verification rates with increasing number of landmarks used for alignment

Figure 3. Additionally, the frontal and profile 2D/3D images of the original sample face and the generic model at different stages of the algorithm are given for comparative visualization.

III. Experiments and Analysis

In order to test the discriminative properties of the extracted feature vectors, we worked on both versions of the FRGC database [15] where the first version is used as the training partition to analyze different techniques and to obtain optimal parameters.

In FRGCv1, the persons with at least 2 images constitute a gallery of 198 classes with a single scan per person. The remaining 668 face scans are appointed to the probe set. The most important property of this version of FRGC is that all facial scans are with a neutral appearance.

For FRGCv2, in which facial models can be either with or without an expression, the gallery is composed of persons with multiple scans in the database with at least one of them being neutral. For 401 subjects that fulfill this condition, there are 2164 neutral and 1359 non-neutral test scans in the probe set.

At the beginning, the face scans in the whole database are preprocessed as explained in Section II-B. After that, the generic model is aligned to each face and finally, TPS warping is applied, resulting in feature vectors of size 140x3 for all face models.

The evaluation is mainly based on three performance metrics: rank-1 identification rate (IR), verification rate with false acceptance rate (FAR) at 0.1% (VR) and equal error rate (EER).

A. Experiments with FRGCv1

When comparing two faces, based on 140 control points, the cosine of the angle and the Euclidean distance between two warping vectors are calculated. In order to obtain a single distance metric from all control points, trimmed mean approach is adopted, where a portion of the values on both extremities are ignored.

The recognition tests are conducted both with and without an additional ICP alignment and the results for Euclidean (TM_{euc}) and Cosine (TM_{cos}) distances and their sum are given in Figure 4. With the additional ICP alignment, the success rates become independent of the number of landmarks and in fact, the best result (95.81%) is achieved using only 3 points (**m2**). On the other hand, without ICP, the recognition rates increase with more points. The best rate is 96.41% with 15 points (**m1**).

These results are obtained with non-normalized distances and without using weights (C). With the addition of weights and normalization (5), mentioned in Section II-E, the accuracy ameliorates. Details are given in Table 1.

$$D(i) = \frac{TM_{cos}(i)}{C(i)} + \frac{TM_{euc}(i)}{C(i)} \quad (5)$$

In a similar approach [20], after transforming a generic model to each face model, adaboosted geodesic distances are calculated additionally for face recognition. The rank-1 identification rate on FRGCv1 is reported as 95.0%. With our much less complex method, we reach 97.31%.

The experiments on FRGCv2 are realized with the parameters that optimize the rank-1 identification rate on FRGCv1. For trimmed mean method, the portion to be ignored is taken as 10% and for the calculation of the coefficients, the possible solution sets are created by rank N, where N equals to 20% of the gallery size.

Table 1: The identification and verification (at 0.001 FAR) rates and EER values for FRGCv1 by adding Cosine and Euclidean distances; with and without weighting and normalization

	IR	VR	EER
m1	96.41%	88.62%	0.021
m2	95.81%	87.72%	0.030
m1 with weights + normalization	97.30%	99.10%	0.005
m2 with weights + normalization	97.31%	98.65%	0.003

B. Experiments with FRGCv2

After the tests and analyses on FRGCv1, the resulting system is applied to the second version. For this experiment, three categories of facial expression are defined: Neutral, small and large [18]. This approach is adopted because the real concern in recognition is more likely the amount of the shape change rather than its type. Small expressions include moderate smiles and

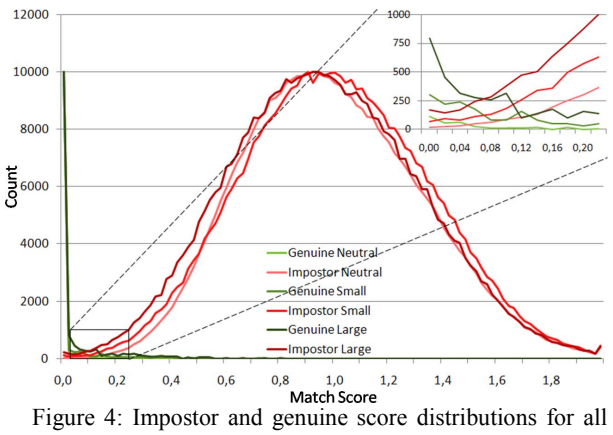


Figure 4: Impostor and genuine score distributions for all comparisons in 3 subsets: neutral, small and large

talking gestures. On the other hand, large contains unnatural expressions like blown cheeks. This classification results in a neutral probe set of size 2053 (~60%), small probe set of size 747 (~20%) and large probe set of size 723 (~20%).

In Figure 5, genuine and impostor score distributions for all-vs.-all comparison in FRGCv2 is given for 3 subsets. As the figure implies, a decline occurs in the accuracy as stronger expressions are introduced into the probe sets. This is mainly due to the control points around the regions that are sensitive to facial expressions.

In Table 2, the rates obtained for both identification and verification (at 0.001 FAR) are given with EER values. As expected, the impact of the expression variations is also visible on the success rates.

Table 2: The identification and verification (at 0.001 FAR) rates and EER values for the whole FRGCv2 database (A) and for three subsets: Neutral (N), small (S) and large (L).

	IR m1	VR m1	EER m1	IR m2	VR m2	EER m2
N	97.13%	99.37%	0.003	92.40%	95.37%	0.015
S	85.81%	91.70%	0.018	84.74%	88.22%	0.031
L	71.92%	76.35%	0.033	72.75%	75.93%	0.050
A	89.55%	94.52%	0.015	86.74%	90.92%	0.026

C. Robustness to Degradation

For the third set of experiments, a gallery of 410 persons is constructed with a single neutral face scan per individual. For the probe set, one test scan is selected randomly from the FRGCv2 database which can be with or without expression and 9 degraded versions of this probe set are created: 3 decimated versions (by 2, 4 and 8), 3 noisy versions (Gaussian noise added to z, with standard deviation of 0.2, 0.4 and 0.8mm) and 3 other versions with 1, 2 or 3 holes on the facial surface (Figure 6). The identification test is conducted on those sets to evaluate the robustness of the extracted feature vector against various degradations (using m1). The results are shown in According to the results, the extracted warping parameters can still perform well with slightly low sample rate or mild noise. But as the degradation becomes stronger, the accuracy begins to drop for both decimation and noise, whereas it does not get affected strongly by the introduction of holes on the surfaces.

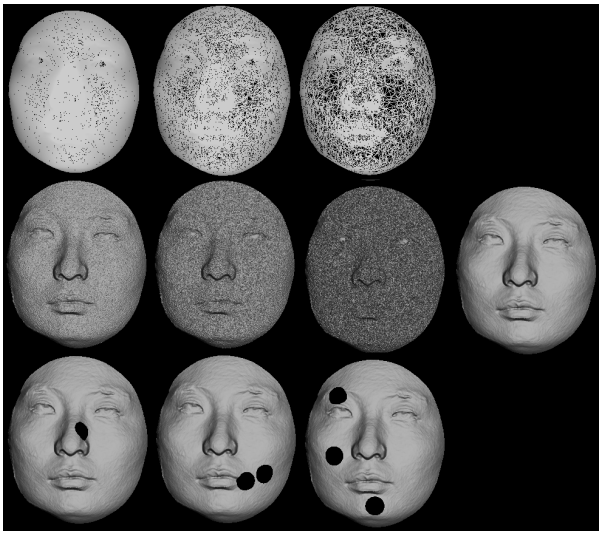


Figure 6: Degradation illustrated on a sample face scan

Table 3. The identification rate for the original probe set is obtained as 89.05%.

According to the results, the extracted warping parameters can still perform well with slightly low sample rate or mild noise. But as the degradation becomes stronger, the accuracy begins to drop for both decimation and noise, whereas it does not get affected strongly by the introduction of holes on the surfaces.

Table 3: Rank-1 identification rates for different probe sets with various degradations

Type	Rate	Type	Rate	Type	Rate
Decimation 2	86.86%	**Noise 0.2**	84.43%	**1 hole**	88.57%
Decimation 4	81.75%	**Noise 0.4**	77.13%	**2 holes**	88.08%
Decimation 6	72.02%	**Noise 0.8**	61.56%	**3 holes**	87.11%

IV. CONCLUSION

This paper presents an evaluation on the distinctive properties of warping parameters that are obtained by aligning a generic model and deforming it to face samples using TPS algorithm.

The alignment procedure is analyzed using different number of landmarks and with/without an additional ICP module. ICP is observed to stabilize the recognition rates as the number of points of interest decreases.

The evaluation is performed on FRGC database with total number of 4596 face scans. A series of tests is conducted on FRGCv1 to analyze different alignment approaches and to obtain optimal parameters for the trimmed mean method and the computation of weights.

FRGCv2 is utilized in order to evaluate the robustness of the extracted feature vectors against facial expressions and degradations; such as noise, holes and decimation. The warping parameters are shown to be robust against moderate decimation and noise and the presence of holes on the facial surface. On the other hand, facial expression variations are observed to be highly detrimental.

Table 4: Comparison with the reported identification and verification (at 0.001 FAR) rates and EER values for the whole FRGCv2 database (A) of some other approaches in all vs. all setup.

	IR	VR	EER
Maurer et al. [18]	-	86.9%	0.021
Berretti et al. [21]	85.8%	94.2%	-
Faltemier et al. [22]	97.2%	93.2%	0.032
Cook et al. [23]	92.9%	92.3%	-
Mian et al. [24]	96.2%	86.6%	-
Our method	89.6%	94.5%	0.015

In Table 4, the results for all vs. all experiments from the recent literature are listed. Considering the fact that extraction of warping parameters is relatively very simple, the obtained result is a harbinger of a very promising 3D face recognition method.

For future work, we would like to adopt a region-based approach using the warping parameters and we plan to explore the quality of the facial surface patches to obtain a reliability score for one point or groups of control points, aiming robustness against expression variations.

ACKNOWLEDGMENT

This work was supported by the project FAR 3D ANR-07-SESU-04.

REFERENCES

[1] G. Pan and Z. Wu, "3D Face Recognition from Range Data," *Int. Journal of Image and Graphics*, 5(3):573-593, 2005.

[2] M.O. Irfanoglu, B. Gokberk, and L. Akarun, "3D Shape-Based Face Recognition Using Automatically Registered Facial Surfaces," *Int. Conference on Pattern Recognition*, 4:183–186, 2004.

[3] B. Gökberk, M.O. Irfanoglu and L. Akarun, "3D shape-based face representation and feature extraction for face recognition," *Image and Vision Computing*, 24(8):857-869, 2006.

[4] X. Lu, D. Colbry and A. K. Jain, "Three-Dimensional Model Based Face Recognition," *Int. Conference on Pattern Recognition*, 1:362-366, 2004.

[5] X. Lu and A. K. Jain, "Deformation Analysis for 3D Face Matching," *IEEE Workshops on Application of Computer Vision*, 1:99-104, 2005.

[6] F. L. Bookstein, "Principal warps: Thin-Plate Splines and Decomposition of Deformations," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 11(6):567–585, 1989.

[7] Y. Hu, M. Zhou and Z. Wu, "An Automatic Dense Point Registration Method for 3D Face Animation," *Int. Congress on Image and Signal Processing*, 1-6, 2009.

[8] D. Schneider and P. Eisert, "Automatic and Robust Semantic Registration of 3D Head Scans," *European Conference on Visual Media Production*, 2008.

[9] Y. Hu, M. Zhou and Z. Wu, "An Automatic Non-rigid Point Matching Method for Dense 3D Face Scans," *Int. Conference on Computational Science and Its Applications*, 215-221, 2009.

[10] X. Ju and J. P. Siebert, "Individualising human animation models," *Eurographics*, 2001.

[11] Z. Mao, P. Siebert, P. Cockshott, and A. Ayoub, "Constructing dense correspondences to analyze 3d facial change," *Int. Conference on Pattern Recognition*, 3:144–148, 2004.

[12] J. R. Tena, M. Hamouz, A. Hilton and J. Illingworth, "A Validated Method for Dense Non-rigid 3D Face Registration," *IEEE Int. Conference on Video and Signal Based Surveillance*, 81-86, 2006.

[13] P. J. Besl and N.D. McKay, "A Method for Registration of 3-D Shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239-256, 1992.

[14] P. Szeptycki, M. Ardabilian and L. Chen, "A coarse-to-fine curvature analysis-based rotation invariant 3D face landmarking," *Int. Conference on Biometrics: Theory, Applications and Systems*, 1-6, 2009.

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

[16] A. Kakadiaris, G. Passalis, G. Toderici, M. N. Murtuza, Y. Lu, N. Karampatziakis and T. Theoharis, "Three-Dimensional Face Recognition in the Presence of Facial Expressions: An Annotated Deformable Model Approach," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 29(4):640-649, 2007.

[17] X. Lu and A. K. Jain, "Multimodal facial feature extraction for automatic 3D face recognition" Department of Computer Science, Michigan State University, East Lansing, Michigan, Tech. Rep. MSU-CSE-05-22, August 2005.

[18] T. Maurer, D. Guigonis, I. Maslov, B. Pesenti, A. Tsaregorodtsev, D. West, and G. Medioni, "Performance of Geometrix ActiveID^TM 3D Face Recognition Engine on the FRGC Data," *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 154-160, 2005.

[19] T. S. Huang, S. D. Blostein, and E. A. Margerum, "Least-squares estimation of motion parameters from 3-D point correspondences," *IEEE Conference on Computer Vision and Pattern Recognition*, 1986.

[20] S. Cadavid, J. Shou and M. Abdel-Mottaleb, "Determining discriminative anatomical point pairings using adaboost for 3D face recognition," *Int. Conference on Image Processing*, 49-52, 2009.

[21] S. Berretti, A. Del Bimbo, P. Pala, "3D Face Recognition Using Isogeodesic Stripes," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 32(12):2162-2177, 2010.

[22] T.C. Faltemier, K.W. Bowyer, and P.J. Flynn, "A Region Ensemble for 3D Face Recognition," *IEEE Trans. Information Forensics and Security*, 3(1): 62-73, 2008.

[23] J. Cook, V. Chandran, and C. Fookes, "3D Face Recognition Using Log-Gabor Templates," Proc. British Machine Vision Conf., pp. 83- 83, 2006.

[24] A.S. Mian, M. Bennamoun, and R. Owens, "An Efficient Multimodal 2D-3D Hybrid Approach to Automatic Face Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 29(11):1927-1943, 2007.