3D FACE RECOGNITION FOR BIOMETRIC APPLICATIONS

L. Akarun, B. Gökberk, A.A. Salah

Department of Computer Engineering
Boğaziçi University, Bebek, İstanbul, Turkey
phone: + (90) 212 3597183, fax: + (90)212 2872461
e-mail: {akarun,gokberk,salah}@boun.edu.tr

ABSTRACT

Face recognition (FR) is the preferred mode of identity recognition by humans: It is natural, robust and unintrusive. However, automatic FR techniques have failed to match up to expectations: Variations in pose, illumination and expression limit the performance of 2D FR techniques. In recent years, 3D FR has shown promise to overcome these challanges. With the availability of cheaper acquisition methods, 3D face recognition can be a way out of these problems, both as a stand-alone method, or as a supplement to 2D face recognition. We review the relevant work on 3D face recognition here, and discuss merits of different representations and recognition algorithms.

1. INTRODUCTION

Recent developments in computer technology and the call for better security applications have brought biometrics into focus. A biometric is a physical property; it cannot be forgotten or mislaid like a password, and it has the potential to identify a person in very different settings: a criminal entering an airport, an unconscious patient without documents for identification, an authorized person accessing a highly-secured system. Be it for purposes of security or human–computer interaction, there is wide application to robust biometrics.

Two different scenarios are of primary importance. In the verification (authentication) scenario, the person claims to be someone, and this claim is verified by ensuring the provided biometric is sufficiently close to the data stored for that person. In the more difficult recognition scenario, the person is searched in a database. The database can be small (e.g. criminals on the wanted list) or large (e.g. photos on registered ID cards). The unobtrusive search for a number of people is called *screening*.

The signature and handwriting have been the oldest biometrics, used in the verification of authentication of documents. Face image and the fingerprint also have a long history, and are still kept by police departments all over the world. More recently, voice, gait, retina and iris scans, hand print, and 3D face information are considered for biometrics. Each of these have different merits, and applicability. When deploying a biometrics based system, we consider its accuracy, cost, ease of use, ease of development, whether it allows integration with other systems, and the ethical consequences of its use. Two other criteria are susceptibility to spoofing (faking an identity) in a verification setting, and susceptibility to evasion (hiding an identity) in a recognition setting.

The purpose of the present study is to discuss the merits and drawbacks of 3D face information as a biometric, and review the state of the art in 3D face recognition. Two things make face recognition especially attractive for our consideration. The acquisition of the face information is easy and non-intrusive, as opposed to iris and retina scans. This is important if the system is going to be used frequently, and by a large number of users. The second point is the relatively low privacy of the information; we expose our faces constantly, and if the stored information is compromised, it does not lend itself to improper use like signatures and fingerprints would. The drawbacks of 3D face recognition include high cost and decreased ease-of-use for laser sensors, low accuracy for other acquisiton types, and the lack of sufficiently powerful algorithms. Figure. 1 presents a summary of different biometrics and their relative strengths.

	2D Face	Signature	Voice	Iris	Finger	Retina	Gait	Hand	3D Face
Accuracy				X	X	X			X
Cost	X		X		X				
Privacy	X			X		X	X		X
Integrity	X		X		X				X
Ease of Use	X	X	X		X			X	X
Development	X	X	X		X			X	X
Hiding identity				X		X		X	
Faking identity				X	X	X		X	X

Figure 1: Biometrics and their relative strengths. Although 2D and 3D face recognition are not as accurate as iris scans, their ease of use and lower cost makes them a preferable choice for some scenarios.

3D face recognition represents an improvement over 2D face recognition in some respects. Recognition of faces from still images is a difficult problem, because the illumination, pose and expression changes in the images create great statistical differences and the identity of the face itself becomes shadowed by these factors. Humans are very capable in this modality, precisely because they learn to deal with these variations. 3D face recognition has the potential to overcome feature localization, pose and illumination problems, and it can be used in conjunction with 2D systems.

In the next section we review the current research on 3D face recognition. We focus on different representations of 3D information, and the fusion of different sources of information. We conclude by a discussion of the future of 3D face recognition.

2. STATE OF THE ART IN 3D FACE RECOGNITION

2.1 3D Acquisition and Preprocessing

We distinguish between a number of range data acquisition techniques. In the *stereo acquisition* technique, two or more cameras that are positioned and calibrated are employed to acquire simultaneous snapshots of the subject. The depth information for each point can be computed from geometrical models and by solving a correspondence problem. This method has the lowest cost and highest ease of use. The *structural light* technique involves a light pattern projected on the face, where the distortion of the pattern reveals depth information. This setup is relatively fast, cheap, and allows a single standard camera to produce 3D and texture information. The last technique employs a *laser sensor*, which is typically more accurate, but also more expensive and slower to use. The acquisition of a single 3D head scan can take more than 30 seconds, a restricting factor for the deployment of laser-based systems.

3D information needs to be preprocessed after acquisition. Depending on the type of sensor, there might be holes and spikes (artifacts) in the range data. Eyes and hair will not reflect the light appropriately, and the structured light approaches will have trouble correctly registering those portions. Illumination still effects the 3D acquisition, unless accurate laser scanners are employed [7].

For patching the holes, missing points can be filled by interpolation or by looking at the other side of the face [22, 33]. Gaussian smoothing and linear interpolation are used for both texture and range information [1, 8, 10, 13, 15, 22, 24, 30]. Clutter is usually removed manually [6, 8, 15, 24, 18, 21, 29, 30] and sometimes parts of the data are completely omitted where the acquisition leads to noise levels that cannot be coped with algorithmically [10, 21, 35]. To help distance calculation, the mesh representations can be regularized [16, 36], or a voxel discretization can be used [2].

Most of the algorithms start by aligning the faces, either by their centres of mass [8, 29], nose tip [15, 18, 19, 22, 26, 30], the eyes [13, 17], or by fitting a plane to the face and aligning it with that of the camera [2]. Registration of the images is important for all local similarity metrics. The key idea in registration is to define the similarity metric and the set of possible transformations. The similarity is measured by point-to-point or point-to-surface distances, or cross correlation between more complex features.

The rigid transformation of a 3D object involves a 3D rotation and translation, but the nonlinearity of the problem calls for iterative methods [11]. The most frequently used ([16, 19, 21, 22, 27, 29]) registration technique is the Iterative Closest Point (ICP) algorithm [3]. Warping and deforming the models (non–rigid registration) for better alignment helps co-locating the landmarks. An important method is the Thin Plate Spline (TPS) algorithm, which establishes perfect correspondence [16, 20]. One should however keep in mind that the deformation may be detrimental to the recognition performance, as discriminatory information is lost proportional to the number of anchor points. Lu and Jain also distinguish between inter-subject and intra-subject deformations, which is found useful for classification [20].

Landmark locations used in registration are either found manually [6, 10, 17, 19, 21, 25, 33] or automatically [12, 16, 37]. The correct localization of the landmarks is crucial to many algorithms, and it is usually not possible to judge the sensitivity of an algorithm to localization errors from its description. Nevertheless, the automatic landmark localization remains an unsolved problem.

2.2 3D Recognition Algorithms

We summarize relevant work in 3D face recognition. We have classified each work according to the primary representation used in the recognition algorithm, much in the spirit of [7]. Table 3 summarizes the recent work on 3D and 2D+3D face recognition.

2.2.1 Curvatures and Surface Features

In one of the early 3D face papers, Gordon proposed a curvature-based method for face recognition from 3D data, kept in a cylindrical coordinate system [13]. Since the curvatures involve second derivatives, they are very sensitive to noise. An adaptive Gaussian smoothing is applied so as not to destroy curvature information. In [31] principal directions of curvatures are used. The advantage of these over surface normals is that they are applicable to free-form surfaces. Moreno et al. extracted a number of features from 3D data, and found that curvature and line features perform better than area features [24]. In [14], the authors have compared different representations on the 3D RMA dataset: point clouds, surface normals, shape-index values, depth images, and facial profile sets. Surface normals are reported to be more discriminative than others, and LDA is found very useful in extracting discriminative features.

2.2.2 Point Clouds and Meshes

Point cloud is the most primitive 3D representation for faces, and it is difficult to work with. Achermann and Bunke employ Hausdorff distance for matching the point clouds [2]. They use a voxel discretization to speed up matching, but it causes some information loss. Lao et al. discard matched points with large distances as noise [17].

When the data are in point cloud representation, ICP is the most widely used registration technique. The similarity of two point sets that is calculated at each iteration of the ICP algorithm is frequently used in point cloud-based face recognizers. Medioni and Waupotitsch present an authentication system that acquires the 3D image of the subject with two calibrated cameras [23] and ICP algorithm is used to define similarity between two face meshes. Lu et al. use a hybrid-ICP based registration using Besl's method and Chen's method successively [19]. The base mesh is also used for alignment in [36], where features are extracted from around landmark points, and nearest neighbour after PCA is used for recognition. Lu and Jain also use ICP for rigid deformations, but they also propose to use TPS for intra-subject and inter-subject nonrigid deformations, with the purpose of handling expression variations [20]. Deformation analysis and combination with appearance based classifiers both increase the recognition accuracy.

In a similar study, İrfanoğlu et al. have used ICP to automatically locate facial landmarks in a coarse alignment step, and then warp faces using TPS algorithm to establish dense point-to-point correspondences [16]. The use of an average face model significantly reduces the complexity of similarity calculation and point-cloud representation of registered faces are more suitable for recognition then depth image-based methods, point signatures, and implicit polynomial-based representation techniques. In a follow-up study, Gökberk et al. have analyzed the effect of registration methods on the classification accuracy [14]. To inspect side effects of warping on discrimination an ICP-based approximate dense registration algorithm is designed that allows only rotation and translation transformations. Experimental results confirmed that ICP without warping leads to better recognition accuracy[1]. Table. 1 summarizes the classification accuracies of different feature extractors for both TPS-based and ICP-based registration algorithms on the 3D RMA dataset. Improvement is visible for all feature extraction methods, except the shape-index.

Table 1: Average classification accuracies (and standard deviations) of different face recognizers for 1) TPS warping-based and 2) ICP-based face representation techniques.

	TPS	ICP
Point Cloud	92.95 ± 1.01	96.48 ± 2.02
Surface Normals	97.72 ± 0.46	99.17 ± 0.87
Shape Index	90.26 ± 2.21	88.91 ± 1.07
Depth PCA	45.39 ± 2.15	50.78 ± 1.10
Depth LDA	75.03 ± 2.87	96.27 ± 0.93
Central Profile	60.48 ± 3.78	82.49 ± 1.34
Profile Set	81.14 ± 2.09	94.30 ± 1.55

2.2.3 Depth Map

Depth maps are usually used in conjunction with subspace methods, although most of the existing 2D techniques are suitable for processing the depth maps. The depth map construction consists of selecting a viewpoint, and smoothing the sampled depth values. In [15], PCA and ICA were compared on the depth maps. ICA was found to perform better, but PCA degraded more gracefully with declining numbers of training samples. In Srivastava et al. the set of all k-dimensional subspaces of the data space is searched with a MCMC simulated annealing algorithm for the optimal linear subspace [30]. The optimal subspace method performs better than PCA, LDA or ICA. Achermann at al. compare an eigenface method with a 5-state left-right HMM on a database of depth maps [1]. They show that the eigenface method outperforms the HMM, and

[1] In [32] texture was found to be more informative than depth; our findings point out to warping as a possible reason.

the smoothing effects the eigenface method positively, while its effect on the HMM is detrimental.

The 3D data are usually more suitable for alignment, and should be preferred if available. In Lee et al. the 3D image is thresholded after alignment to obtain the depth map, and a number of small windows are sampled from around the nose [18]. The statistical features extracted from these windows are used in recognition.

2.2.4 *Profile*

The most important problem for the profile-based schemes is the extraction of the profile. In an early paper Cartoux et al. use an iterative scheme to find the symmetry plane that cuts the face into two similar parts [9]. The nose tip and a second point are used to extract the profiles. Nagamine et al. use various heuristics to find feature points and align the faces by looking at the symmetry [25]. Then the faces are intersected with different kinds of planes (vertical, horizontal or cylindrical around the nose tip), and the intersection curve is used in recognition. Vertical planes around $\pm$20mm. of the central region and selecting a cylinder with $20-30$mm. radius around the nose (crossing the inner corners of the eyes) produced the best results. In [4], Beumier and Acheroy detail the acquisition of the popular 3D RMA dataset with structural light and report profile based recognition results. In addition to the central profile, they use the average of two lateral profiles in recognition.

Once the profiles are obtained, there are several ways of matching them. In [9], corresponding points of two profiles are selected to maximize a matching coefficient that uses the curvature on the profile curve. Then a correlation coefficient and the mean quadratic distance is calculated between the coordinates of the aligned profile curves, as two alternative measures. In [4], the area between the profile curves is used. In [14] distances calculated with L_1 norm, L_2 norm, and generalized Hausdorff distance were compared for aligned profiles, and the L_1 norm is found to perform better.

2.2.5 *Analysis by Synthesis*

In [6] the analysis-by-synthesis approach that uses morphable models is detailed. A morphable model is defined as a convex combination of shape and texture vectors of a number of samples that are placed in dense correspondence. A single 3D model face is used to render an image similar to the test image, which leads to the estimation of viewpoint parameters (pose angles, 3D translation, focal length of the camera), illumination parameters (ambient and directed light intensities, direction angles of the light, colour contrast, gains and offsets of the colour channels), and deformation parameters (shape and texture). In [22] a system is proposed to work with 2D colour images and corresponding 3D depth maps. The idea is to synthesize a pose and illumination corrected image pair for recognition. The depth images performed significantly better (by 4-7 per cent) than colour images, and the combination increased the accuracy as well (by 1-2 per cent). Pose correction is found to be more important than illumination correction.

2.2.6 *Combinations of Representations*

Most of the work that uses 3D face data use a combination of representations. The enriched variety of features, when combined with classifiers with different statistical properties, produce more accurate and more robust results. In Tsutsumi et al. surface normals and intensities are concatenated to form a single feature vector, and the dimensionality is reduced with PCA [34]. In [35], the 3D data are described by point signatures, and the 2D data by Gabor wavelet responses, respectively. 3D intensities and texture were combined to form the 4D representation in [29]. Bronstein et al. point out to the non-rigid nature of the face, and to the necessity of using a suitable similarity metric that takes this deformability into account [8]. For this purpose, they use multi-dimensional scaling projection algorithm for both shape and texture information.

Apart from techniques that fuse the representations at the feature level, there are a number of systems that employ combination at the decision level. Chang et al. propose in [10] to use Mahalanobis distance-based nearest-neighbor classifiers on the 2D intensity and 3D range images separately, and fuse the decisions with a rank-based approach at the decision level. In [32] the depth map and colour maps (one for each YUV channel) are projected via PCA and the distances in four subspaces are combined by multiplication. In [33] the depth map and the intensity image are processed with embedded HMMs separately, and weighted score summation is proposed for the combination. In [21], Lu and Jain combine texture (LDA) and surface (point-to-plane distance) with weighted voting, but only the difficult samples are classified via the combined system.

Profiles are also used in conjunction with other features. In [5], 3D central and lateral profiles, gray level central and lateral profiles were evaluated separately, and then fused with Fisher's method. In [26] a surface-based recognizer and a profile-based recognizer are combined at the decision level. Surface-matcher's similarity is based on a point cloud distance approach, and profile similarity is calculated using Hausdorff distance. In [27], a number of methods are tested on the depth map (Eigenface, Fisherface, and kernel Fisherface), and the depth map expert is fused with three profile experts with Max, Min, Sum, Product, Median and Majority Vote rules, out of which the Sum rule was selected.

Gökberk et al. have proposed two combination schemes that use 3D facial shape information [14]. In the first scheme, called *parallel fusion*, different pattern classifiers are trained using different features such as point clouds, surface normals, facial profiles, and PCA/LDA of depth images. The outputs of these pattern classifiers are merged using a rank-based decision level fusion algorithm. As combination rules, consensus voting, a non-linear variation of a rank-sum method, and a highest rank majority method are used. Table. 2 shows the recognition accuracies of individual pattern recognizers together with the accuracies of the parallel ensemble methods for the 3D RMA dataset. It is seen that while the best individual pattern classifier (Depth-LDA) can accurately classify 96.27 per cent of the test examples, a non-linear rank-sum fusion of Depth-LDA, surface normals, and point cloud classifiers improves the accuracy to 99.07 per cent. Paired t-test results indicate that all of the accuracies of the parallel fusion schemes are statistically better than individual classifier's performances. The second scheme is called *serial fusion* where the class outputs of a filtering first classifier is passed to a second more complex classifier. The ranked output lists of these classifiers are fused. The first classifier in the pipeline should be fast and accurate. Therefore a point cloud-based pattern classifier was selected. As the second classifier, Depth-LDA was chosen because of its discriminatory power. This system has 98.14 per cent recognition accuracy, significantly better than the single best classifier.

Table 2: Classification accuracies of single face classifiers (top part), and the combined classifiers (bottom part).

Performances of Pattern Classifiers		
	Dimensionality	Acc.
Point Cloud	$3,389\times 3$	95.96
Surface Normals	$3,389\times 3$	95.54
Depth PCA	300	50.78
Depth LDA	30	96.27
Profile Set	$1,557$	94.30

Performances of Combined Classifiers		
	Pattern Classifiers	Acc.
Consensus Voting	LDA, PC, SN	98.76
Nonlinear Rank-Sum	Profile, LDA, SN	99.07
Highest Rank Majority	Profile, LDA, SN, PC	98.13
Serial Fusion	PC, LDA	98.14

3. CONCLUSIONS

There are a number of questions 3D face recognition research needs to address. In acquisition, the accuracy of cheaper and less intrusive systems needs to be improved, temporal sequences should be considered. For registration, automatic landmark localization, artifact removal, scaling, and elimination of errors due to occlusions, glasses, beard, etc. need to be worked out. Ways of deforming the face without losing discriminative information might be beneficial.

It is obvious that information fusion is the future of 3D face recognition. There are many ways of representing and combining texture and shape information. We also distinguish between local and configural processing, where the ideal face recognizer makes use of both. For realistic systems, single training instance cases should be considered, which is a great hurdle to some of the more successful discriminative algorithms. Publicly available 3D datasets are necessary to encourage further research on these topics.

REFERENCES

[1] Achermann, B., X. Jiang, H. Bunke, "Face recognition using range images," in *Proc. Int. Conf. on Virtual Systems and MultiMedia*, pp.129-136, 1997.

[2] Achermann, B., H. Bunke, "Classifying range images of human faces with Hausdorff distance," in *Proc. ICPR*, pp.809-813, 2000.

[3] Besl, P., N. McKay, "A Method for Registration of 3-D Shapes," *IEEE Trans. PAMI*, vol.14, no.2, pp.239-256, 1992.

[4] Beumier, C., M. Acheroy, "Automatic 3D face authentication," *Image and Vision Computing*, vol.18, no.4, pp.315-321, 2000.

[5] Beumier, C., M. Acheroy, "Face verification from 3D and grey level cues," *Pattern Recognition Letters*, vol.22, pp.1321-1329, 2001.

[6] Blanz, V., T. Vetter, "Face Recognition Based on Fitting a 3D Morphable Model," *IEEE Trans. PAMI*, vol.25, no.9, pp.1063-1074, 2003.

[7] Bowyer, K., Chang K., P. Flynn, "A survey of multi-modal 2D+3D face recognition," Technical Report, Notre Dame Department of Computer Science and Engineering, 2004.

[8] Bronstein, A.M., M.M. Bronstein, R. Kimmel, "Expression-invariant 3D face recognition," in J. Kittler, M.S. Nixon (eds.) *Audio- and Video-Based Person Authentication*, pp.62-70, 2003.

[9] Cartoux, J.Y., J.T. LaPreste, M. Richetin, "Face authentication or recognition by profile extraction from range images," in *Proc. of the Workshop on Interpretation of 3D Scenes*, pp.194-199, 1989.

[10] Chang, K., K. Bowyer, P. Flynn, "Multi-modal 2D and 3D biometrics for face recognition," in *Proc. IEEE Int. Workshop on Analysis and Modeling of Faces and Gestures*, 2003.

[11] Chen, Y., G. Medioni, "Object Modeling by Registration of Multiple Range Images," *Image and Vision Computing*, vol.10, no.3, pp.145-155, 1992.

[12] Colbry, D., X. Lu, A. Jain, G. Stockman, "3D face feature extraction for recognition," Technical Report MSU-CSE-04-39, Computer Science and Engineering, Michigan State University, 2004.

[13] Gordon, G. "Face recognition based on depth and curvature features," in *SPIE Proc.: Geometric Methods in Computer Vision*, vol.1570, pp.234-247, 1991.

[14] Gökberk, B., A.A. Salah, L. Akarun, "Rank-based Decision Fusion for 3D Shape-based Face Recognition," *submitted for publication*.

[15] Hesher, C., A. Srivastava, G. Erlebacher, "A novel technique for face recognition using range imaging," in *Proc. 7th Int. Symposium on Signal Processing and Its Applications*, pp.201-204, 2003.

[16] İrfanoğlu, M.O., B. Gökberk, L. Akarun, "3D Shape-Based Face Recognition Using Automatically Registered Facial Surfaces," in *Proc. ICPR*, vol.4, pp.183-186, 2004.

[17] Lao, S., Y. Sumi, M. Kawade, F. Tomita, "3D template matching for pose invariant face recognition using 3D facial model built with iso-luminance line based stereo vision," in *Proc. ICPR*, vol.2, pp.911-916, 2000.

[18] Lee, Y., K. Park, J. Shim, T. Yi, "3D face recognition using statistical multiple features for the local depth information," in *Proc. ICVI*, 2003.

[19] Lu, X., D. Colbry, A.K. Jain, "Three-Dimensional Model Based Face Recognition," in *Proc. ICPR*, 2004.

[20] Lu, X. A.K. Jain, "Deformation Analysis for 3D Face Matching," to appear in *Proc. IEEE WACV*, 2005.

[21] Lu, X. A.K. Jain, "Integrating Range and Texture Information for 3D Face Recognition," to appear in *Proc. IEEE WACV*, 2005.

[22] Malassiotis, S., M.G. Strintzis, "Pose And Illumination Compensation For 3D Face Recognition," in *Proc. ICIP*, 2004.

[23] Medioni, G., R. Waupotitsch, "Face recognition and modeling in 3D," *IEEE Int. Workshop on Analysis and Modeling of Faces and Gestures*, pp.232-233, 2003.

[24] Moreno, A.B., Á. Sánchez, J.F. Vélez, F.J. Díaz, "Face recognition using 3D surface-extracted descriptors," in *Proc. IMVIPC*, 2003.

[25] Nagamine, T., T. Uemura, I. Masuda, "3D facial image analysis for human identification," in *Proc. ICPR*, pp.324-327, 1992.

[26] Pan, G., Y. Wu, Z. Wu, W. Liu, "3D Face recognition by profile and surface matching," in *Proc. IJCNN*, vol.3, pp.2169-2174, 2003.

[27] Pan, G., Z. Wu, "3D Face Recognition From Range Data," submitted to *Int. Journal of Image and Graphics*, 2004.

[28] Pankanti, S., R.M. Bolle, A. Jain "Biometrics: The Future of Identification," *IEEE Computer*, pp.46–49, 2000.

[29] Papatheodorou, T., D. Rueckert, "Evaluation of automatic 4D face recognition using surface and texture registration," in *Proc. AFGR*, pp.321-326, 2004.

[30] Srivastava, A., X. Liu, C. Hesher, "Face Recognition Using Optimal Linear Components of Range Images," *submitted for publication*, 2003.

[31] Tanaka, H., M. Ikeda, H. Chiaki, "Curvature-based face surface recognition using spherical correlation," in *Proc. ICFG*, pp.372-377, 1998.

[32] Tsalakanidou, F., D. Tzovaras, M. Strinzis, "Use of depth and colour eigenfaces for face recognition," *Pattern Recognition Letters*, vol.24, pp.1427-1435, 2003.

[33] Tsalakanidou, F., S. Malassiotis, M. Strinzis, "Integration of 2D and 3D images for enhanced face authentication," in *Proc. AFGR*, pp.266-271, 2004.

[34] Tsutsumi, S., S. Kikuchi, M. Nakajima, "Face identification using a 3D gray-scale image-a method for lessening restrictions on facial directions," in *Proc. AFGR*, pp.306-311, 1998.

[35] Wang, Y., C. Chua, Y. Ho, "Facial feature detection and face recognition from 2D and 3D images," *Pattern Recognition Letters*, vol.23, pp.1191-1202, 2002.

[36] Xu, C., Y. Wang, T. Tan, L. Quan, "Automatic 3D face recognition combining global geometric features with local shape variation information," in *Proc. AFGR*, pp. 308-313, 2004.

[37] Yacoob, Y., L.S. Davis, "Labeling of human face components from range data," *CVGIP: Image Understanding*, vol.60, no.2, pp.168-178, 1994.

Table 3: Overview of 3D face recognition systems

Group	**Representation**	**Database**	**Algorithm**	**Notes**
Gordon [13]	curvatures	26 training 24 test	Euclidean nearest neighbour	Curvatures can be used for feature detection but they are sensitive to smoothing.
Tanaka et al. [31]	curvature based EGI	NRCC	Fisher's spherical correlation	Use principal curvatures instead of surface normals for non-polyhedral objects.
Moreno et al. [24]	Curvature, line, region features	7 img.× 60 persons	Euclidean nearest neighbour	Angle, distance and curvature features work better than area based features.
Achermann and Bunke [2]	point cloud	120 training 120 test	Hausdorff nearest neighbour	Hausdorff distance can be speeded up by voxel discretization.
Lao et al. [17]	curve segments	36 img.× 10 persons	Euclidean nearest neighbour	Points with bad correspondence are not used in distance calculation.
Medioni and Waupotitsch [23]	mesh	7 img.× 100 persons	normalized cross-correlation	After alignment, a distance map is found. Statistics of the map are used in similarity.
İrfanoğlu et al. [16]	point cloud	3D_RMA	Point set difference (PSD)	ICP used to align point clouds with a base mesh. PSD outperforms PCA on depth map.
Lu et al. [19]	mesh	90 training 113 test	hybrid ICP and cross-correlation	ICP distances and shape index based correlation can be usefully combined.
Xu et al. [36]	regular mesh	3D_RMA	Feature extraction, PCA and NN	Feature derivation + PCA around landmarks worked better than aligned mesh distances.
Lu and Jain [20]	deformation points	500 training 196 test	ICP + TPS, nearest neighbour	Distinguishing between inter-subject and intra-subject deformations helps recognition.
Achermann et al. [1]	depth map	120 training 120 test	eigenface vs. HMM	Eigenface outperforms HMM. Smoothing is good for eigenface, bad for HMM.
Hesher et al. [15]	mesh	FSU	ICA or PCA+ nearest neighbour	ICA outperforms PCA, PCA degrades more gracefully as training samples are decreased.
Lee et al. [18]	depth map	2 img.× 35 persons	feature extraction+ nearest neighbour	Mean and variance of depth from windows around the nose are used as features.
Srivastava et al. [30]	depth map	6 img.× 67 persons	subspace projection + SVM	Optimal subspace found with MCMC simulated annealing outperforms PCA, ICA and LDA.
Cartoux et al. [9]	profile	3/4 img.× 5 persons	curvature based nearest neigbour	High quality images needed for principal curvatures.
Nagamine et al. [25]	vertical, horiz., circular profiles	10 img.× 16 persons	Euclidean nearest neigbour	Central vertical profile and circular sections touching eye corners are most informative.
Beumier and Acheroy [4]	vertical profiles	3D RMA	area based nearest neigbour	Central profile and mean lateral profiles are fused by averaging.
Blanz and Vetter [6]	2D+viewpoint parameters	CMU-PIE, FERET	analysis by synthesis	Using a generic 3D model, 2D viewpoint parameters are found.
Malassiotis and Strinzis [22]	texture+ depth map	110 img.× 20 persons	embedded HMM +fusion	Depth is better than colour, fusion is best. Pose correction is better than illumination correction.
Tsutsumi et al. [34]	texture + depth map	35 img.× 24 persons	concatenated features+PCA	Adding perturbed versions of training images reduces sensitivity of PCA.
Beumier and Acheroy [5]	2D and 3D vertical profiles	3D RMA	nearest neighbour +fusion	Combination of 2D and 3D helps. Temporal fusion (snapshots taken in time) helps too.
Wang et al. [35]	point signature Gabor features	6 img.× 50 persons	concatenation after PCA+SVM	Omit 3D info from the eyes, eyebrows (missing elements) and mouth (expression sensitivity)
Bronstein et al. [8]	texture+ depth map	157 persons	concatenation after PCA+near. neigh.	Bending-invariant canonical representation is robust to facial expressions.
Chang et al. [10]	texture+ depth map	278 training 166 test	Mahalanobis based near.neigh.+fusion	Pose correction through 3D is not better than rotation-corrected 2D.
Pan et al. [26]	profile+ point cloud	3D RMA	ICP+Hausdorff +fusion	Surface and profile combined usefully. Discard worst points (10 per cent) during registration.
Tsalakanidou et al. [32]	texture+ depth map	XM2VTS	nearest neighbour +fusion	Fusion of frontal colour and depth images with colour faces from profile.
Tsalakanidou et al. [33]	texture+ depth map	60 img.× 50 persons	embedded HMM +fusion	Appropriately processed texture is more informative than warped depth maps.
Pan and Wu [27]	depth map +profile	6 img.× 120 persons	(kernel) Fisherface +Eigenface+fusion	Sum rule is preferred to max, min, product, median and majority vote for fusion.
Papatheodorou and Rückert [29]	dense mesh + texture	12 img.× 62 persons	nearest neighbour +fusion	3D helps 2D especially for profile views. Texture has small relative weight.
Lu and Jain [21]	mesh +texture	598 test scans	ICP(3D), LDA(2D) + fusion	Difficult samples are evaluated by the combined scheme.
Gökberk et al. [14]	surface normals, profiles, depth map, point cloud	3D RMA	PCA, LDA, nearest neighbour, rank based fusion	Best single classifier is depth-LDA. Combining it with surface normals and profiles with nonlinear rank sum increases accuracy.