Rank-based Decision Fusion for 3D Shape-based Face Recognition

Berk Gökberk, Albert Ali Salah, and Lale Akarun

Boğaziçi University
Computer Engineering Department
TURKEY
{gokberk,salah,akarun}@boun.edu.tr

Abstract. In 3D face recognition systems, 3D facial shape information plays an important role. Various shape representations have been proposed in the literature. The most popular techniques are based on *point clouds*, *surface normals*, *facial profiles*, and statistical analysis of *depth images*. The contribution of the presented work can be divided into two parts: In the first part, we have developed face classifiers which use these popular techniques. A comprehensive comparison of these representation methods are given using 3D RMA dataset. Experimental results show that the *linear discriminant analysis-based* representation of depth images and *point cloud* representation perform best. In the second part of the paper, two different multiple-classifier architectures are developed to fuse individual shape-based face recognizers in parallel and hierarchical fashions at the decision level. It is shown that a significant performance improvement is possible when using rank-based decision fusion in ensemble methods.

1 Introduction

Despite two decades of intensive study, the challenges of face recognition remain: changes in the illumination and in-depth pose problems make this a difficult problem. Recently, 3D approaches to face recognition have shown promise to overcome these problems [1]. 3D face data essentially contains multi-modal information: *shape* and *texture*. Initial attempts in 3D research have mainly focused on *shape* information, and combined systems have emerged which fuse shape and texture information.

Surface normal-based approaches use facial surface normals to align and match faces. A popular method is to use the EGI representation [2, 3]. *Curvature-based* approaches generally segment the facial surface into patches and use curvatures or shape-index values to represent faces [4]. *Iterative Closest Point-based (ICP)* approaches perform the registration of faces using the popular ICP algorithm [5], and then define a similarity according to the quality of the fitness computed by the ICP algorithm [6–8]. *Principal Component Analysis-based (PCA)* methods first project the 3D face data into a 2D intensity image where the intensities are determined by the depth function. Projected 2D depth images can later

be processed as standard intensity images [9–11]. *Profile-based* or *contour-based* approaches try to extract salient 2D/3D curves from face data, and match these curves to find the identity of a person [12, 13]. *Point signature-based* methods encode the facial points using the relative depths according to their neighbor points [14, 15].

In addition to the pure shape-based approaches, 2D texture information has been combined with 3D shape information. These multi-modal techniques generally use PCA of intensity images [16, 17], facial profile intensities [13], ICP [18, 19], and Gabor wavelets [14]. These studies indicate that combining shape and texture information reduces the misclassification rate of a face recognizer.

One aim in this paper is to evaluate the usefulness of state-of-the-art shape-based representations and to compare their performance on a standard database. For this purpose, we have developed five different 3D shape-based face recognizers. They use: *ICP-based point cloud* representation, *surface normal-based* representation, *profile-based* representation, and two *depth image-based* representations: PCA and Linear Discriminant Analysis (LDA), respectively. Our second aim is to analyze whether combining these distinct 3D shape representation approaches can improve the classification performance of a face recognizer. To accomplish the fusion, we have designed two fusion schemes, *parallel* and *hierarchical*, at the sensor decision level. Although it has been shown in the literature that fusion of texture and shape information can increase the performance of the system, the fusion of different 3D shape-based classifiers has remained as an open problem. In this work, we show that the integration of distinct shape-based classifiers by using a rank-based decision scheme can greatly improve the overall performance of a 3D face recognition system.

2 3D Shape-based Face Recognizers

2.1 Registration

Registration of facial data involves two steps: a preprocessing step and a transformation step. In the preprocessing step, a surface is fitted to the raw 3D facial point data. Surface fitting is carried out to sample the facial data regularly. After surface fitting, central facial region is cropped and only the points inside the cropped ellipsoid are retained. In order to determine the central cropping region, nose tip coordinates are used. Figure 1 shows a sample of the original facial data, and the cropped region. Cropped faces are translated so that the nose tip locations are at the same coordinates. In the rest of the paper, we refer to the cropped region as the facial data.

After preprocessing of faces, a transformation step is used to align them. In the alignment step, our aim is to rotate and translate faces such that later on we can define acceptable similarity measures between different faces. For this purpose, we define a *template face model* in a specific position in the 3D coordinate system. Template face is defined as the average of the training faces. Each face is rigidly rotated and translated to fit the template. Iterative Closest Point

(ICP) algorithm is used to find rotation and translation parameters. The correspondences found between the template face and any two faces F_i and F_j by the ICP algorithm are then used to establish point-to-point dense correspondences.

2.2 3D Facial Shape Representations

Several 3D features can be extracted from registered faces. The simplest feature consists of the 3D coordinates of each point in the registered facial data (*point cloud representation*). Another representation, *surface normal representation*, is based on surface normals calculated at each 3D facial point. Both point cloud and surface normal-based approaches are related to whole facial surfaces. Besides surface-based features, facial profiles are also found to be important for discriminating 3D faces. In this work, we have extracted seven equally spaced vertical profiles, one central and three from either side of the profile (*profile set representation*). See Figure 1.b for the extracted profiles.

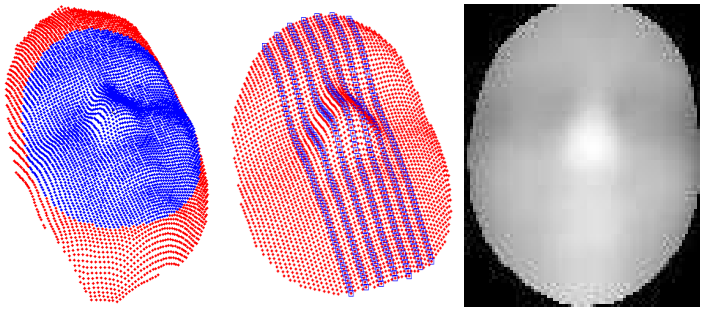

Fig. 1. (a) Cropped region, (b) extracted facial profiles, and (c) depth image

Facial profile can be found by using the 3D symmetry property of faces. However, in this work, we have used the nose region to find the central profile. We use the (x, y) coordinates of the topmost points over the nose. These points form an approximately ellipsoid cluster on the xy-plane. The vertical line passing through the center of nose can then be easily found by calculating the principal direction. To find the principal direction, we have performed PCA on the x and y coordinates of the topmost k nose points. Since all faces are registered to the template face, we can speed up the profile extraction process by simply finding the first principal direction in the face template once, and searching for closest points in a given registered 3D face image. This approach performs better since average template face is more robust to irregular nose shapes.

Registration of profile contours is performed by translating profile curves in such a way that nose tips of profiles are always at the same xy- coordinates. After aligning profile curves, a spline is fitted to the profile curve, and it is regularly sampled in order to be able to compute Euclidean distances between two profiles.

In the PCA and LDA techniques, the 3D face points are projected to a 2D image where the intensity of a pixel denotes the depth of a 3D point. Figure 1.c shows a sample depth image. Statistical feature extraction methods can be used to extract features from depth images. In this work, we have employed PCA (*Depth-PCA*) and LDA (*Depth-LDA*) to extract features from depth images.

2.3 Similarity Measures and Classifiers

In our system, we have used k-nearest neighbor algorithm (k-NN) as a pattern classifier which is intensively used in face recognition systems due to its high recognition accuracy. In order to use k-NN, we have to define a similarity measure for each representation used. Let Φ_i be a 3D face. We can represent Φ_i in *point cloud representation* as $\Phi_i^P = \{p_1^i, p_2^i, ...p_N^i\}$, where N is the number of points in the face and p^is are 3D coordinates. We define the distance between two faces Φ_i and Φ_j as:

$$D(\Phi_i^P, \Phi_j^P) = \sum_{k=1}^{n} ||p_k^i - p_k^j|| \tag{1}$$

where $||.||$ denotes Euclidean norm. Similarly, in *surface normal representation*, face Φ_i is represented by $\Phi_i^N = \{n_1^i, n_2^i, ...n_N^i\}$, where n^is are surface normals calculated at points p^is. Distance between two faces Φ_i and Φ_j in surface normal representations can be defined as in the above formula, but by replacing Φ^Ps with Φ^Ns.

In *profile set representation*, we have seven equally spaced vertical profiles, $C_k, (k = 1..7)$. Each profile curve C_k is a vector and contains n_k depth coordinates: $C_k = \{z_1, z_2, ..., z_{n_k}\}$. Therefore, we represent a face in profile set representation as $\Phi_i^R = \bigcup C_k$. The distance between two corresponding k^{th} profile curves of face i and face j can be determined by $d(C_k^i, C_k^j) = \sum_{m=1}^{n_k} ||z_m^i - z_m^j||$. Then, the distance between faces Φ_i and Φ_j is defined as the sum of the distances between each corresponding profile curve. In depth image-based face representations, the distance between two faces is calculated as the Euclidean distance between extracted feature vectors.

3 Combination of Shape-based Face Recognizers

When working on different representations, classifiers can be made more accurate through combination. Classifier combination has caught the attention of many researchers due to its potential for improving the performance in many applications [20–22]. In classifier fusion, the outputs of individual classifiers (*pattern classifiers*) are fused by a second classifier (*combination classifier*) according to a combination rule. In order to produce a successful ensemble classifier, individual pattern classifiers should be highly tuned, diverse, and should not be redundant. In this work, the diversity of the pattern classifiers is provided by letting them use a different face representation. In our system, the outputs of individual pattern classifiers are the ranked class labels and their associated similarity scores.

However, we only use the rank information because of the variability of the score functions produced by different representations.

In our fusion schemes, a *combination set* is formed by selecting the most similar k classes for each pattern classifier and by feeding these into the combining classifier. As *combination rules* for rank-output classifiers, we have used *consensus voting*, *rank-based combination* and *highest-rank majority* methods [23]. In *consensus voting*, the class labels from the combination set of each pattern classifier are pooled, and the most frequent class label is selected as the output. In *rank-based combination*, the sum of the rankings of each class in all combination sets are used to compute a final ranking (*rank-sum* method). A generalization of the rank-sum method is to transform ranks by a function f which maps ranks $\{1, 2, 3, ..., K\}$ to $\{f(1), f(2), f(3), ..., f(K)\}$. f may be any nonlinear monotonically increasing function. The motivation to use such a function f is to penalize the classes at the bottom of a ranked list. In this work, $f(x) = x^n$ is used as a mapping function. In *highest-rank majority*, a consensus voting is performed among the rank-1 results of each pattern classifier.

3.1 Parallel Fusion of Face Classifiers

We have designed a parallel ensemble classifier which fuses the rank-outputs of different face pattern classifiers. Profile set, Depth-LDA, point cloud and surface normal-based face representations are chosen in these pattern classifiers. Combination set is formed by selecting the most similar N classes in the rank outputs of each classifier. As a combination rule, four different types of rules are used: consensus voting, rank-sum, nonlinear rank-sum and highest-rank majority rule. In nonlinear rank-sum method, $f(x) = x^n$ function is used. If $n = 1$, nonlinear rank-sum method is identical to the standard rank-sum method. As a generalization of the rank outputs of individual classifiers, we have also used the ranking of each training instance whereas in standard rank-output classifiers, classes are assigned a single rank. In the rest of the paper, we will refer to the generalized method as *instance-based ranking*, and the standard method as *class-based ranking*. See Figure 2 for a schematic diagram of the parallel fusion scheme.

3.2 Hierarchical Fusion of Face Classifiers

In addition to the parallel fusion scheme, we have also designed a hierarchical fusion methodology. The main motivation of the hierarchical architecture is to filter out the most similar K classes using a simple classifier, and then to feed these K classes into a more complex and powerful second classifier. For this purpose, we have used the point cloud-based nearest neighbor classifier as the first classifier C_1, and depth map-based LDA classifier as the second classifier C_2. The use of LDA as a second classifier is based on the idea that it can boost the differences between similar classes in the transformed feature space. See Figure 3 for a schematic diagram of the hierarchical fusion.

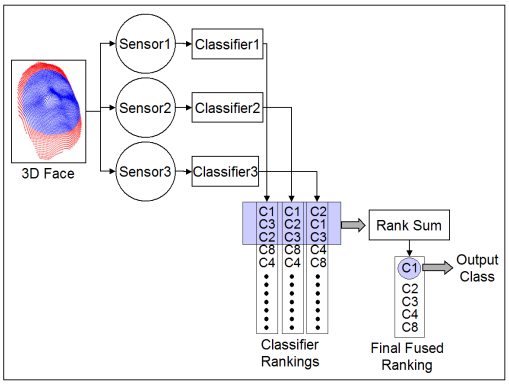

Fig. 2. A schematic diagram of the parallel combination scheme

As in the previous section, C_1 produces an instance-based ranking R_1, and then class labels of the top K instances are passed to C_2. C_2 then performs a linear discriminant analysis on the depth images of the training examples of these classes, and forms a feature space. Nearest neighbor classifier is used in this feature space to produce a new instance-based ranking R_2. If only the rank-1 class output of R_2 is used, the information in C_1 is discarded. We use a nonlinear rank-sum method to fuse R_1 and R_2 which is superior to using R_2 alone.

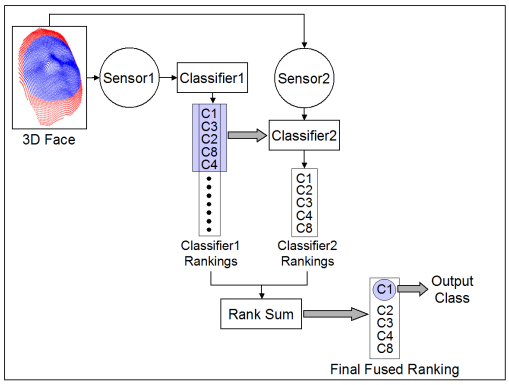

Fig. 3. A schematic diagram of the hierarchical combination scheme

4 Experimental Results

In our experiments, we have used the 3D RMA dataset [13]. Specifically, a subset of the automatically prepared faces were used in experiments, which consists of 106 subjects each having five or six shots. The data is obtained with a stereo vision assisted structured light system. On the average, faces contain about 4000 3D points, and they cover different portions of the faces and the entire data is subject to expression and rotation changes. To be able to statistically compare the algorithms, we have designed five experimental sessions.

Table 1 shows which shots of a subject are placed into the training and test sets for each session. At each session, there are exactly 193 test shots in total.

Table 1. Training and test set configurations

Session	Training Set Shots	Test Set Shots
S_1	{1, 2, 3, 4}	{5, 6}
S_2	{1, 2, 3, 5}	{4, 6}
S_3	{1, 2, 4, 5}	{3, 6}
S_4	{1, 3, 4, 5}	{2, 6}
S_5	{2, 3, 4, 5}	{1, 6}

4.1 Performance of Different Shape Features

Table 2 summarizes the classification accuracies of each representation method. Best performance is obtained using *Depth-LDA* which has an average recognition accuracy of 96.27 per cent. The dimensionality of the reduced feature vector of the Depth-LDA method is 30. Point cloud and surface normal representations 95.96 and 95.54 per cent correct recognition rate on the test set, respectively. In each of these representation schemes, feature vector size is $3,389 \times 3 = 10167$, since there are 3,389 points in each representation method and each point is a 3D vector. Profile set representation has a recognition accuracy of 94.30 per cent. The feature dimensionality of the profile set representation is the sum of the number of sampled points for each individual profile curve. In our representation, this dimensionality is 1,557. Depth-PCA method performed worst with a 50.78 per cent recognition accuracy, using 300 dimensional feature vectors.

4.2 Performance of Parallel and Hierarchical Decision Fusion Schemes

In our experiments on the parallel fusion scheme, we have tested all possible combinations of *point-cloud*, *surface normal*, *profile-set*, and *Depth-LDA* based classifiers. We have also analyzed the effect of the combination set size (N) in

Table 2. Classification accuracies of each classifier for each experimental session. d denotes the feature dimensionality of the representations

Session	Point Cloud ($d = 3,389 \times 3$)	Surface N. ($d = 3,389 \times 3$)	Depth-PCA ($d = 300$)	Depth-LDA ($d = 30$)	Profile Set ($d = 1,557$)
S_1	93.26	93.26	49.74	95.34	94.30
S_2	94.82	97.93	52.33	97.41	92.75
S_3	96.89	93.26	49.74	95.34	92.75
S_4	97.41	96.89	51.30	96.37	95.86
S_5	97.41	96.37	50.78	96.89	95.86
Mean	95.96	95.54	50.78	96.27	94.30
STD	1.85	2.16	1.10	0.93	1.55

the fusion process. Average recognition accuracies of different ensemble architectures are shown in Table 3. The best classification accuracy is obtained by a nonlinear rank-sum combination rule where the pattern classifiers are *profile set*, *Depth-LDA* and *surface normal*-based representations. In this architecture, combination set size is $N = 6$, and the nonlinear function used is $f(x) = x^3$. It is seen that instance-based ranking outperforms class-based ranking except for the highest rank majority rule. As a combination rule, nonlinear rank-sum method consistently outperforms its alternatives. We observe that parallel combination of different pattern classifiers which rely on distinct feature sets significantly improves the recognition accuracies in all cases. We confirm this finding with paired t-test on five-fold experiments.

Table 3. Mean classification accuracies of hierarchical fusion methods. S denotes the selected individual classifiers in the ensemble, where S = { 1: Profile set, 2: Depth-LDA, 3: Point cloud, 4: Surface Normals}

	Instance-based Ranking	Class-based Ranking
Consensus Voting	98.76 (N=2) S={2,3,4}	98.34 (N=1) S={1,2,3,4}
Nonlinear Rank-Sum	99.07 (N=6) S={1,2,4}	98.86 (N=1) S={1,2,3,4}
Highest Rank Majority	98.13 (N=1), S={1,2,3,4}	98.34 (N=1) S={1,2,3,4}

In hierarchical fusion experiments, point-cloud-based first classifier C_1 produces an *instance-based* rank list. On the average, first rank-80 instances provide 100 per cent recognition accuracy in C_1. We have seen that 80 training instances in the combination set corresponds to approximately 25 classes. Therefore, our Depth-LDA based second classifier C_2 dynamically constructs a feature space using these 25 classes. Finally, the ranks produced by C_1 and C_2 are integrated using nonlinear rank-sum technique where $f(x) = x^3$. The average performance of the hierarchically combined classifiers is found to be 98.13 per cent, which is statistically significantly different from all individual classifier's accuracies. As

in the parallel case, hierarchical fusion is found to be beneficial when compared to individual classifier accuracies. The accuracy of the parallel fusion of point cloud and Depth-LDA using nonlinear rank-sum is 98.45 per cent and is better than hierarchical fusion.

5 Conclusion

In this work, we have compared some of the state-of-the-art 3D shape-based face representation techniques frequently used in 3D face recognition systems. They include ICP-based point cloud representations, surface normal-based representations, PCA and LDA-based depth map techniques and facial profile-based approaches. It has been shown that among these methods, Depth-LDA method performs best, and point cloud and surface normal-based classifiers have a comparable recognition accuracy. Our results on Depth-PCA confirmed the sensitivity of PCA to alignment procedure. To obtain better results, facial landmarks need to be correctly aligned, possibly by warping of faces. In our work, we choose not to warp facial surfaces since it is known that such a warping process suppresses discriminative features [8].

We have also developed parallel and hierarchical combination schemes to fuse the outputs of individual shape-based classifiers. In the parallel architecture, a subset of the rank outputs of surface-normal, Depth-LDA, and profile-based classifiers are fused using nonlinear rank-sum method, and the recognition accuracy improved to 99.07 per cent from 96.27 per cent which is the best individual classifier's (Depth-LDA) accuracy. In the hierarchical fusion scheme, we transfer the most probable classes found by our first point-cloud based classifier to a Depth-LDA based second classifier, where LDA makes use of the differences between similar classes in the transformed feature space. The hierarchical architecture reaches a 98.13 per cent recognition accuracy which is statistically superior to all individual performances according to paired t-tests. As a conclusion, we observe that the combination of separate shape-based face classifiers improves the classification accuracy of the whole system, when compared to using individual classifiers alone. As a future work, we plan to investigate the fusion of shape-based ensemble classifiers with texture-based ensemble methods.

References

1. Bowyer, K.W., Chang, K., Flynn, P.J.: A survey of 3D and multi-modal 3D+2D face recognition. In: International Conference on Pattern Recognition. (2004)
2. Lee, J.C., Milios, E.: Matching range images of human faces. In: International Conference on Computer Vision. (1990) 722–726
3. Tanaka, H.T., Ikeda, M., Chiaki, H.: Curvature-based face surface recognition using spherical correlation principal directions for curved object recognition. In: International Conference on Automated Face and Gesture Recognition. (1998) 372–377

4. Moreno, A.B., Sanchez, A., Velez, J.F., Diaz, F.J.: Face recognition using 3D surface-extracted descriptors. In: Irish Machine Vision and Image Processing Conference. (2003)
5. Besl, P., McKay, N.: A method for registration of 3D shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence **14** (1992) 239–256
6. Medioni, G., Waupotitsch, R.: Face recognition and modeling in 3D. In: IEEE International Workshop on Analysis and Modeling of Faces and Gestures. (2003) 232–233
7. Lu, X., Colbry, D., Jain, A.: Matching 2.5d scans for face recognition. In: International Conference on Pattern Recognition. (2004) 30–36
8. Irfanoglu, M.O., Gokberk, B., Akarun, L.: 3D shape based face recognition using automatically registered facial surfaces. In: International Conference on Pattern Recognition. (2004) 183–186
9. Hesher, C., Srivastava, A., Erlebacher, G.: A novel technique for face recognition using range imaging. In: International Symposium on Signal Processing and Its Applications. (2003) 201–204
10. Pan, G., Wu, Z., Pan., Y.: Automatic 3d face verification from range data. In: International Conference on Acoustics, Speech, and Signal Processing. Volume 3. (2003) 193–196
11. Xu, C., Wang, Y., Tan, T., Quan, L.: Automatic 3D face recognition combining global geometric features with local shape variation information. In: International Conference on Automated Face and Gesture Recognition. (2004) 308–313
12. Lee, Y., Park, K., Shim, J., Yi, T.: 3D face recognition using statistical multiple features for the local depth information. In: International Conference on Vision Interface. (2003)
13. Beumier, C., Acheroy, M.: Face verification from 3D and grey level cues. Pattern Recognition Letters **22** (2001) 1321–1329
14. Y.Wang, Chua, C., Ho, Y.: Facial feature detection and face recognition from 2D and 3D images. Pattern Recognition Letters **23** (2002) 1191–1202
15. Chua, C.S., Han, F., Ho, Y.K.: 3D human face recognition using point signature. In: Proceedings of Int. Conf. on Automatic Face and Gesture Recognition. (2000) 233–237
16. Tsalakanidou, F., Tzocaras, D., Strintzis, M.: Use of depth and colour eigenfaces for face recognition. Pattern Recognition Letters **24** (2003) 1427–1435
17. Chang, K., Bowyer, K., Flynn, P.: Face recognition using 2D and 3D facial data. In: Multimodal User Authentication Workshop. (2003) 25–32
18. Papatheodorou, T., Reuckert, D.: Evaluation of automatic 4d face recognition using surface and texture registration. In: International Conference on Automated Face and Gesture Recognition. (2004) 321–326
19. Lu, X., Jain, A.K.: Integrating range and texture information for 3D face recognition. In: IEEE Workshop on Applications of Computer Vision. (2005) To appear
20. Toygar, O., Acan, A.: Multiple classifier implementation of a divide-and-conquer approach using appearance-based statistical methods for face recognition. Pattern Recognition Letters **25** (2004) 1421–1430
21. Khuwaja, G.A.: An adaptive combined classifier system for invariant face recognition. Digital Signal Processing **12** (2002) 21–46
22. Jing, X., Zhang, D.: Face recognition basedon linear classifiers combination. Neurocomputing **50** (2003) 485–488
23. Melnik, O., Vardi, Y., Zhang, C.H.: Mixed group ranks: Preference and confidence in classifier combination. IEEE Transactions on Pattern Analysis and Machine Intelligence **26** (2004) 973–981