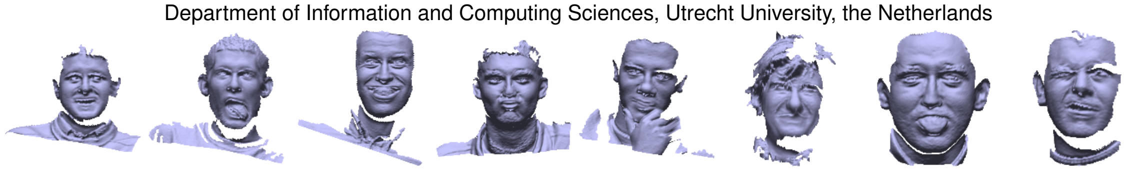
SHape REtrieval Contest 2008: 3D Face Scans

Frank B. ter Haar*
Department of Information and Computing Sciences, Utrecht University, the Netherlands
Mohamed Daoudi†
Institut Telecom, Telecom Lille 1, LIFL (UMR USTL/CNRS 8022), France
Remco C. Veltkamp‡
Department of Information and Computing Sciences, Utrecht University, the Netherlands



1 Introduction

Three-Dimensional face recognition is a challenging task with a large number of proposed solutions [1, 2]. With variations in pose and expression the identification of a face scan based on 3D geometry is difficult. To improve on this task and to evaluate existing face matching methods large sets of 3D faces were constructed, such as the FRGC [3], BU-3DFE [4], and the GavabDB [5] database. When used in the same experimental way, these publicly available sets allow for a fair comparison of different methods. Usually, researchers compare the recognition rates (or identification rates) of different methods. To identify a person, its 3D face scan is enrolled as query in the database and if the most similar scan (other than the query) in the database belongs to the same person, he or she is identified correctly. For a set of queries, the recognition rate is computed as the average of zeros (no identification) and ones (correct identification). However, the recognition rate is a limited evaluation measure, because it considers merely the closest match of each query. In case you are using a database that contains two scans per expression per subject and you use each scan as query once, you are bound to find the similar scan on top of the ranked list. Such an experiment boosts the recognition rate, but gives no insight of the expression invariance of different methods. For that, an evaluation measure is required that takes a larger part of the ranked list into account. In this contest we compare different face matching methods using a large number of performance measures. As a test set we have used a processed subset of the GavabDB [5], which contains several expressions and pose variations per subject.

2 Database

For the retrieval contest of 3D faces we have used a subset of the GavabDB [5]. The GavabDB consists of Minolta Vi-700 laser range scans from 61 different subjects. The subjects, of which 45 are male and 16 are female, are all Caucasian. Each subject was scanned nine times for different poses and expressions, namely six neutral expression scans and three scans with an expression. The neutral scans include two different frontal scans, one scan while looking up (≈+35°), one scan while looking down (≈-35°), one scan from the right side (≈+90°), and one from the left side (≈-90°). The expression scans include one with a smile, one with a pronounced laugh, and an "arbitrary expression" freely chosen by the subject.

*e-mail: fhaar@cs.uu.nl
†e-mail: mohamed.daoudi@telecom-lille1.eu
‡e-mail: Remco.Veltkamp@cs.uu.nl

Participant	Affiliation	Reference	Runs
Amberg et al.	University of Basel, Switzerland	[7]	4
Berretti et al.	University of Florence, Italy	[8]	2
Nair et al.	University of London, England	[9]	4
ter Haar et al.	University of Utrecht, the Netherlands	[10]	4
Xu et al.	Chinese Academy of Sciences, China	[11]	5

Table 1: Participants of SHREC 2008 - Shape Retrieval Contest of 3D Face Scans

A basic hole filling technique was applied to fill some simple holes in the scan data.

In our face retrieval contest we have focussed on the retrieval of all subject's scans, regardless of the query. Since it is impossible to retrieve a left side scan using a right side scan as a query (without the assumption of facial symmetry), we excluded the side scans from the contest's database. All other 7 times 61 (427) scans were converted to triangular meshes in OFF-format and renamed to a random index number (D1xxx) in the final SHREC'08 3D face database.

The problem with 3D face scans in general is their variation in pose. Even when a subject is asked to gaze at the scanner, two sequential scans will most likely show small pose variations. Because most 3D face matching algorithms are not invariant to changes in pose and some methods require the tip of the nose as a reference point, we applied an automatic method to normalize the pose and localize the tip of the nose. This way the contest emphasizes on the actual 3D face matching, because the influence of different pose normalization methods on the retrieval performance is reduced. The applied pose normalization method is based on [6] and can be briefly summarized as follows. Each surface mesh is randomly sampled such that every ≈2.0 mm^2 of the surface is approximately sampled once. These locations, in combination with their surface normals, are used as initial placements for a small nose tip template. To locations where this template fits well, a second template of global face features is fitted to normalize the face's pose and to select the tip of the nose. After that, the pose normalized scan was placed with its selected tip of the nose in the origin. The normalized poses of a subject's seven scans are shown in Figure 1.

The set of 427 pose normalized face scans were distributed among the participants listed in Table 1. Participant used each of the 3D face scans as a query once, constructing a ranked list with dissimilarity values in the process. For a method to perform well on the contest's database it needs to cope with the facial expressions. In particular querying with the scan where the subject shows a pronounced expression requires a certain level of expression invariance, otherwise it is hard to retrieve the relevant frontal scans. Several of these scans are shown in the figure below the title.

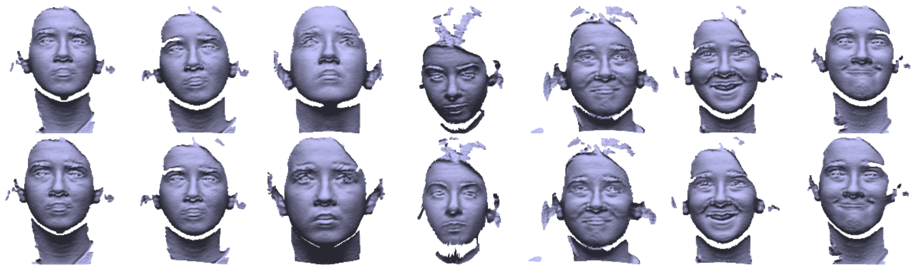


Figure 1: 3D face scans corresponding to the different variations of one individual in the database, before (top) and after (bottom) face pose normalization. From left to right *front1, front2, up, down, smile, laugh* and the *arbitrary expression*

	MAP	RR	1st Tier	2nd Tier	MADP	MNDCG@10
ideal	1.000	100.0%	100.0%	100.0%	1.000	1.000
Berretti run1	0.935	99.5%	90.4%	94.2%	0.977	0.952
Amberg run2	0.897	98.6%	85.5%	90.6%	0.946	0.919
Amberg run1	0.850	98.8%	80.8%	86.1%	0.938	0.892
Amberg run3	0.753	92.5%	69.3%	77.5%	0.866	0.806
Amberg run4	0.720	93.2%	66.5%	73.5%	0.860	0.788
ter Haar run4	0.693	91.1%	62.4%	72.4%	0.823	0.757
Xu run5	0.674	81.7%	61.2%	71.9%	0.793	0.734
ter Haar run2	0.668	86.9%	60.7%	70.4%	0.809	0.739
Nair run4	0.661	82.2%	60.5%	67.9%	0.791	0.721
Berretti run2	0.654	83.8%	59.5%	69.0%	0.791	0.722
Nair run1	0.641	80.8%	58.4%	66.7%	0.776	0.703
ter Haar run3	0.615	81.5%	55.7%	65.5%	0.766	0.691
ter Haar run1	0.561	76.1%	50.6%	60.5%	0.729	0.646
Nair run3	0.539	75.2%	48.5%	56.8%	0.712	0.625
Xu run2	0.534	72.4%	47.7%	57.5%	0.699	0.616
Nair run2	0.522	74.0%	47.3%	55.0%	0.702	0.610
Xu run3	0.504	69.6%	45.5%	53.9%	0.683	0.594
Xu run4	0.474	60.2%	43.3%	51.8%	0.645	0.559
Xu run1	0.460	63.5%	42.4%	49.5%	0.653	0.556

Table 2: Retrieval results of all participants sorted on the mean average precision

3 Performance measures

The results of the 427×427 face matches were submitted to the origanizers by means of 427 ranked lists. These ranked lists were turned into gain vectors by replacing item IDs by their relevance scores. Scans from the same subject are all classified as relevant (score 1) and scans from other subjects as non-relevant (score 0). Several performance measures were used to evaluate each gain vector, e.g. the average precision. The mean results over all queries are used to evaluate and compare different (settings of) algorithms. Table 2 shows the results of this contest for six different distance measures, namely the recognition rate (RR), the first and second Tier, the mean average precision (MAP), the mean average dynamic Precision (MADP), and the mean normalized discounted cumulative gain for the top ten rankings (MNDCG@10). For more performance measures and their definitions we refer to our website [12].

4 Results

Results in Table 2 show high performances for the methods of Amberg and Berretti. The recognition rates of most methods are very high, which means that for most of the queries a relevant face (other than the query) is retrieved on top of the ranked list. For the frontal scans with neutral expression, recognition should be relatively easy, because each subject has two of them in the database. More interesting are the retrieval results using one of the other performance measures. The First Tier, for instance, shows the percentage of relevant items within the top seven (class size) of all ranked lists. The MADP shows the mean average precision within the top seven. The Second Tier shows the percentage within the top 14 (twice the class size). These measures are similar in the sense that they all take a rather small scope into account. The MAP takes the entire ranked list (scope of 427) into account, and together with for instance the MADP and the RR they could give a more complete view on the overall retrieval performance.

5 Concluding remarks

For this contest we have adapted an existing publicly available dataset of 3D face scans, namely the GavabDB. To focus on the retrieval of relevant scans, the scans were automatically pose normalized in a preprocessing step. Five participant submitted retrieval results with a total number of 19 runs, which were evaluated using performance measure such as the recognition rate, the mean average precision and the mean average dynamic precision. The contest enabled participants to effectively compare different algorithms to their own and to those of other participants. The highest MAP was obtained by Berretti et al. (run1 0.935) closely followed by Amberg et al. (run2 0.897). In particular, the arbitrary expression and the pronounced laugh cause large geometric changes between the query face and its relevant faces. The two runs of these participants achieve good retrieval results for these cases, proving a high level of expression invariance of their methods.

Acknowledgements

This research was partially supported by the FP6 IST Network of Excellence 506766 AIM@SHAPE and FOCUS-K3D FP7-ICT-2007-214993.

References

[1] K. W. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition," *CVIU*, vol. 101, no. 1, pp. 1–15, 2006.

[2] A. Scheenstra, A. Ruifrok, and R. C. Veltkamp, "A Survey of 3D Face Recognition Methods," in *AVBPA*, pp. 891–899, 2005.

[3] Face Recognition Grand Challenge, "FRGC."

[4] Binghamton University 3D Facial Expression database, "BU-3DFE."

[5] Gavab Department of Computing, "GavabDB."

[6] F. B. ter Haar and R. C. Veltkamp, "A 3D Face Matching Framework," in *Proc. Shape Modeling International (SMI'08)*, 2008.

[7] B. Amberg, R. Knothe, and T. Vetter, "SHREC'08 Entry: Shape Based Face Recognition with a Morphable Model," in *These proceedings*.

[8] S. Berretti, A. D. Bimbo, and P. Pala, "SHREC'08 Entry: 3D Face Recognition using Integral Shape Information," in *These proceedings*.

[9] P. Nair and A. Cavallaro, "SHREC'08 Entry: Registration and Retrieval of 3D Faces using a Point Distribution Model," in *These proceedings*.

[10] F. B. ter Haar and R. C. Veltkamp, "SHREC'08 Entry: 3D Face Recognition using Facial Contour Curves," in *These proceedings*.

[11] D. Xu, P. Hu, W. Cao, and H. Li, "SHREC'08 Entry: 3D Face Recognition using Moment Invariants," in *These proceedings*.

[12] Utrecht University, "SHREC 2008 - Shape Retrieval Contest of 3D Face Scans at `http://give-lab.cs.uu.nl/shrec/shrec2008`," March 2008.