3D Facial Similarity: Automatic Assessment versus Perceptual Judgments

Anush K. Moorthy, Anish Mittal, Sina Jahanbin, Kristen Grauman and Alan C. Bovik

***Abstract*— We develop algorithms that seek to assess the similarity of 3D faces, such that similar and dissimilar faces may be classified with high correlation relative to human perception of facial similarity. To obtain human facial similarity ratings, we conduct a subjective study, where a set of human subjects rate the similarity of pairs of faces. Such similarity scores are obtained from 12 subjects on 180 3D faces, with a total of 5490 pairs of similarity scores. We then extract Gabor features from automatically detected fiducial points on the range and texture images from the 3D face and demonstrate that these features correlate well with human judgements of similarity. Finally, we demonstrate the application of using such facial similarity ratings for scalable face recognition.**

I. Introduction

Humans naturally perceive similarity between different objects. Research has demonstrated that humans are particularly sensitive to similarity in faces [19]. Indeed, it has been suggested that humans tend to seek partners with similar facial attributes [9]. Given that humans are adept at gauging the similarity of faces, our goal is to algorithmically mimic this process. Applications of such similarity ratings are varied and include determining familial relationships, identification at security installations, judging facial aesthetics [20] and so on. One particularly interesting application is the use of face similarity ratings for scalable face recognition.

Face recognition has a rich history and a multitude of methods have been proposed to recognize faces [21]. Recently, research has moved from 2D face recognition to 3D face recognition [7]. The earliest approaches to 3D face recognition were simple extensions of 2D techniques. However, recently, researchers have proposed targeted approaches for 3D face recognition that utilize range and 2D texture information available from the 3D scan of a face with great success [12], [15], [17]. We are concerned with 3D faces here, and all future reference to a face assumes that range and texture information is available.

The goal of face recognition can be summarized as follows. Given a probe face, one needs to identify that face from the gallery of faces in the dataset that *exactly* matches the probe face. Our concern is not with *exact* matches, but with *similar* matches. In short, our goal is to create an algorithm that will receive as input a probe face, and produce as output a similarity measure (similar/dissimilar) for that probe face with respect to all faces in the database. An application of such a similarity measure is to gauge the degree of change in a face over time. Another application, which we are exploring, is to use such algorithms to gauge the efficacy of reconstructive plastic surgery on the face. As we mentioned before, one interesting application of gauging facial similarity is scalable face recognition.

Currently, to accomplish face recognition, given a probe face, one needs to scan the entire database of faces to recognize the most corresponding face. Instead, if one had a method of judging the similarity of faces, one could group faces in the database into similar clusters. In this case, the probe face needs to be examined only with a 'representative' face from each of the clusters. Once the cluster to which the face belongs to has been found, search through only the faces in that cluster needs to be performed for face recognition, thereby reducing computational cost considerably.

Face similarity has been studied for 2D faces [10], however, to the best of our knowledge, 3D face similarity has not yet been studied. Further, the algorithm to determine similarity of faces in [10] was simplistic. The authors only utilized a patch around manually chosen fiducial points as features for face similarity - this clearly is sub-optimal. Further, the human study conducted in [10] asked people to make a binary decision on whether the two faces were similar or not. If one were interested in the degree of similarity, such a rating process is insufficient.

Here, we propose to algorithmically rate the similarity of faces by extracting Gabor features at automatically chosen fiducial points from the range and texture images for each face. Similar features have been used for face recognition [12] and our motivation for their use was partly the success that the face recognition algorithm demonstrated [12] and partly our application to scalable face recognition. Using the same set of pre-computed face recognition features for face similarity reduces the amount of computation as well as storage requirements. Further, unlike the simple image patches in [10], the chosen features are robust to illumination and pose changes [12].

In order to evaluate the application of these features to gauge facial similarity, we conducted a human study where human observers were asked to rate the similarity of 3D faces on a 7-point Likert scale [18]. The mean similarity scores for pairs of faces are representative of human perceived similarity and we demonstrate that the chosen features correlate well with this human perceived similarity. Further, we also demonstrate how scalable face recognition may be achieved using these facial similarity ratings.

The rest of the paper is organized as follows. In Section II, we detail the human study to gauge facial similarity. Section III describes the automatic fiducial point detection approach

A. K. Moorthy, A. Mittal, S. Jahanbin and A. C. Bovik are with the Dept. of Elec. & Comp. Engg., K. Grauman is with the Dept. of Computer Science, The University of Texas at Austin, Austin, TX 78712, USA.

as well as the Gabor features extracted. In Section IV we evaluate the features' ability to agree with human facial similarity judgments and demonstrate how facial similarity measures may be used for scalable 3D face recognition. We conclude the paper in Section V.

II. Human study for Facial Similarity

In order to evaluate 3D facial similarity, we use the Texas 3D Face Recognition (FR) database [6]. This publicly available database of 3D facial images was acquired using a stereo imaging system. The database contains 1196 3D models - i.e., portrait (texture) and range pairs - of 116 adult human subjects. The subjects' ages range from 22 to 75 years. The database includes images of both males and females from major ethnic groups. The faces are in neutral and expressive modes. All subjects were requested to remove hats and eye-glasses prior to image acquisition. Processed range and portrait pairs are available to researchers at no cost [6]. For further details on the database, the reader is referred to [6].

From the 118 individual faces, we chose 60 faces with neutral expression so as to enable an easier morph (see below). The faces were divided into 3 groups of 20 faces each. Within each group, pairs of faces were morphed to create a total of 60 faces for a group, leading to a total of 180 faces. Human studies were conducted separately for each of the groups in order to minimize the amount of time needed for the study (see below). Hence, pair-wise similarity scores are available for faces within each group but not across groups.

A. Creating similar faces

It should be obvious to the reader that a random subset of faces chosen from a face recognition database will not provide a set of similar faces at the level of granularity that this project demands. Hence, in order to create similar faces, we chose to morph portrait-range pairs; such morphing was performed using the algorithm in [16].

As described above, we selected 60 faces from the Texas 3D FR database. In order to create a set of similar faces, we selected a pair of faces at random and manually marked 25 fiducial points on these faces as shown in Fig. 1. The chosen points were at readily identifiable locations on the face and such that they were uniformly distributed across the face in order to enable a better morph. We attempted the morphs using our automated fiducial point detection algorithm (see below), however, as we shall see, since the number of fiducial points from our algorithm are limited to 11, the obtained morphs were less than satisfactory. Hence, we chose to manually mark fiducial points on pairs of faces.

Once fiducial points from a particular pair were marked, the two faces were not considered for future morphs. Thus, not all possible pairs were morphed. Once the fiducial points were selected, we applied the algorithm in [16] to morph the faces. The morphing algorithm allows one to specify the degree of morph. For eg., assume that F_1 and F_2 represent the two faces, then the morphed faces $F_m = (1-\lambda)F_1+\lambda F_2$,

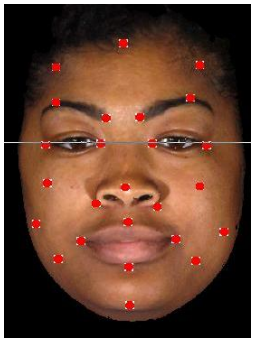


Fig. 1. Figure demonstrating the 25 manually chosen fiducial points used to produce morphs.

where λ is a parameter which controls how much the faces need to be mixed, i.e., the degree of the morph. $\lambda = 0$ implies that the morphed face is the same as F_1, while $\lambda = 1$ implies that the morphed face is the same as F_2. In our study, we chose $\lambda = \{0.2, 0.4, 0.6, 0.8\}$ to create four morphed faces for each pair of faces considered. Fig. 2 shows two individual faces (a) and (f) and various morphed faces for the different values of λ considered here.

B. Portrait-range to left-right pairs

The Texas 3DFR database contains portrait and range pairs as described before. However, in order to display the faces on a 3D screen, we need left-right pairs. In order to go from the portrait-range pairs to left-right pairs, we simulate a stereo setup.

The simulated stereo setup consists of a pair of vergent cameras with focal length $f = 80mm$ at a distance of $d = 30mm$ from the tip of the nose[1] such that the angle of vergence is $0.25°$. The simulated camera setup can then be used to create left-right pairs [14].

The simulated camera geometry was chosen arbitrarily so that the area occupied by the face in the image accounted for a sizeable portion of the image. Missing values were interpolated using bicubic interpolation. Median filtering was applied in order to smooth out 'pepper' noise that was observed in some of the left-right pairs.

Fig. 3 shows a portrait-range pair and the corresponding left-right pair obtained using the simulated stereo camera setup described above.

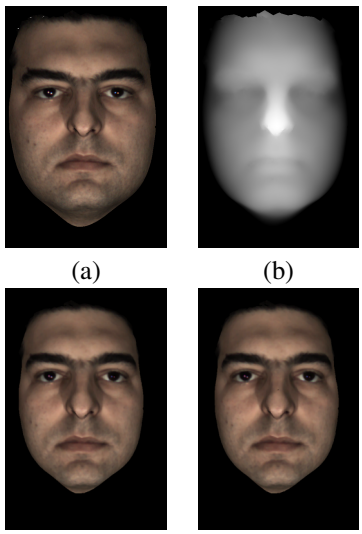
C. Human study

The subjective study was conducted over a period of two weeks at the University of Texas at Austin. The participants were mostly graduate student researchers in the field of image and video processing and voluntarily undertook the study. A total of 12 subjects participated in the study, with each group of 60 images viewed and rated by four subjects. The study was conducted in a room lit as per ITU recommendations [11]. The stimuli were shown on an IZ3D 22" monitor with a viewing distance corresponding to 5 times the height of the screen [11]. The monitor resolution was set

[1]The tip of the nose is assumed closest to the camera.

(a) (b) (c) (d) (e) (f)

Fig. 2. Figure demonstrating the morphed portrait faces - (a) and (f) are the two faces to be morphed, (b)-(e) morphed faces with $\lambda = \{0.2, 0.4, 0.6, 0.8\}$.



(a) (b) (c) (d)

Fig. 3. Figure shows how the portrait and range images in (a), (b) are converted to the left-right views in (c) and (d) using the described camera geometry.

Fig. 4. Figure shows the study setup. Subject compares the reference face at the top with each of the 6 faces at the bottom and rates the similarity on a scale of 1-7.

at 1680×1050 and the refresh rate was 60 Hz. The 3D display was circularly polarized, and the subjects viewed the images on the screen while wearing glasses.

The study took approximately 45 minutes, and the subject was free to take a break after about 20 minutes to avoid fatigue. The reference face was shown at the top of the screen and the faces whose similarity to the reference face needed to be assessed were shown at the bottom as seen in Fig. 4. A 7-point Likert scale [18] was used to gauge similarity. A value of 1 indicated that the two faces were completely dissimilar, while a value of 7 indicated that the two faces being compared were exactly the same.

The subject was instructed as follows: *You are taking part in a study to assess the similarity of faces. You will be shown a reference face and six other faces, whose similarity you need to assess with respect to the reference face. You will assess this similarity on a scale of 1-7, where 1 implies that the faces are completely dissimilar while 7 implies that the two faces are exactly the same.* None of the subjects had any trouble understanding the instructions.

Each subject made a total of 1830 comparisons[2]. Given that there were a total of 4 subjects for each set of 60 faces, each pair amongst the 60 faces received 4 human similarity ratings. We analyzed the subjective ratings provided for the variance between subjects and in Fig. 5 we plot a histogram of the standard deviation of similarity scores across subjects for the entire dataset. In general, the subjects agree with each other, with 73.65% of the ratings differing only by a standard deviation of ≤ 1 and 95.19% of the ratings differing by a standard deviation of ≤ 2. Hence we chose to simply compute the mean of the similarity scores across subjects. The mean of these similarity scores represents the human perceived similarity for each pair of faces in the database and is referred to as the mean similarity score (MSS) through the rest of the paper. MSS acts as the ground truth for our dataset.

Thus, we have three sets - set 1, set 2 and set 3 - each containing 60 faces with MSS for each pair of faces within a set. We also note that the study was designed such that the similarity matrix formed from the similarity scores is symmetric i.e., $sim(F_1, F_2) = sim(F_2, F_1)$, where $sim(\cdot, \cdot)$ is the MSS for a pair of faces.

III. Feature Extraction

In order to algorithmically judge facial similarity, we extract Gabor features from automatically detected feature points. Next, we describe in brief the automatic fiducial point detection algorithm and the feature vector extracted. The reader is referred to [13] for details.

Local appearance around a point, $\vec{x}$, from the range or portrait image $I(\vec{x})$ can be encoded using a set of Gabor coefficients $J_j(\vec{x})$ [3]. Each coefficient $J_j(\vec{x})$ is derived by

[2]There are a total of 60 images, and pairs of them can be chosen in $\binom{60}{2} = 1770$ ways. We also allow for reference-reference comparisons, hence the total number of comparisons made = $1770 + 60 = 1830$.

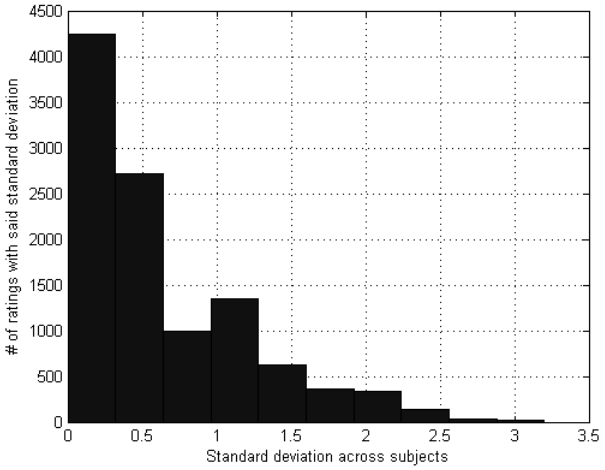


Fig. 5. Histogram of the standard deviation of similarity scores across subjects for the entire dataset. 73.65% of the ratings differ by a standard deviation of ≤ 1 and 95.19% of the ratings differ by a standard deviation of ≤ 2.

convolving the input image $I(\vec{x})$ with a family of Gabor kernels

$$\psi_j(\vec{x}) = \frac{k_j^2}{\sigma^2} \exp\left(\frac{-k_j^2 x^2}{2\sigma^2}\right) \left[\exp(i\vec{k}_j \cdot \vec{x}) - \exp(\frac{-\sigma^2}{2})\right] \quad (1)$$

Gabor kernels are plane waves modulated by a 2-D Gaussian function. One of the main motivations behind the use of Gabor filters is that they have been shown to model the receptive fields in area V1 of the primary visual cortex [4]; further, Gabor filters have been used in the past for a multitude of tasks [3], [2], [8].

In our implementation, $\sigma = 2\pi$ and each kernel is characterized by a wave vector $\vec{k}_j = [k_v \cos\phi_u \;\; k_v \sin\phi_u]^T$ where $k_v = 2^{-(v+1)}$, $v = 0, 1 \ldots, 4$ denote spatial frequencies and $\phi_u = (\phi/8)u$, $u = 0, 1 \ldots, 7$ are the orientations of the Gabor kernels.

A "jet" $\vec{J}$ is defined as a set $\{J_j, j = u + 8v\}$ of 40 complex Gabor coefficients obtained from a single image point. Complex Gabor coefficients can be represented in exponential form $J_j = a_j \exp(i\phi_j)$ where $a_j(\vec{x})$ is the slowly varying magnitude and $\phi_j(\vec{x})$ is the phase of the jth Gabor coefficient at pixel $\vec{x}$. These jets are extracted around a regions defined by the face geometry that contains the fiducial point that we wish to detect. Once these jets are extracted, the corresponding fiducial point is the one that is most similar to the trained Gabor 'bunch' template[3]. The similarity between a bunch and a jet is defined as the maximum value of the similarity between the jet and the constituent jets of the bunch:

$$S_B(\vec{J}, \vec{\vec{B}}) = \max_{i=1}^{50} S(\vec{J}, \vec{B}_{(i)}) \quad (2)$$

[3]A bunch consists of Gabor coefficients collected from several training images by manually marking that specific fiducial.

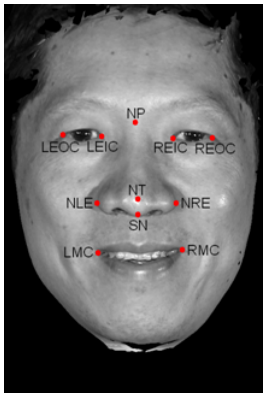


Fig. 6. Figure showing the 11 automatically detected fiducial points.

where in (2), $\vec{\vec{B}}$ represents a bunch with 50 constituent jets $\vec{B}_{(i)}$ and

$$S(\vec{J}, \vec{J'}) = \frac{\sum_{i=1}^{40} a_i \acute{a}_i \cos(\phi_i - \acute{\phi}_i)}{\sqrt{\sum_{i=1}^{40} a_i^2 \sum_{i=1}^{40} \acute{a}_i{}^2}} \quad (3)$$

defines the similarity between two jets. Such automatic fiducial point detection is undertaken for 11 fiducial points as seen in Fig. 6.

Once the fiducial points have been isolated, the magnitude of the Gabor jet at each of these points forms our feature vector. Since fiducial points are isolated on the range and portrait image we have a set of features from the range as well as the portrait image. Since there are 11 fiducial points, and 40 Gabor coefficients are extracted from these points, each feature is a 440 dimensional vector. Two such 440 dimensional feature vectors correspond to the range and portrait images for each face, which we believe are sufficient to characterize facial similarity.

In order to demonstrate that this feature vector gauges similarity well, we stack feature vectors from the range and portrait images to form an 880-dimensional feature vector (range+texture feature), and plot the cosine distance[4] between the feature vectors for a particular pair of faces in Fig. 7. Note how the distance varies smoothly as a function of reducing similarity.

IV. Performance Evaluation

We have demonstrated that the feature extracted captures facial similarity measurements when compared with the morphing parameter. In this section we describe how we use the extracted features along the simple cosine similarity to classify pairs of faces as being similar or dissimilar. Further, we explore the importance of range and portrait features and also of the individual features themselves. Finally, we also demonstrate how scalable face recognition can be achieved.

[4]Note that we experimented with the use of a metric learning algorithm [5], instead of using cosine similarity and we found that the use of the metric does not improve results significantly and hence we utilized the cosine measure as an indicator of similarity throughout.

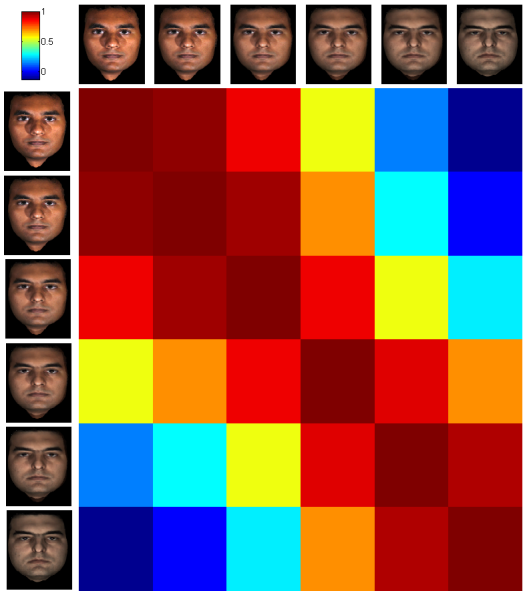


Fig. 7. Color coded similarity matrix for a pair of faces and their morphed versions from the cosine distance between respective feature vectors. Note that the similarity matrix is symmetric.

A. *Classifying similar and dissimilar faces*

Recall that we conducted a human study and have mean similarity scores (MSS) for pairs of faces. Also recall that we have 3 such sets of MSS scores for 3 subsets of the dataset spanning 60 faces each. Since the original rating was on a scale of 1-7, the MSS also falls within this range. Our goal now is to utilize the obtained MSS for pairs of faces, along with their respective features to classify pairs of faces as being similar or dissimilar. For this purpose, we first divide the obtained MSS ratings into similar-dissimilar classes by thresholding the MSS value, such that all pairs of faces with $MSS < 4-\theta$ are dissimilar while those with $MSS \geq 4+\theta$ are similar.

Given the ground truth similar-dissimilar classes, we compute the cosine similarity as before between pairs of points. We then use a generalized linear model [1] to map the cosine similarity scores onto either the 'similar' or 'dissimilar' class. This cosine similarity is computed for 3 cases: (1) Range features only, (2) Texture features only and (3) Range+Texture features. Since we have 3 disjoint sets, we train on two of the datasets and test the classifier on the third. For each test-set we report the mean and standard deviation of classification accuracy over 100 train-test trials[5].

In order to evaluate the effect of the dead-band obtained due to thresholding the MSS scores, in Fig. 8 we plot the classification accuracy and the area under the ROC curve (AUC) of the above described classifier as a function of θ for each of the three cases (range only, texture only, range+texture) described above. A larger value of θ indicates a larger deadband, and one would expect that the classification accuracy as well as the AUC would increase with increase in θ as seen.

[5]Since the class distributions are unequal - there are $\sim$ 10 times as many dissimilar samples as there are similar ones - we randomly sample the dissimilar class samples so that the two classes have identical number of points. Thus, the logistic fitting tool produces varying levels of classification accuracy on each trial and hence we report mean and standard deviation across multiple trials.

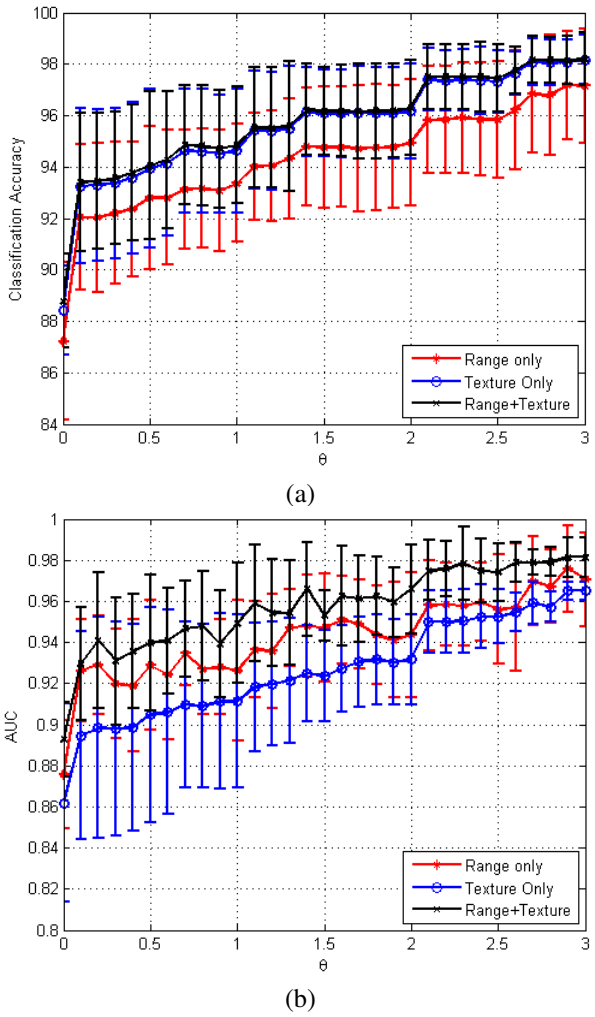


Fig. 8. (a) Classification accuracy, (b) area under ROC curve (AUC) with standard error bars across 100 train-test trials. Plotted are the mean values across these 100 trials and across the three test-sets as well as the standard error bars.

In Table I we report the mean and standard deviation of the classification accuracy across 100 train-test trials with $\theta = 2$, for the three cases (range only, texture only, range+texture). Based on the results from Table I, one would conjecture that texture information is probably more important for gauging facial similarity than range information, but only marginally so. Also, combining the two features together does not seem to harm performance, but does not boost performance substantially either.

As Table I demonstrates, the simple cosine measure of similarity between Gabor features achieves high classification accuracy when similarity is gauged by human observers. This implies that we can model perceptual similarity with high levels of accuracy using the features extracted here.

Given that these features are capable of classifying similar

	Range only		Texture only		Range+Texture	
Test Set	Mean acc.(%)	std. dev. (%)	Mean acc.(%)	std. dev. (%)	Mean acc.(%)	std. dev. (%)
Set 3	96.9048	0.5035	97.8571	0.7192	97.8571	0.7781
Set 2	92.0849	0.9337	96.3320	0.4555	96.5251	0.5191
Set 1	95.0570	0.4630	94.2966	0.1715	94.1065	0.2272
Overall	94.6822	0.6334	96.1619	0.4487	96.1629	0.5081

TABLE I

MEAN AND STD. DEV. OF CLASSIFICATION ACCURACY OVER 100 TRAIN-TEST TRIALS FOR EACH TEST SET; FOR RANGE ONLY, TEXTURE ONLY AND RANGE+TEXTURE FEATURES. ALSO REPORTED IS THE AVERAGE MEAN PERFORMANCE AND THE AVERAGE STANDARD DEVIATION.

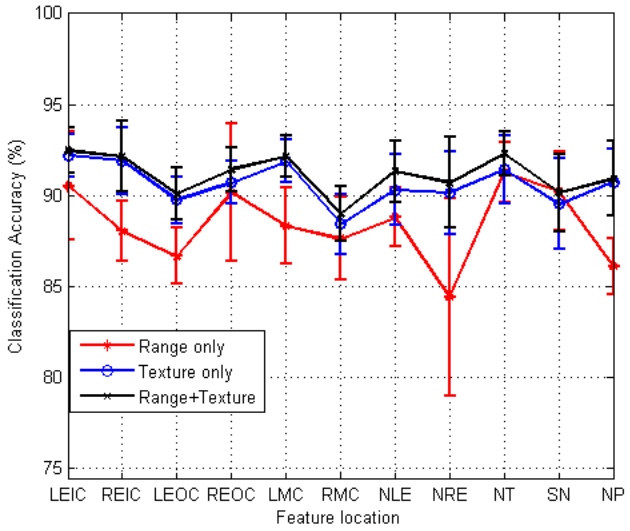


Fig. 9. Figure demonstrating the performance of individual Gabor responses at each of the fiducial points. Labels correspond to the labels in Fig. 6. Shown here are the median classification accuracy and the standard error bars for features from the 11 fiducial points for 3 cases: range only, texture only and range+texture.

and dissimilar faces, it is of interest to study how humans gauge similarity. One can indirectly infer which regions on the face are most important when gauging similarity by evaluating how features from a particular fiducial point affect prediction accuracy. In Fig. 9, we plot the mean accuracy and the standard error bars for Gabor features extracted at each of the 11 fiducial points separately for the range-only, texture-only and range+texture cases. Reported results are the mean and standard deviation for 100 train-test trials as described, across the three train-test combinations possible.

From the figure, one would conjecture that features extracted from regions around the eyes are seemingly more important than those from other regions in predicting accuracy. Surprisingly enough, features around the left-mouth-corner matter more than those from the right-mouth-corner and similarly for the nose. One could conjecture that humans scan the face from left-right and possibly assume facial symmetry when making a decision - hence features on the left of the face are seemingly more important than those on the right.

B. Using similarity scores for scalable face recognition

As described in the introduction, one of our main motivations for estimating similar faces was to enable faster face recognition. Here we describe a simple embodiment of the general principle. Our setup is as follows.

We have a large dataset of faces (180 faces from the 3 sets described above). The probe face is a random face from this dataset and cosine similarity (between the range+texture features) is used for face recognition. Note that in this setup the accuracy of face recognition is always 100%, since the probe face vector is exactly the same as a face from the gallery. Our goal here is to increase efficiency of search and hence the setup has been designed so that accuracy is set at 100%. Without the use of face similarity, face recognition would need a linear scan through the dataset to find the feature which correlates the best with that of the probe face.

In our demonstration of scalable face recognition, we first cluster the faces in the dataset into k clusters. Such clustering is performed on the Gabor feature vectors for the faces using k-means. The intuition behind such clustering is that similar faces will have similar feature vectors and hence will be clustered together. Within each cluster, we pick a representative face for that cluster. This face is the one represented by the centroid of the cluster.

Thus, we have k clusters and k representative faces - one for each cluster. Given a probe face, we first compare the probe face to each of the k representative faces. Given the similarity ratings from such a comparison, we only search through the cluster whose representative face is maximally similar to the probe face to find the match.

The reader will notice that the complexity of such a search is a function of k and hence, we explore various values of k. In Fig. 10 we plot the mean time taken to recognize a face as a function of k. The time taken for a full-search (corresponding to k =1) is also plotted in Fig. 10 for comparison purposes. Note that we plot the time taken for face recognition using the similarity approach as a fraction of the time taken for full-search so as to make the results independent of any implementation or processor used. From the figure one would conjecture that with approximately 17-18 clusters for 180 faces (i.e., $1/10^{th}$ the number of samples), a 10-fold increase in efficiency may be achieved over a general full-search.

In order to demonstrate that the gains achieved in computational performance is not dependent on the presence of artificially simulated 'similar' faces, we undertake the same experiment as described above - clustering faces before face recognition - for the entire Texas 3D FR database. The experimental setup is the same as described above, and in

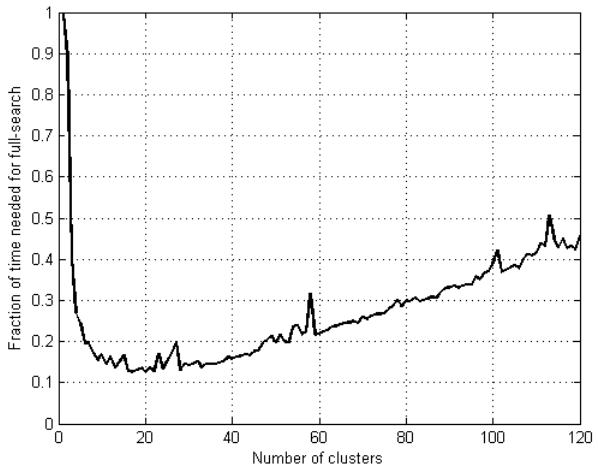


Fig. 10. Mean time taken to recognize a face as a function of k. The time taken for a full-search corresponds to k = 1. Note that we plot the time take for face recognition using the similarity approach as a fraction of the time taken for full-search (corresponding to k = 1) so as to make the results independent of any implementation or processor used.

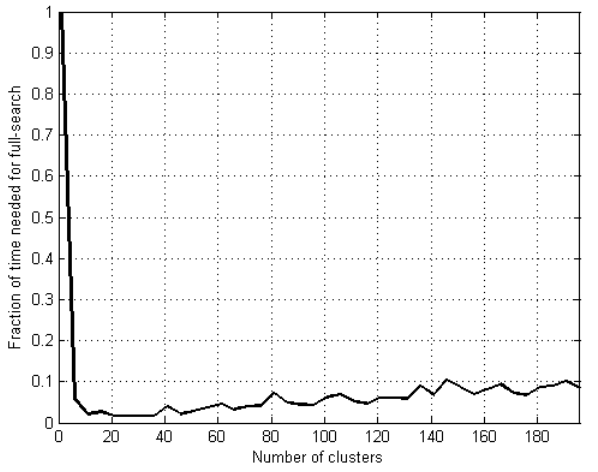


Fig. 11. Mean time taken to recognize a face as a function of k for the entire Texas 3D FR dataset. The time taken for a full-search corresponds to k = 1. Note that we plot the time take for face recognition using the similarity approach as a fraction of the time taken for full-search (corresponding to k = 1) so as to make the results independent of any implementation or processor used.

Fig. 11 we plot the mean time taken to recognize a face as a function of k. The time taken for a full-search (corresponding to k =1) is also plotted in Fig. 11. As can be seen, a 15-fold increase in computational accuracy is obtained with approximately 20-25 clusters for 1196 faces.

V. Conclusion

We proposed an algorithm to assess the similarity of pairs of 3D faces given the portrait and range images. In order to obtain perceptual similarity scores, we created a database of similar faces using a subset of the Texas 3D Face Recognition database and pairwise face morphs. A subjective study was undertaken to obtain human ratings of pairwise similarity. We demonstrated that the proposed algorithm, based on automatically detecting fiducial points and extracting Gabor features at fiducial points, produces features that correlate well with human perception of similarity. We explored the importance of regions on the face for predicting facial similarity. Further, we demonstrated how face similarity measures can be used to enable faster face recognition on larger datasets. Based on our results a 10-fold increase in efficiency may be achieved by using a small number of clusters.

Future work would involve extending the feature set to include geodesic distances [13], as well as exploring other similarity measures. Applications of such similarity measures for gauging effectiveness of facial reconstruction remains attractive.

References

[1] C. M. Bishop. *Pattern recognition and machine learning*. Springer New York:, 2006.

[2] A. C. Bovik, M. Clark, and W. S. Geisler. Multichannel texture analysis using localized spatial filters. *IEEE Trans. Pat. Ana. Mach. Intel.*, 12:55–73, 1990.

[3] M. Clark, A. C. Bovik, and W. S. Geisler. Texture segmentation using Gabor modulation/demodulation. *Pattern Recognition Letters*, 6(4):261–267, 1987.

[4] J. G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am. A*, 2(7):1160–1169, 1985.

[5] J. Davis, B. Kulis, P. Jain, S. Sra, and I. Dhillon. Information-theoretic metric learning. *Proc. of Intl. Conf. Mach. Learn.*, page 216, 2007.

[6] S. Gupta, K. R. Castleman, M. K. Markey, and A. C. Bovik. Texas 3d face recognition database. *IEEE Southwest Symp. on Img. Analy. and Interp.*, (to appear), 2010.

[7] S. Gupta, M. K. Markey, and A. C. Bovik. Advances and challenges in 3D and 2D+3D human face recognition. *Pattern recognition in biology*, page 63, 2007.

[8] J. P. Havlicek, D. S. Harding, and A. C. Bovik. Multidimensional quasi-eigenfunction approximations andmulticomponent AM-FM models. *IEEE Tran. Image Proc.*, 9(2):227–242, 2000.

[9] V. Hinsz. Facial resemblance in engaged and married couples. *Journal of Social and Personal Relationships*, 6(2):223–229, 1989.

[10] A. Holub, Y. Liu, and P. Perona. On constructing facial similarity maps. *IEEE Conf. Comp. Vis. Pat. Recog.*, pages 1–8, 2007.

[11] International Telecommunications Union. BT 500-11: Methodology for the subjective assessment of the quality of television pictures.

[12] S. Jahanbin, H. Choi, R. Jahanbin, and A. C. Bovik. Automated facial feature detection and face recognition using Gabor features on range and portrait images. In *IEEE Intl. Conf. on Image proc.*, pages 2768–2771, 2008.

[13] S. Jahanbin, H. Choi, Y. Liu, and A. C. Bovik. Three Dimensional Face Recognition Using Iso-Geodesic and Iso-Depth Curves. In *IEEE Intl. Conf. Biometrics: Theory, Appl. and Sys.*, pages 1–6, 2008.

[14] N. H. Kim, A. C. Bovik, and S. J. Aggarwal. Shape description of biological objects via stereo light microscopy. *IEEE Tran. Sys., Man and Cyber.*, 20(2):475–489, 1990.

[15] Y. Lee and T. Yi. 3D face recognition using multiple features for local depth information. In *EURASIP Conf. on. Video/Image Proc. Mult. Comm.*, volume 1, pages 429–434, 2-5 July 2003.

[16] S. Mullens and S. Notley. Image morphing. *http://www.stephenmullens.co.uk/image_morphing/*.

[17] G. Pan, S. Han, Z. Wu, and Y. Wang. 3d face recognition using mapped depth images. In *IEEE Intl. Conf. Comp. Vis.*, volume 3, pages 175–175, 20-26 June 2005.

[18] J. Rasmussen. Analysis of Likert-scale data: A reinterpretation of Gregoire and Driver. *Psychological Bulletin*, 105(1):167–170, 1989.

[19] C. Tredoux. A direct measure of facial similarity and its relation to human similarity perceptions. *Journal of Experimental Psychology: Applied*, 8(3):180–193, 2002.

[20] I. Voak, D. Perrett, and J. Peirce. Computer graphic studies of the role of facial similarity in judgements of attractiveness. *Current Psychology*, 18(1):104–117, 1999.

[21] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. Face recognition: a literature survey. *ACM Computing Surveys*, 35(4):399–459, 2003.