FACIAL SIMILARITY ACROSS AGE,DISGUISE,ILLUMINATION AND POSE

Narayanan Ramanathan, Rama Chellappa

Center for Automation Research and
Dept. of Electrical & Computer Engineering
University of Maryland, College Park, MD 20742

*Amit K. Roy Chowdhury**

Dept. of Electrical Engineering
University of California,
Riverside, CA 92521

ABSTRACT

Illumination, pose variations, disguises, aging effects and expression variations are some of the key factors that affect the performance of face recognition systems. Face recognition systems have always been studied from a recognition perspective. Our emphasis is on deriving a measure of similarity between faces. The similarity measure provides insights into the role each of the above mentioned variations play in affecting the performance of face recognition systems. In the process of computing the similarity measure between faces, we suggest a framework to compensate for pose variations and introduce the notion of 'Half-faces' to circumvent the problem of non-uniform illumination. We used the similarity measure to retrieve similar faces from a database containing multiple images of individuals. Moreover, we devised experiments to study the effect age plays in affecting facial similarity. In conclusion, the similarity measure helps in studying the significance facial features play in affecting the performance of face recognition systems.

1. INTRODUCTION

Face recognition has been an active area of research in computer vision and psychophysics, over the past decade. Unlike other biometric person-identification methods such as fingerprint analysis, retinal or iris scans, face recognition systems do not rely heavily on the co-operation of the participants. Applications of face recognition systems range from comparison of mug-shot images of frontal faces to comparison of faces obtained through surveillance video images. Recent improvements in the performance of still-image based and video-based face recognition systems, coupled with the availability of standardized performance evaluation methodologies[1, 2], have resulted in face recognition systems gaining commercial significance.

From a recognition perspective, still-image based face recognition systems can be classified into three categories - systems that take into account holistic facial features [3, 4, 5], systems that associate significance to local features (fiducial points) on faces [6] and systems that consider both holistic and local features of a face. A qualitative analysis of the above mentioned face recognition systems is provided at [2]. Eigenfaces [3], Fischerfaces [5], Subspace LDA (Linear Discriminant Analysis) [4] and Elastic Bunch Graph Matching (EBGM) [6] are some of the well known face recognition algorithms. The Eigenfaces method performs recognition by linearly projecting the image space onto a low dimensional feature space that spans the significant variations among known face images. The significant features are nothing but the eigenvectors of the set of faces.The Eigenfaces method yields projection directions that maximize the total scatter across all classes,i.e.,across all images of all faces. Thus unwanted variations due to illumination changes are retained. The Fischerfaces method selects the projection direction by maximizing the ratio of the inter-class scatter matrix to the intra-class scatter matrix. Thus the Fischerfaces method is better equipped to handle illumination variations than the Eigenfaces method. The Subspace LDA method projects face images from the original vector space to a face subspace using Principal Component Analysis and then uses Linear Discriminant Analysis to obtain a linear classifier in the subspace. Subspace LDA and Fischerfaces [4, 5] report better recognition results than the Eigenfaces method. However both the Subspace LDA method and the Fisherfaces method need more training images per subject unlike the Eigenfaces method. Pose variations affect the performance of all the three algorithms. EBGM [6] method is a simplified implementation of dynamic link architecture methods based on a neural network and a geometric measure. In the EBGM method faces are represented in the form of labeled graphs. Edges are labeled with distance information and nodes are labeled with wavelet responses locally bundled in jets. The model graphs are translated,scaled or deformed to perform matching, thus accounting for a large part of the variance of the images. EBGM is robust to moderate variations in illumination conditions. But the EBGM method is computationally intensive.

1.1. Problem Statement

Some of the key problems that affect the performance of face recognition systems are illumination changes, pose variations, aging effects, disguises and expression variations. Face recognition systems have been deployed to identify one or more individuals from a database of faces. The effectiveness of a face recognition algorithm is measured in terms of the number of correct matches. Usually, the database (gallery) consists of uniformly illuminated, non-disguised frontal face images of subjects with neutral expressions and the test images carry one of the above mentioned variations. The emphasis has always been on correctly matching a test image with its corresponding gallery image. Our emphasis is rather on computing a similarity measure between different face images in a gallery, as in an indexing problem.

The database for the first experiment comprises of multiple images of each individual, shot under different conditions. When presented with a test image, the similarity measure will be used to retrieve similar faces from the database. Ideally, the most similar

*The author performed the work at University of Maryland, College Park.

The authors would like to thank Todd James and George Steinmetz of National Geographic for allowing us to use their image database in our paper.

images retrieved from the database should be the images of that individual. But due to variations such as illumination, disguises etc., images of other individuals might be retrieved as more similar images. Thus the results of this experiment would assist in studying the significance of facial features over the above mentioned varying conditions in affecting facial similarity. The database for the second experiment comprises of images of an individual, spanning a number of variations. Illumination conditions vary drastically within the database. The database comprises of frontal views and profile views. Some of the images in the database contain disguises such as beard, mustache, hat etc thereby resulting in occlusion of facial features. Expression variations are notably significant in the database. More significantly, the database comprises of the individual's images that were taken many years apart. Thus, studying the similarity measure across images in this database shall help us analyze the effect age plays in affecting facial similarity, apart from the other factors that were explored in the previous experiment.

We used the images from the AR Face database [7] for the first experiment. This database contains frontal images. The conditions that interested us were - neutral expressions, neutral expressions with different illuminations on each half of the face, dark glasses, dark glasses with different illuminations on each half of the face and expressions such as smile under uniform illumination. For the second experiment, we acquired a series of images of an individual that differ as described above, from National Geographic. Some of our results on the second experiment were published in National Geographic's November 2003 issue [8].

2. BASIC FRAMEWORK

We suggest the following framework to compute the similarity measure. When we are presented with a non-frontal image we propose an idea to compensate for pose variation. We also introduce the notion of 'Half-Faces' to circumvent the problem of non-uniform illumination.

2.1. Pose Estimation

A contour-based pose estimation method proposed in [9] is used to estimate the pose of the test image. A generic 3D face model,texture mapped with an average human texture, is rotated about the azimuth angle. The edges of the 2D projections of the rotated 3D models are extracted. To estimate the pose of the test image, we extract the edge image of each of the test image and compute its disparity from the edge map of the 2D projections of the rotated 3D face models. The disparity between edge maps is computed using the Euclidean Distance Transform (DT). For each pixel in the binary edge map of the test image, the distance transform assigns the distance between that pixel and the nearest nonzero pixel of the edge map. The cost function,F, which is the measure of disparity between the 3D model edge map and the edges of the test image is of the form

$$F \triangleq \frac{1}{N} \sum_{i,j \epsilon A_{EM}} DT(i,j) \tag{1}$$

where $A_{EM} \equiv (i,j) : EM(i,j) = 1$ and N is the size of set A_{EM} (total number of nonzero pixels in the 3D model edge map EM). The pose of the test image is estimated to be the pose of the 3D face model, the edge map of which, minimizes the function F.

An equivalent frontal image is generated for those images where the estimated pose is beyond a threshold, indicating that it is non-

Fig. 1. (a)Top row : Profile views (b)Bottom row : Generated frontal views

frontal. We map a generic 3D mesh-model of a face onto the test image and texture map it with this image. By rotating the texture mapped 3D model and by using a mirror reflection of the half-face that had all the texture information, we generate the frontal image. Figure 1 shows instances where the above mentioned method was adopted to generate frontal images.

2.2. Half-Faces

Human faces can be considered as 3D objects that are symmetric about the longitudinal axis.The symmetry in human faces has been exploited in applications such as 3D face modeling [9]. Performance of face recognition systems reveal the significance of intrinsic features of a human face coupled with the outer contour of the face in enhancing the variability between human faces. The loss of either of the features affects recognition. Facial symmetry suggests that the intrinsic facial features and the outer contour captured in one half of a human face is sufficient for recognition. Thus, symmetry of a human face suggests that under uniform illumination conditions, performance of a face recognition system that uses only one half of the human face will be comparable to its performance while using the full face. When a face recognition system is presented with a non-uniformly illuminated human face, usually one half of the face is well illuminated. Thus the recognition results stand to gain if recognition is performed using the better illuminated half of the face instead of the complete face that is non-uniformly illuminated. Thus the notion of 'Half-Faces' can be employed to circumvent the problem of non-uniform illumination on faces.

What is the criteria for picking the best half of a frontal face? Distinctiveness of facial features depends a great deal on the nature of illumination. Facial features are less distinctive both in the case of poor illumination and in the case of illumination saturation. We need to pick that half of the face where illumination is neither poor nor saturated. We compute the difference in the average pixel intensities on each half of the face. If the difference is beyond a threshold we classify the image as a non-uniformly illuminated image. To pick the better half-face we analyze the edge densities on each half-face. The edge density is higher on that half-face where illumination is bright and uniform. The edge density is lower if the half-face is poorly illuminated or if the half-face is saturated with illumination. Figure 2 illustrates the above idea. In the first

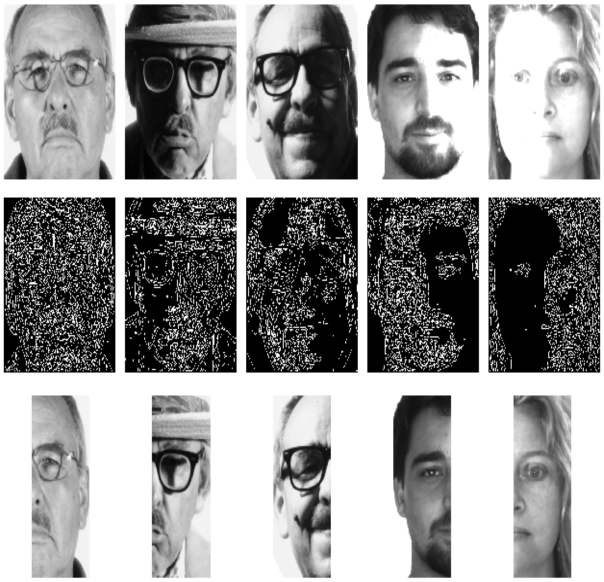

Fig. 2. (a) First row : Test Images (b) Second row : Edge Densities (c) Third row : Picked Half-faces

image illumination is uniform. The edge densities are prominent on either half of the face. Thus both the half-faces are equally informative. In the second and the fourth image, the left half (from the viewing direction of the subject) of the face has higher illumination than the right half. But the edge densities reveal that the illumination on the left half of the second face and the right half of the fourth face are optimal. Similarly, in the third and the fifth image, the right half of the face has higher illumination than the left half. Again, the edge densities indicate that the right half of the third face and the left half of the fifth face have optimal illumination. The third row of images indicate the half-faces that are picked. This shows that the method does not always pick the more brightly illuminated half of the face, but is able to mitigate the effect of saturation (the right and the left half of the faces were not selected for the fourth and the fifth image)

2.3. Similarity Measure

To compute the similarity measure between faces we adopt the eigenfaces framework and incorporate the notion of 'half-faces' to circumvent non-uniform illumination. As in the Eigenfaces framework we have a gallery of face images. We create two eigenspaces - one using the left-half of the gallery images and the other using the right half. Given a test image, we detect the optimally illuminated half-face as explained in the previous section. We project the half-face on the appropriate eigenspace. We define the similarity measure between the test image and the images in the gallery as the Cosine Mahalanobis distance [10] between the projection of the test image and the projections of the gallery images.

Let Γ_1,Γ_2,Γ_3,...,Γ_N be the vector of half-faces picked from the gallery images. Let $\Theta = \frac{1}{N}\sum_{i=1}^{N}\Gamma_i$ be the average half-face. Let $\Phi_i = \Gamma_i - \Theta$ be the mean subtracted half faces. Let the data matrix A be defined as $A = [\Phi_1\Phi_2.....\Phi_N]$. Eigenfaces are nothing but the eigenvectors of AA^T. The eigenvectors of A^TA can be computed as $A^TAv_i = \mu_i v_i$. Pre-multiplying both sides by A, we have $AA^TAv_i = \mu_i Av_i$. Thus Av_i are the eigenvectors of AA^T. To compute the projection of a half-faced image onto the above space, the average half face is subtracted from the test image. If w_i is the projection of the mean subtracted half face on the i^{th} eigenface, then the projection co-efficients of the half face are $u = [w_1, w_2, ...w_N]$. We use the Cosine-Mahalanobis distance to measure the similarity between projection co-efficients. The use of the Cosine Mahalanobis distance is motivated by the results in [10]

The Eigenfaces span the image space. The eigenvalues correspond to the variance along each Eigenface. We need to understand the transformation between the image space and the Mahalanobis space before computing the Cosine Mahalanobis distance. The Mahalanobis space has unit variance along each dimension. Let u and v be two vectors in the Eigenspace. Let $\mu_i = \sigma_i^2$ be the variance along the i^{th} dimension. Let m and n be the corresponding vectors in the Mahalanobis space. The relationship between the vectors is defined as:

$$m_i = \frac{u_i}{\sigma_i} \qquad n_i = \frac{v_i}{\sigma_i}. \tag{2}$$

Mahalanobis cosine is the cosine of the angle between the projection of the images on the Mahalanobis space. So, the Cosine Mahalanobis distance between u and v is computed in terms of m and n.

$$D_{MahCosine(u,v)} = \cos(\theta_{mn}) = \frac{mn}{|m||n|} \tag{3}$$

3. EXPERIMENTAL RESULTS

3.1. Experiment 1

The objective of this experiment was retrieval of similar faces from a database that comprises of multiple images of every individual. The images from the AR Face database [7] were used to compile the gallery and the test set. The gallery comprises of images that belong to the categories neutral-uniform illumination, neutral-left illumination and neutral-right illumination. The test set was compiled from the images under the categories such as dark-glasses, dark-glasses with left illumination, dark-glasses with right illumination and smile. Detection of the optimally illuminated half-face revealed that the right-half face and the left-half face were more informative for the images under the categories neutral-left illumination and neutral-right illumination, respectively. A left eigenspace was created using the images from the categories neutral-uniform illumination and neutral-right illumination and a right eigenspace was created using images from the category neutral-uniform illumination and neutral-left illumination. The images from the test set were projected on the appropriate eigenspace. Similarity scores, computed as explained in the previous section, were used for retrieval of similar faces. Ideally, the top two retrievals for a test image should be the images of the same individual. The face retrieval results were as follows.

3.2. Experiment 2

The objective of the second experiment was the computation of a similarity measure between images of the same individual taken under varying conditions. The images of an individual, provided

Table 1. Accuracy(%) in top N retrieved faces in Experiment 1

Condition	Top 2	Top 5	Top 10
Smile	39	59	66
Glasses	20	37	51
Glasses Left Light	21	37	55
Glasses Right Light	22	40	60

by National Geographic, were used for the experiment. A frontal well illuminated neutral image of the individual was added to a gallery of frontal well illuminated images. The left and the right eigenspaces were created. The similarity scores of all the test images with respect to the image that was added to the gallery was computed as discussed in the previous sections. The result of this experiment is shown in 3. The value below each image indicates the similarity score calculated using the Cosine Manalanobis distance. The effects of various parameters are seen in the gradually decreasing similarity scores. The ordering of the similarity scores is close to what could be expected be expected of a human observer. The effects of disguise and aging are prominent. The similarity measures are different from the ones published in the National Geographic issue [**?**] because of the use of the Cosine Mahalanobis distance instead of the Euclidean distance. However, we believe that the main conclusions of the study do not change because of this.

Fig. 3. Similarity Measure in Experiment 2

4. CONCLUSIONS AND FUTURE WORK

Our work addresses the issue of understanding facial similarity under varying conditions. Illumination, disguise and aging play a significant role in computing facial similarity. Retrieval of faces from a database is more effective when half-faces with optimal illumination are used, rather than full-faces with non-uniform illumination. Illumination variations affected facial similarity more than aging effects. Pose variations can be taken care of by generating the frontal view using the notion of facial symmetry. Expression variations do not affect facial similarity as much as occlusions or disguises do.

In future we wish to study facial similarity under arbitrary illumination conditions. We wish to address the issue of compensation for age and disguises. We wish to bring in a notion of familiarity in our algorithm for the retrieval of similar faces from a database.

5. REFERENCES

[1] P.J. Phillips, H. Moon, S.A. Rizvi, and P.J. Rauss, “The feret evalutation methodology for face-recognition algorithms,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22(10), pp. 1090–1104, October 2000.

[2] W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld, “Face recognition : A literature survey,” *ACM Computing Surveys*, vol. 35(4), pp. 399–458, December 2003.

[3] M. Turk and A. Pentland, “Eigenfaces for recognition,” *Journal of Cognitive Neuroscience*, vol. 3, pp. 72–86, 1991.

[4] W. Zhao, R. Chellappa, and P.J. Philips, “Subspace linear discriminant analysis for face recogniton,” Tech. Rep. CAR-TR-914, Centre for Automation Research, University of Maryland, College Park, 1999.

[5] Peter N. Belhumeur, P. Hespanha, and David J. Kriegman, “Eigenfaces vs fischerfaces : Recognition using class specific linear projection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19(7), July 1997.

[6] Laurenz Wiskott, J.M. Fellous, N. Kruger, and C. Malsburg, “Face recognition by elastic graph matching,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 775–779, 1997.

[7] A.M. Martinez and R. Benavente, “The AR face database CVC technical report,” Tech. Rep., june 1998.

[8] Roy Chowdhury A., R. Chellappa, N. Ramanathan, and N. Koterba, “About face,” in *National Geographic*, November 2003, pp. 18–19.

[9] Himaanshu Gupta, A. Roy Chowdhury, and R. Chellappa, “Contour-based 3d face modelling from monocular video,” *International Conference on Pattern Recognition*, submitted.

[10] Ross Beveridge, D. Bolme, Marico Teixeira, and B. Draper, “The CSU face indentification evaluation system user’s guide: Version 5,” Tech. Rep., Colorado State University, may 2003.