Face Sketch Recognition

Xiaoou Tang, *Senior Member, IEEE,* and Xiaogang Wang, *Student Member, IEEE*

***Abstract*—Automatic retrieval of face images from police mug-shot databases is critically important for law enforcement agencies. It can effectively help investigators to locate or narrow down potential suspects. However, in many cases, the photo image of a suspect is not available and the best substitute is often a sketch drawing based on the recollection of an eyewitness. In this paper, we present a novel photo retrieval system using face sketches. By transforming a photo image into a sketch, we reduce the difference between photo and sketch significantly, thus allowing effective matching between the two. Experiments over a data set containing 188 people clearly demonstrate the efficacy of the algorithm.**

***Index Terms*—Eigenface, face recognition, face sketch synthesis, sketch recognition.**

I. Introduction

DUE to growing demands in such application areas as law enforcement, video surveillance, banking, and security system access authentication, automatic face recognition has attracted great attention in recent years. The advantages of facial identification over alternative methods, such as fingerprint identification, are based primarily on the fact that face recognition does not require those being checked to cooperate. In addition, face recognition systems are more convenient to use and are more cost-effective, since recognition results can be corrected in uncertain cases by people without extensive training.

An important application of face recognition is to assist law enforcement. For example, automatic retrieval of photos of suspects from police mug-shot databases can help police narrow down potential suspects quickly. However, in most cases, the photo image of a suspect is not available. The best substitute available is often an artist drawing based on the recollection of an eyewitness. Automatic searching of a photo database using a sketch drawing is potentially very useful. It will not only help the police to locate a group of potential suspects, but may also help the witness and the artist to modify the sketch drawing of the suspect interactively based on similar images retrieved.

Despite the great need of such an automatic sketch-based photo retrieval system, few effective systems can be found in previous research, probably due to the difficulties in constructing a face sketch database. In psychology study, researchers have long been using various face drawings, especially the line drawings of faces, to investigate face recognition by the human visual system [1]–[3], [7], [13]. Human beings can easily recognize caricature, which is a special kind of line drawings of human faces, with particular details of a face accentuated. Presumably, the details which get accentuated in caricaturing are those which are characteristic of that individual. Some even question whether caricatures are in any way better representations than natural images, since the caricature may contain not only the essential minimum of information but also some kind of "super-fidelity" due to the accentuated structures [1]. Bruce *et al.* [2] have also shown that computer-drawn "cartoons" with edge, pigmentation, and shading of the original image can be well recognized by human beings.

It is of great interest to investigate whether automatic recognition of sketches using computers can achieve similar performance as human beings. Toward this purpose, we recently construct a database of face photos and sketches of 188 people [15]. Some examples are shown in Fig. 1. The size of this database is comparable to those of many conventional photo-based face recognition studies [5]. Using such a database, we develop a novel photo-to-sketch transformation method for face sketch recognition. The method is shown to be much more effective than directly using conventional methods such as geometrical measures and the eigenface method. We also compare the performance of the new algorithm with the sketch recognition performance of human beings and find that the new method outperforms human beings fairly consistently.

II. Geometrical Measures and Direct Eigenface Method

A. Geometrical Measures

The geometrical feature method is intuitively the most straightforward method. A great amount of geometrical face recognition researches focus on extracting relative positions and other parameters of face components such as eyes, mouth, and chin [4]–[6], [14], [19]. Although the geometrical features are easy to understand, they do not seem to contain enough information for stable face recognition. In particular, geometrical features change with different facial expressions and scales, thus vary greatly for different images of the same person. A comparison between geometric features and template features greatly favors the template features [4].

For face sketch recognition, it is intuitively reasonable to use geometrical measures as features since, for a mug-shot matching application, there are no expression or scale changes to distort the face geometrical structure. A photo and its corresponding sketch are expected to be relatively similar to each other in terms of geometrical measures if they were to look alike. The major difference between a photo and a sketch is the grayscale texture due to the drastically different image generation process. Using the geometrical measures as recognition features, we can avoid such texture difference.

Manuscript received November 17, 2002; revised June 8, 2003. This work was supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region under Project CUHK 4190/01E, Project CUHK 4224/03E, and Project AOE/E-01/99).

The authors are with the Department of Information Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong (e-mail: xtang@ie.cuhk.edu.hk; xgwang1@ie.cuhk.edu.hk).

Digital Object Identifier 10.1109/TCSVT.2003.818353

Fig. 1. Sample face photos (top two rows) and sketches (bottom two rows).

In order to compare geometrical features extracted under ideal conditions, we design a fiducial grid model. We defined 35 fiducial points over a face image, as shown in Fig. 2. Our definition differs from that of previous researches. We first define a set of anchor points over salient locations such as nose tip and mouth corners on the face image, so that they can be

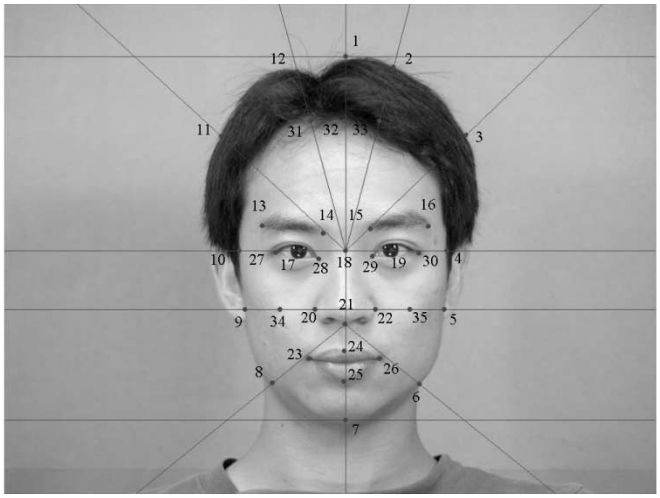

Fig. 2. Fiducial graph model.

located easily and accurately. Then other fiducial points can be derived from the anchor points. For example, in Fig. 2, point 8 is found through intersection of the face contour with a line anchored by the nose tip (point 21) and the mid-point of points 7 and 9. Because of such a definition, all points can be precisely located for different faces, and the drawing process is very fast even by manual dragging of points along anchor lines. Accurate geometrical measures can then be computed from the fiducial points.

B. Eigenface Method

The eigenface method is a classic face recognition method [9], [16], [17], [20]. It has been extensively tested in the FERET test [11], [12]. Even though the eigenface method is sensitive to illumination, expression, and rotation changes, it is not an important concern for our application given our focus on mug-shot photo identification.

The eigenface approach uses the Karhunen–Loeve Transform (KLT) for the representation and recognition of face images. Once a set of eigenvectors, also called eigenfaces, is computed from the ensemble face covariance matrix, a face image can be approximately reconstructed using a weighted combination of the eigenfaces. The weights that characterize the expansion of the given image in terms of eigenfaces constitute the feature vector. When a new test image is given, the weights are computed by projecting the image onto the eigenface vectors. The classification is then carried out by comparing the distances between the weight vectors of the test image and the images from the database.

To compute the KLT, let $\vec{Q}_i$ be a column vector representation of a sample face image with the mean face computed as $\vec{m}_p = 1/M \sum_{i=1}^{M} \vec{Q}_i$, where M is the number of training samples. Removing the mean face from each image, we have $\vec{P}_i = \vec{Q}_i - \vec{m}_p$. The photo training set then forms an N by M matrix $A_p = \left[\vec{P}_1, \vec{P}_2, \ldots, \vec{P}_M\right]$, where N is the total number of pixels in the image. The sample covariance matrix can be estimated by

$$W = A_p A_p^T. \tag{1}$$

Given the large size of a photo image, direct computation of the eigenvectors of W is not practical. The dominant eigenvector estimation method [8] is generally used. Because of the relatively small sample image number M, the rank of W is only $M-1$. So the eigenvector of the smaller matrix $A_p^T A_p$ can be computed first as

$$\left(A_p^T A_p\right) V_p = V_p \Lambda_p \tag{2}$$

where V_p is the eigenvector matrix and Λ_p is the diagonal eigenvalue matrix. Multiplying both sides by A_p, we have

$$\left(A_p A_p^T\right) A_p V_p = A_p V_p \Lambda_p. \tag{3}$$

Therefore, the orthonormal eigenvector matrix of the covariance matrix W is

$$U_p = A_p V_p \Lambda_p^{-1/2}. \tag{4}$$

For a new face photo $\vec{P}_k$, its projection coefficients in the eigenvector space form the vector $\vec{\mathbf{b}}_p = U_p^T \vec{P}_k$, which is used as a feature vector for the classification.

Because of the structural similarity across all face images, strong correlation exists among face images. Through the KLT, the eigenface method takes advantage of such a high correlation to produce a highly compressed representation of face images, thus improving the face classification efficiency.

However, because of the large difference between face photos and sketches, direct application of the eigenface method for sketch-based face identification may not work. The distance between a photo and a sketch of the same person is in general much larger than the distance between two photos of two different people. In order to overcome such a difference, we develop a photo-to-sketch transformation algorithm to convert a photo into a sketch first and then perform the classification using eigensketch features.

III. Sketch Transformation and Recognition

A. Photo-to-Sketch Transformation

For the conventional eigenface method, a face image can be reconstructed from the eigenfaces by

$$\vec{P}_r = U_p \vec{\mathbf{b}}_p. \tag{5}$$

Since $U_p = A_p V_p \Lambda_p^{-1/2}$, we can represent the reconstructed photo by

$$\vec{P}_r = A_p V_p \Lambda_p^{-(1/2)} \vec{\mathbf{b}}_p = A_p \vec{\mathbf{c}}_p \tag{6}$$

where $\vec{\mathbf{c}}_p = V_p \Lambda_p^{-(1/2)} \vec{\mathbf{b}}_p = [c_{p_1}, c_{p_2}, \ldots, c_{p_M}]^T$ is a column vector of dimension M. We can rewrite (6) in summation form as

$$\vec{P}_r = A_p \vec{\mathbf{c}}_p = \sum_{i=1}^{M} c_{p_i} \vec{P}_i. \tag{7}$$

This shows that the reconstructed photo is in fact the best approximation of the original image with the least mean-square error using an optimal linear combination of the M training

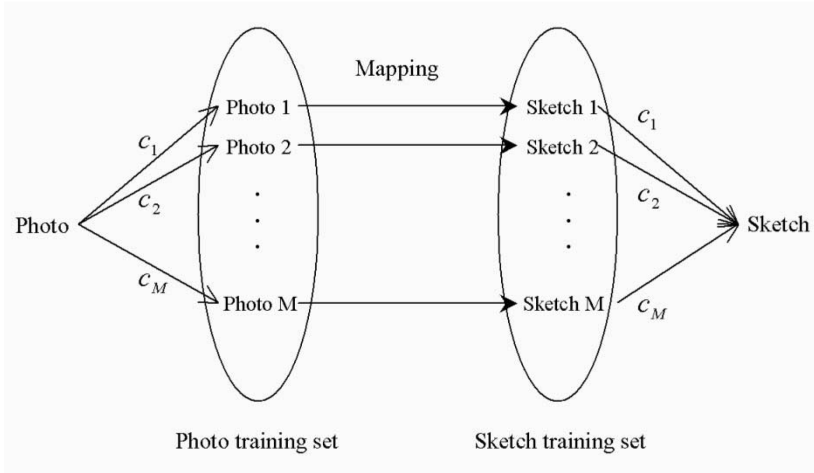

Fig. 3. Photo-to-sketch transformation.

sample images. The coefficients in $\vec{\mathbf{c}}_p$ describe the contribution weight of each sample image.

For each training photo image $\vec{P}_i$, there is a corresponding sketch $\vec{S}_i$, where $\vec{S}_i$ is a column vector representation of a sample sketch with the mean sketch $\vec{m}_s$ removed. Similar to $A_p = \left[\vec{P}_1, \vec{P}_2, \ldots, \vec{P}_M\right]$ for photo image training set, we have a corresponding sketch training set, $A_s = \left[\vec{S}_1, \vec{S}_2, \ldots, \vec{S}_M\right]$. If we map each sample photo image $\vec{P}_i$ in (7) to its corresponding sketch $\vec{S}_i$, as illustrated in Fig. 3, we obtain

$$\vec{S}_r = \sum_{i=1}^{M} c_{p_i} \vec{S}_i = A_s \vec{\mathbf{c}}_p = A_s V_p \Lambda_p^{-(1/2)} \vec{\mathbf{b}}_p. \tag{8}$$

Given the structural resemblance between photos and sketches, it is reasonable to expect the reconstructed sketch $\vec{S}_r$ to resemble the real sketch. For such a reconstruction, a sample sketch $\vec{S}_i$ contributes more weight to the reconstruction if its corresponding photo sample $\vec{P}_i$ contributes more weight to the reconstructed face photo. For an extreme example, if a reconstructed photo $\vec{P}_i$ has a unit weight $c_{p_k} = 1$ for a particular sample photo $\vec{P}_k$ and zero weights for all other sample photos, i.e., the reconstructed photo looks exactly like the sample photo $\vec{P}_k$, then the reconstructed sketch $\vec{S}_r$ is simply reconstructed by replacing it with the corresponding sketch $\vec{S}_k$. Through such a mapping, we successfully transform a photo image into a pseudo-sketch.

In summary, the photo-to-sketch transformation is computed through the following steps.

1) Compute the average photo image $\vec{m}_p$ for the photo training set, and the average sketch $\vec{m}_s$ for the sketch training set.
2) Compute the photo training set eigenvector matrix U_p by first computing the eigenvectors V_p of $A_p^T A_p$.
3) Remove the photo mean $\vec{m}_p$ from the input photo image $\vec{Q}_k$ to get $\vec{P}_k = \vec{Q}_k - \vec{m}_p$.
4) Project $\vec{P}_k$ in the eigenspace U_p to compute the eigenface weight vector $\vec{\mathbf{b}}_p$.
5) Reconstruct the pseudo-sketch by $\vec{S}_r = A_s V_p \Lambda_p^{-(1/2)} \vec{\mathbf{b}}_p$.
6) Finally, add back the average sketch $\vec{m}_s$ to get the final reconstructed sketch $\vec{T}_r = \vec{S}_r + \vec{m}_s$.

Fig. 4 shows the comparison between the real sketch and the reconstructed sketch. We can clearly see the similarity between the two.

B. Sketch Recognition

After such a photo-to-sketch transformation, sketch recognition becomes straightforward. We can compare the pseudo-sketch with the real sketch using the conventional eigenface method. In fact, we can use any other conventional methods including the elastic graph matching method for the recognition [10], [18]. We first compute the eigenvectors using the sketch training samples. Then the probe sketch and the generated pseudo-sketches from the photo gallery are projected onto the eigensketch vectors. The projection coefficients are then used as feature vectors for final classification. The detail algorithm can be summarized as follows.

1) Compute the photo eigenspace U_p using the photo training set A_p.
2) Compute the sketch eigenspace U_s using the sketch training set A_s.
3) Use U_p to compute the pseudo-sketch $\vec{S}_r$ for each photo $\vec{P}_i$ in the photo gallery by the sketch transformation algorithm described in Section III-A
4) Compute the eigensketch weight vector $\vec{\mathbf{b}}_r = U_s^T \vec{S}_r$ for each pseudo-sketch $\vec{S}_r$ by projecting $\vec{S}_r$ in the sketch eigenspace U_s.
5) Compute the eigensketch weight vector $\vec{\mathbf{b}}_s = U_s^T \vec{S}_k$ for the probe sketch $\vec{S}_k$ by projecting $\vec{S}_k$ in the sketch eigenspace U_s.
6) Compute the distance between $\vec{\mathbf{b}}_s$ and each $\vec{\mathbf{b}}_r$ generated from the photo gallery, the sketch is classified as the face with minimum distance between the two vectors.

In the algorithm, the photos in the galley are first transformed to pesudo-sketches based on the photo eigenspace. Then the

Fig. 4. Photo-to-sketch transformation examples. (a) Original photo. (b) Reconstructed photo. (c) Reconstructed sketch. (d) Original sketch.

recognition is conducted in the sketch eigenspace. Similarly, we can also transform each probe sketch into a pseudo-photo based on the sketch eigenspace, then use the photo eigenspace for recognition.

For both approaches, we rely on two sets of reconstruction coefficients $\vec{\mathbf{c}}_p$ and $\vec{\mathbf{c}}_s$, where $\vec{\mathbf{c}}_p$ represents the weights for reconstructing a photo using the photo training set and $\vec{\mathbf{c}}_s$ represents the weights for reconstructing a sketch using the sketch training set. In fact, to compare a photo with a sketch, we can also use their corresponding reconstruction coefficients $\vec{\mathbf{c}}_p$ and $\vec{\mathbf{c}}_s$ directly as feature vectors for recognition. To see the differences of the three approaches, we conduct the following analysis.

As shown in Section II, for an input photo, its reconstruction coefficient vector on the photo training set is $\vec{\mathbf{c}}_p = V_p \Lambda_p^{-(1/2)} \vec{\mathbf{b}}_p$, where $\vec{\mathbf{b}}_p$ is the projection weight vector of the photo in the photo eigenspace. Similarly, for an input sketch, its reconstruction coefficient vector on the sketch training set is $\vec{\mathbf{c}}_s = V_s \Lambda_s^{-(1/2)} \vec{\mathbf{b}}_s$, where $\vec{\mathbf{b}}_s$ is the projection weight vector of the input sketch in the sketch eigenspace.

If we compare photo with a sketch using $\vec{\mathbf{c}}_p$ and $\vec{\mathbf{c}}_s$ directly, the distance is defined as

$$d_1 = \|\vec{\mathbf{c}}_p - \vec{\mathbf{c}}_s\|. \tag{9}$$

If we first generate a pseudo-sketch for a photo, then calculate the distance in the sketch eigenspace, the distance is defined as $d_2 = \|\vec{\mathbf{b}}_r - \vec{\mathbf{b}}_s\|$, where $\vec{\mathbf{b}}_r$ is the weight vector of the generated pseudo-sketch projected in the sketch eigenspace, and $\vec{\mathbf{b}}_s$ is the weight vector of the real sketch projected in the sketch eigenspace. Given $\vec{\mathbf{b}}_r = U_s^T \vec{\mathbf{S}}_r, U_s = A_s V_s \Lambda_s^{-(1/2)}$ and $\vec{\mathbf{S}}_r = A_s \vec{\mathbf{c}}_p$, we can compute $\vec{\mathbf{b}}_r$ as

$$\vec{\mathbf{b}}_r = \Lambda_s^{-(1/2)} V_s^T A_s^T A_s \vec{\mathbf{c}}_p. \tag{10}$$

Since $V_s^T (A_s^T A_s) V_s = \Lambda_s$, we have

$$\vec{b}_r = \Lambda_s^{(1/2)} V_s^T \vec{\mathbf{c}}_p. \tag{11}$$

To compute $\vec{\mathbf{b}}_s$, we can use relation $\vec{\mathbf{c}}_s = V_s \Lambda_s^{-(1/2)} \vec{\mathbf{b}}_s$ to get $\vec{\mathbf{b}}_s = \Lambda_s^{(1/2)} V_s^T \vec{\mathbf{c}}_s$. Finally, the distance d_2 becomes,

$$\begin{aligned} d_2 &= \|\vec{\mathbf{b}}_r - \vec{\mathbf{b}}_s\| = \left\| \Lambda_s^{1/2} V_s^T \vec{\mathbf{c}}_p - \Lambda_s^{1/2} V_s^T \vec{\mathbf{c}}_s \right\| \\ &= \left\| \Lambda_s^{1/2} V_s^T (\vec{\mathbf{c}}_p - \vec{\mathbf{c}}_s) \right\|. \end{aligned} \tag{12}$$

Similarly, if we first generate a pseudo-photo for a sketch, then calculate the distance in the photo eigenspace, the distance d_3 can be obtained by

$$\begin{aligned} d_3 &= \|\vec{\mathbf{b}}_r - \vec{\mathbf{b}}_p\| = \left\| \Lambda_p^{1/2} V_p^T \vec{\mathbf{c}}_s - \Lambda_p^{1/2} V_p^T \vec{\mathbf{c}}_p \right\| \\ &= \left\| \Lambda_p^{1/2} V_p^T (\vec{\mathbf{c}}_p - \vec{\mathbf{c}}_s) \right\| \end{aligned} \tag{13}$$

where $\vec{\mathbf{b}}_r$ is the weight vector of the generated pseudo-photo projected in the photo eigenspace, and $\vec{\mathbf{b}}_p$ is the weight vector of the original photo. The distances for recognition are different for the three cases. We will compare their performances in the experiments.

IV. Experiments

In order to demonstrate the effectiveness of the new algorithm, we conduct a set of experiments to compare with the geometrical measures and the conventional eigenface method. A database containing 188 pairs of photo and sketch of 188 people is used for the experiment. We use 88 photo-sketch pairs as training data and the other 100 photo-sketch pairs for testing.

For the geometrical method, we use 26 measures of geometrical distances between key fiducial points shown in Fig. 2. They include the sizes of nose, eyes, mouth, eyebrows, face contour, and their relative positions. For the traditional eigenface method, we simply treat a probing sketch as a regular photo. We adopt the recognition test protocol used in the FERET test [12]. Thus, our gallery set consists of 100 face photos. The probe set consists of 100 face sketches. The cumulative match score is used to evaluate the performance of the algorithms. It measures the percentage of "the correct answer is in the top n matches," where n is called the rank.

TABLE I
CUMULATIVE MATCH SCORES FOR THE THREE METHODS

Rank	1	2	3	4	5	6	7	8	9	10
Geometry Method	30	37	45	48	53	59	62	66	67	70
Eigenface Method	31	43	48	55	61	63	65	65	67	67
Sketch Transform Method	71	78	81	84	88	90	94	94	95	96

TABLE II
CUMULATIVE MATCH SCORES USING THREE DIFFERENT DISTANCES

Rank	1	2	3	4	5	6	7	8	9	10
d_1	20	49	59	65	69	73	75	76	81	82
d_2	71	78	81	84	88	90	94	94	95	96
d_3	57	70	77	79	83	84	85	86	87	88

A. Comparison With Traditional Methods

Table I shows the cumulative match scores of the first ten ranks for the three methods. Both the geometrical method and the eigenface method perform poorly in the experiment. Only around 30% accuracy is obtained for the first match. The accuracy for the tenth rank match is 70%. The poor performance of the eigenface method can be expected given the large differences between photo and sketch. As for the geometrical measure, the results show that the reason that photo and sketch look alike is not mainly because of the geometrical similarity of the facial components. As pointed out earlier, like caricature, a sketch exaggerates the sizes of facial components. If a person has a larger than average nose, the sketch will depict an even larger nose. On the contrary, if a person has a smaller than normal nose, he will be drawn with a nose with a further reduced size. The results demonstrate the effect of such exaggeration.

The eigensketch transform method greatly improves the recognition accuracy to 96% for the top ten match. The first match more than doubles the other two methods. This clearly shows the advantage of the new approach. It should be pointed out that the absolute accuracy of the algorithm should not be given too much emphasis, given the still relatively small size of the database. The results also depend on the quality of the sketch drawing. As shown in Fig. 1, not all sketches look exactly like the original photo. The first row of sketches in Fig. 1 is quite similar to the corresponding photos, yet sketches in the second row are much less so. The significance of the results lies upon the large gap between the new methods and the traditional face recognition methods.

B. Comparison of the Three Distance Measures

In this section, we conduct a set of experiments to compare the performance of the three distance measures defined in Section III-B. The same dataset described above is used for the comparison. Experimental results are shown in Table II. From the results we can see that $d_1 = ||\vec{\mathbf{c}}_p - \vec{\mathbf{c}}_s||$ is least effective among the three distances. This is not surprising, since both $\vec{\mathbf{c}}_p$ and $\vec{\mathbf{c}}_s$ represent coefficients projected in nonorthogonal spaces spanned by the training photos and sketches respectively, they cannot properly reflect the distance between face images. Both d_2 and d_3 are distances computed in orthogonal eigen-spaces and thus give much better performance. An interesting observation is that d_2 is consistently better than d_3. This seems to suggest that the sketch eigenspace can characterize the difference among different people better than photo eigenspace. This is possible since in the drawing process, an artist tend to capture and highlight the distinct characteristics of a face thus makes it easier to be distinguished. The above experiment seems to confirm this point, since $||\vec{\mathbf{c}}_p - \vec{\mathbf{c}}_s||$ has better recognition performance after projection to the sketch eigenspace than to the photo eigenspace.

Alternatively, we may also consider another explanation for the better performance of d_2. In order to compute d_2, a photo needs to be transformed into a pseudo-sketch, while to compute d_3 a sketch has to be converted to a pseudo-photo. In general, compressing more information into a smaller compact representation is a more stable operation than enlarging a compact representation to a full representation. Since photos contain much more detail information than sketches, it should be easier to convert a photo into sketch. For an extreme example, suppose the sketch only contains some simple outlines of facial features, it is quite easy to draw the outlines from the face photo, however it will be very difficult to reconstruct the photo from the simple line drawings. Therefore, for d_2 computation, better performance is achieved because of the more stable photo-to-sketch transformation.

C. Comparison With Human Performance

We conduct two experiments to compare the new method with sketch recognition by human beings. Such a comparison is important since in current law enforcement application, the sketch of a suspect is usually widely distributed through mass media.

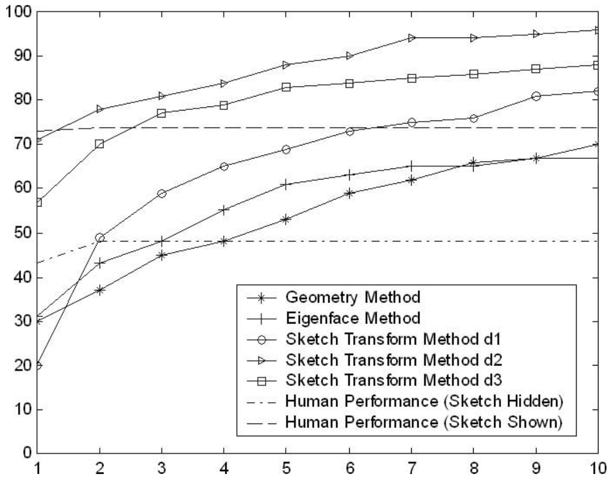

Fig. 5. Comparison of cumulative match scores between various automatic recognition methods and the human performance.

It is expected that a match with the real person can be found by people who have seen the sketch. If we can demonstrate that automatic recognition by computers can perform comparably as well as human beings, we can then use computers to systematically conduct large-scale search in a large photo-ID database.

In the first experiment, a sketch is shown to a human test candidate for a period of time, then the sketch is taken away before the photo search starts. The candidate tries to memorize the sketch, then goes on to search the photo database without the sketch reference in front. The candidate can go through the database and is allowed to select up to 10 photos that are similar to the sketch. He can then rank the selected photos according to the similarity level to the sketch. This is closer to a real application scenario, since people usually see the sketch of a criminal suspect in a newspaper or on TV briefly and then must rely on their memory to match the sketch with the suspect in real life.

For the second experiment, we allow the test candidate to look at the sketch while he searches through the photo database. The result can be considered as a benchmark for the automatic recognition system to compare. Experimental results of both tests are shown in Fig. 5. The human performance for the first experiment is much lower than the computer recognition result. This is not only because of the difference between photo and sketch, but also because of the memory distortion, since it is difficult to precisely memorize the sketch. In fact, people are very good at distinguishing familiar faces, such as relatives and famous public figures, but are not very good at distinguishing strangers. Without putting the sketch and photo together for a detailed comparison, it is hard for a person to recognize the two.

When the candidate is allowed to see the sketch while searching through the database, the accuracy rate rises to 73%, which is comparable to the computer recognition rate. However, unlike the computer recognition rate, which increases to 96% for the tenth rank, the human performance does not increase much with the rank. These encouraging results show that a computer can perform sketch matching with accuracy at least as comparable as that obtained by a human being. Given this, we can now perform automatic searching of a large database using a sketch just like using a regular photo. This is extremely important for law enforcement application where a photo is often not available.

V. Conclusion

A novel face sketch recognition algorithm is developed in this paper. The photo-to-sketch transformation method is shown to be an effective approach for automatic matching between a photo and a sketch. Surprisingly, the recognition performance of the new approach is even better than that of human beings. Of course, like most of the regular photo-based face recognition researches, further verification of our conclusions are needed on a larger scale of test. Nevertheless, without considering the absolute recognition accuracy, the relative superior performance of the new method compared to the human performance and the conventional photo based methods clearly demonstrates the advantage of the new algorithm.

Acknowledgment

The authors would like to thank Prof. E. Grimson of MIT for his encouragement of this research and many fruitful discussions. They would also like to thank Dr. Q. Yang for many useful comments. K. Ma and Z. Li contributed a great amount of time in the construction of the database. They would also like to acknowledge all the contributions from the student helpers and volunteers at the Department of Information Engineering of the Chinese University of Hong Kong, especially all the students from classes IEG 4190 and IEG 2051B.

References

[1] P. J. Benson and D. I. Perrett, “Perception and recognition of photographic quality facial caricatures: Implications for the recognition of natural images,” *Eur. J. Cognitive Psychol.*, vol. 3, no. 1, pp. 105–135, 1991.

[2] V. Bruce, E. Hanna, N. Dench, P. Healy, and A. M. Burton, “The importance of ‘mass’ in line drawings of faces,” *Appl. Cognitive Psychol.*, vol. 6, pp. 619–628, 1992.

[3] V. Bruce and G. W. Humphreys, “Recognizing objects and faces,” *Visual Cognition*, vol. 1, no. 2/3, pp. 141–180, 1994.

[4] R. Brunelli and T. Poggio, “Face recognition: Features versus template,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, pp. 1042–1052, Oct. 1993.

[5] R. Chellappa, C. L. Wilson, and S. Sirohey, “Human and machine recognition of faces: A survey,” *Proc. IEEE*, vol. 83, pp. 705–741, May 1995.

[6] I. Craw, H. Ellis, and J. R. Lishman, “Automatic extraction of face-features,” *Patt. Recognit. Lett.*, vol. 5, pp. 183–187, Feb. 1987.

[7] G. M. Davies, H. D. Ellis, and J. W. Shepherd, “Face recognition accuracy as a function of mode of representation,” *J. Appl. Psychol.*, vol. 63, pp. 180–187, 1978.

[8] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. New York: Academic, 1972.

[9] M. Kirby and L. Sirovich, “Application of the karhunen-loeve procedure for the characterization of human faces,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 103–108, Jan. 1990.

[10] M. Lades, J. C. Vorbriiggen, J. Buhmann, J. Lange, C. Malsburg, R. P. Wurtz, and W. Konen, “Distortion invariant object recognition in the dynamic link architecture,” *IEEE Trans. Computers*, vol. 42, pp. 300–311, Mar. 1993.

[11] H. Moon and P. J. Phillips, “Analysis of PCA-based face recognition algorithms,” in *Empirical Evaluation Techniques in Computer Vision*, K. W. Bowyer and P. J. Phillips, Eds. Los Alamitos, CA: IEEE Computer Soc. Press, 1998.

[12] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, “The FERET evaluation,” in *Face Recognition: From Theory to Applications*, H. Wechsler, P. J. Phillips, V. Bruce, F. F. Soulie, and T. S. Huang, Eds. Berlin, Germany: Springer-Verlag, 1998.
[13] G. Rhodes and T. Tremewan, “Understanding face recognition: Caricature effects, inversion, and the homogeneity problem,” *Visual Cognition*, vol. 1, no. 2/3, pp. 275–311, 1994.
[14] A. Samal and P. A. Iyengar, “Automatic recognition and analysis of human faces and facial expressions: A survey,” *Patt. Recognit.*, vol. 25, no. 1, pp. 65–77, 1992.
[15] X. Tang and X. Wang, “Face photo recognition using sketch,” *Image Processing*, pp. I-257–I-260, 2002.
[16] M. Turk and A. Pentland, “Eigenfaces for recognition,” *J. Cognitive Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.
[17] ——, “Face recognition using eigenfaces,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1991, pp. 586–591.
[18] L. Wiskott, J. Fellous, N. Kruger, and C. Malsburg, “Face recognition by elastic bunch graph matching,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 775–779, July 1997.
[19] A. Yuille, D. Cohen, and P. Hallinan, “Feature extraction from faces using deformable templates,” in *Proc. IEEE Computer Soc. Conf. Computer Vision and Pattern Recognition*, 1989, pp. 104–109.
[20] J. Zhang, Y. Yan, and M. Lades, “Face recognition: Eigenface, elastic matching, and neural nets,” *Proc. IEEE*, vol. 85, pp. 1423–1435, Sept. 1997.

Xiaoou Tang (S’93–M’96–SM’02) received the B.S. degree from the University of Science and Technology of China, Hefei, in 1990 and the M.S. degree from the University of Rochester, Rochester, NY, 1991, and the Ph.D. degree from the Massachusetts Institute of Technology, Cambridge, in 1996.

He is currently an Associate Professor and the founding Director of the Multimedia Lab in the Department of Information Engineering, the Chinese University of Hong Kong. His research interests include video processing and pattern recognition.

Xiaogang Wang (S’02) received the B.S. degree from the University of Science and Technology of China, Hefei, in 2001 and the M.Phil. degree from the Chinese University of Hong Kong in 2003.

He is currently a Research Assistant with the Multimedia Lab in the Department of Information Engineering of the Chinese University of Hong Kong. His research interests include face recognition, image processing, and pattern recognition.