Face Sketch Synthesis and Recognition

Xiaoou Tang and Xiaogang Wang

Department of Information Engineering
The Chinese University of Hong Kong
Shatin, Hong Kong
{xtang, xgwang1}@ie.cuhk.edu.hk

Abstract

In this paper, we propose a novel face photo retrieval system using sketch drawings. By transforming a photo image into a sketch, we reduce the difference between photo and sketch significantly, thus allow effective matching between the two. To improve the synthesis performance, we separate shape and texture information in a face photo, and conduct transformation on them respectively. Finally a Bayesian classifier is used to recognize the probing sketch from the synthesized pseudo-sketches. Experiments on a data set containing 606 people clearly demonstrate the efficacy of the algorithm.

1. Introduction

Face recognition has attracted great attention in recent years. An important application of face recognition is to assist law enforcement. Automatic retrieval of photos of suspects from police mug-shot database can help the police narrow down potential suspects quickly. However, in most cases, the photo image of a suspect is not available. The best substitute is often a sketch drawing based on the recollection of an eyewitness. Therefore, automatically searching through a photo database using a sketch drawing is 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 the similar photos retrieved.

However, due to the great difference between sketches and photos, and the unknown psychological mechanism of sketch generation, face sketch recognition is much more difficult than the normal face recognition based on photo image. During the past three decades, many face recognition techniques have been proposed [13], however, few effective face sketch recognition systems can be found in previous researches. Methods using traditional photo-based face recognition techniques such as the eigenface method [10] and the elastic graph matching method [4] have been tested with very small sketch datasets. In [10], the sketch was normalized in

geometry and blurred by a Gaussian filter in preprocessing, and then recognized by the eigenface method. The experiment includes only 7 sketches. The method in [4] recognized face sketches using Elastic Graph Matching [12], and was tested on 13 sketches only.

Photo and sketch have different modalities as shown by examples in Figure 1. The key for sketch-based face photo recognition is to reduce the difference between the two modalities. In this paper, using the eigentransformation algorithm we developed earlier [9], we propose a sketch synthesis method based on separate transformation of photo texture and shape. This method significantly reduces the difference between photo and sketch. We show that the synthesized sketch by the separate transformation is a good approximation to the real one when the transformation procedure can be approximated as linear. A Bayesian classifier combining texture and shape features is then designed to recognize the probing real sketch from the synthesized pseudo-sketches. To evaluate face sketch recognition performance on a large database, we construct a database containing photo-sketch pairs of 606 people. Experiments show that our method is much more effective than using the conventional photo-based methods directly. The new method is also shown to outperform human beings.

2. Face sketch synthesis

It is difficult to directly match photo and sketch since they are in different modalities. The starting point of our algorithm is to transform photo into sketch, so that recognition can be performed in the same modality. The relationship between photo and sketch is learnt from a set of training photo-sketch pairs using an eigentransformation procedure.

2.1. Algorithm for eigentransformation

Face image can be reconstructed from eigenfaces in the PCA representation. Since eigenface is computed from the training set, we can show that the reconstructed face image can also be expressed as the linear combination of training samples.
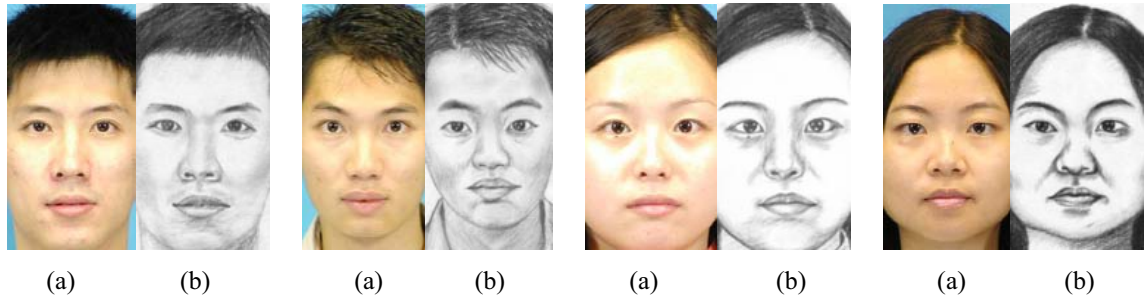
Figure 1. Samples of photo-sketch pairs. (a) is photo, and (b) is the corresponding sketch.

We represent the photo training set by a N by M matrix, $[\vec{P}_1, \vec{P}_2, \ldots, \vec{P}_M]$, where $\vec{P}_i$ is the photo vector, N is the number of image pixel, and M is the number of training samples. In PCA, a set of eigenvectors $E_p = [e_1, \ldots, e_K]$, also called eigenfaces, are computed from the covariance matrix,

$$C = \sum_{i=1}^{M} (\vec{P}_i - \vec{m}_p)(\vec{P}_i - \vec{m}_p)^T = A_p A_p^T, \quad (1)$$

where $\vec{m}_p$ is the photo mean face, and A_p is the photo sample matrix,

$$A_p = [\vec{P}_1 - \vec{m}_p, \ldots, \vec{P}_M - \vec{m}_p] = [\vec{P}'_1, \ldots \vec{P}'_M]. \quad (2)$$

According to the singular value decomposition theorem, E_p can also be computed from,

$$E_p = A_p V_p \Lambda_p^{-1/2}, \quad (3)$$

where V_p and Λ_p are the eigenvector and eigenvalue matrix for $A_p^T A_p$.

For a new face photo $\vec{P}$, it can be reconstructed from the eigenfaces by,

$$\vec{P}_r = E_p \vec{w}_p + \vec{m}_p, \quad (4)$$

where $\vec{w}_p$ is the weight vector computed by projecting the face photo onto the eigenfaces,

$$\vec{w}_p = E_p^T (\vec{P} - \vec{m}_p). \quad (5)$$

From (3) and (4), the reconstructed photo can be represented by

$$\vec{P}_r = A_p V_p \Lambda_p^{-\frac{1}{2}} \vec{w}_p + \vec{m}_p = A_p \vec{c} + \vec{m}_p, \quad (6)$$

where $\vec{c} = V_p \Lambda_p^{-1/2} \vec{w}_p = [c_1, \ldots, c_M]^T$. Equation (6) can be rewritten as,

$$\vec{P}_r = \sum_{i=1}^{M} c_i \vec{P}'_i + \vec{m}_p. \quad (7)$$

This shows that the reconstructed photo is an optimal approximation to the original face photo using a linear combination of the M training photos. Replacing each training photo with its corresponding sketch $\vec{S}'_i$, and replacing the mean photo $\vec{m}_p$ with the mean sketch $\vec{m}_s$, we get,

$$\vec{S}_r = \sum_{i=1}^{M} c_i \vec{S}'_i + \vec{m}_s. \quad (8)$$

$\vec{S}_r$ is the synthesized sketch expected to resemble the real sketch.

2.2. Linear assumption

Two conditions are needed for the eigentransformation to work:

- A new face can be reconstructed from training samples by PCA.
- The transformation between photo and sketch can be approximated as a linear process.

Assuming that there is a linear transformation matrix T that can ideally transform a photo to sketch, the transformation can be expressed as

$$\vec{S} = T\vec{P}. \quad (9)$$

For the training set, we have

$$\vec{S}'_i = T\vec{P}'_i, \quad (10)$$

$$\vec{m}_s = T\vec{m}_p. \quad (11)$$

As shown in Figure 2, when $\vec{P}$ is projected onto the photo eigenspace, a group of coefficients $\{c_1, c_2, \ldots, c_M\}$ on the training set can be obtained, from which $\vec{P}_r$ and $\vec{S}_r$ can be reconstructed by the linear combination of
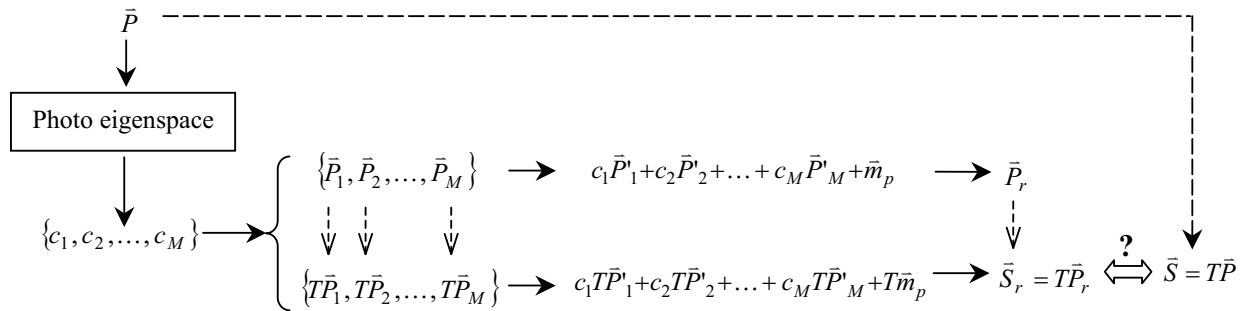
Figure 2. Eigentransformation with the assumption that the transformation between photo and sketch is linear. We use the dotted arrow to represent the drawing process.

training photos and training sketches respectively. From Eq. (7) and (8), replacing $\bar{S}'_i$ and $\bar{m}_s$ with (10) and (11), we have

$$\bar{S}_r = \sum_{i=1}^{M} c_i T\bar{P}'_i + T\bar{m}_p = T\left(\sum_{i=1}^{M} c_i \bar{P}'_i + \bar{m}_p\right) = T\bar{P}_r . \quad (12)$$

This shows that the reconstructed sketch is in fact a sketch drawn based on $\bar{P}_r$. Photo-based face recognition studies [7] have shown that for eigenface reconstruction $\bar{P}_r$ is close to $\bar{P}$ because of the facial structural similarity. Therefore, comparing Eq. (9) and (12), we see that the reconstructed sketch $\bar{S}_r$ should be similar to $\bar{S}$. The linear requirement is critical for Eq. (12) to hold. The linear assumption is not unreasonable since some highpass-filered images are actually sketch-like. For a simple example, the edge gradient map obtained by linear edge detector can be seen as a line drawing sketch. Thus it is possible to use linear operator to generate sketch-like images from the original photo. Of course, for the real sketch drawn manually, the transform cannot be strictly linear, but an approximation.

Even though the transformation may be simplified to be linear, the transformation matrix T is still too complicated to be expressed explicitly. For an image of size 128 by 128, the length of the image vector is 128^2. So the matrix T has 128^4 elements to be defined. Eigentransformation takes advantage of the linear property and face structural similarity to generate the sketch by using only a small number of training samples without actually deriving the large transformation matrix.

2.3. Framework of face sketch synthesis

The performance of the sketch synthesis by eigentransformation depends on the linear assumption. Given the fairly complex structure of human face, this assumption is rather difficult to be demonstrated. Since the difference between sketch and photo exists in both

texture and shape, corresponding points in the sketch and photo can be quite different without proper alignment, thus the process becomes difficult to be described by a linear process. However, if we separate the texture and shape, and then treat them independently, a closer linear correspondence can then be established.

The shape distortion is somewhat caused by that the artist tries to exaggerate some distinctive features just like caricature. For example, if a face has a big nose in a photo, the nose drawn in the sketch will be even bigger. A study in [5] suggested that the shape exaggeration could be approximated as,

$$G_s = E\left(G_p - G_p^m\right) + G_p , \quad (13)$$

where G_p is the photo shape vector, G_s is the sketch shape vector, G_p^m is the mean photo shape, and E is the exaggeration matrix. The difference between the photo shape and the mean shape is exaggerated. Averaging both sides of Eq. (13), we get the mean shape of sketch,

$$G_s^m = E\left(G_p^m - G_p^m\right) + G_p^m = G_p^m . \quad (14)$$

The mean shapes of photo and sketch are actually the same. Subtract Eq. (14) from Eq. (13), we have,

$$\left(G_s - G_s^m\right) = \left(E + I\right)\left(G_p - G_p^m\right). \quad (15)$$

Thus we prove that the shape transformation between photo and sketch can be approximated as linear.

The texture in a sketch is formed by the grayscale changes in small local areas. It is reasonable to assume that the grayscale around a fiducial point in the sketch is mainly influenced by the grayscale around the same fiducial point in the photo. However, because of shape distortion, the same fiducial points have different coordinates in different photos and sketches, thus it is difficult to derive the linear relation for the local texture transformation. Therefore it is necessary to separate the shape from texture.
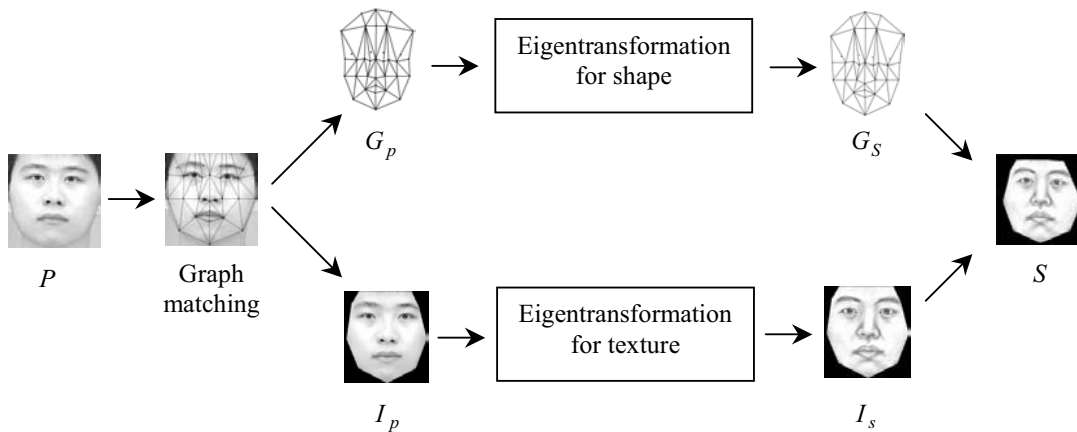
Figure 3. Framework of the face sketch synthesis system.

We represent face shape with a graph containing the coordinates of a set of fiducial points. A mean shape is computed from the training set. In order to remove shape factor, we warp the face image to the mean shape using the affine interpolation based on a set of triangles. After alignment, the fiducial points in different face photos and sketches finally correspond to the same position. We observe that the sketch grayscales after shape alignment also has a similar style of exaggeration as the sketch shape as shown by Eq. (13). If an area in photo is light color, the artist will leave it blank in the sketch; if an area is relatively dark, the artist tends to emphasize it more with shade texture. Therefore, at least within a small local neighbor, there is a linear trend, thus a linear relation similar to Eq. (15) can be derived for texture transform. Of course, this is a very rough approximation, since an artist will not decide on the grayscale of a small area only based on the grayscales of the same area in the photo. For precise description of the texture transformation, the whole picture has to be taken into consideration.

Finally, the sketch synthesis system based on separate shape and texture eigentransformation can be implemented through the following steps, as shown in Figure 3:

- For an input face photo P, locate all the fiducial points on the face graph model to extract shape information.
- Warp the face image to a mean face shape derived from training set to separate the texture I_p and shape G_p from the photo image.
- Apply eigentransformation to the photo texture and shape respectively to generate texture I_s and shape G_s for the sketch.
- Warp the generated texture from the mean shape to the sketch shape to produce the final synthesized sketch S.

3. Face sketch recognition

Face sketch recognition is based on the matching between the probing real sketch and the synthesized pseudo-sketch from photo. In this section, we present the PCA and Bayesian classifiers for recognition. For classification, we extract a set of salient geometric measures from the face graph to represent the shape feature, including the sizes and relative positions of nose, eyes, eyebrows, and face contours etc., and the texture vector is normalized by the shape.

3.1. PCA classifier

Let $G \in R^{N_1}$ and $I \in R^{N_2}$ represent the shape and texture vectors, where N_1 and N_2 are the vector length for shape and texture. The feature vectors used here is similar to the features used for photo-based recognition in active shape models [2][3]. Eigenspaces for shape and texture are computed from the sketch training set. In the PCA classifier, feature vectors are projected to eigenspaces to get the low dimensional features,

$$\bar{x} = E_G(G - m_G), \qquad (16)$$

$$\bar{y} = E_I(I - m_I), \qquad (17)$$

where E_G and E_I are the eigenvector matrices of shape and texture respectively, and m_G and m_I are the averages of shape and texture respectively. The shape and texture features are normalized to unit norms, and form an integrated feature,

$$\bar{z} = \left(\frac{\bar{x}^T}{\| \bar{x} \|} \; \frac{\bar{y}^T}{\| \bar{y} \|} \right)^T . \qquad (18)$$

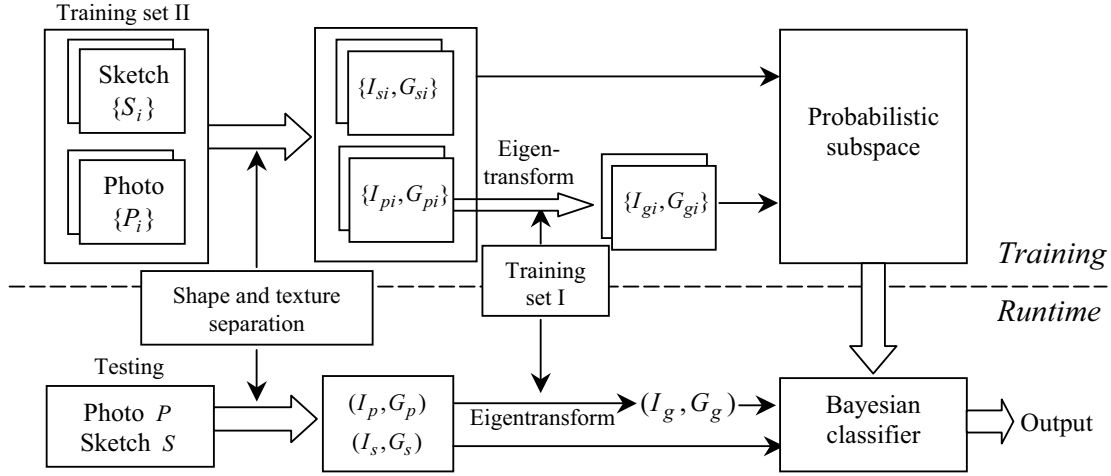Classification is based on the Euclid distance,

Figure 4. Framework of the face sketch recognition system.

$$d = \| \vec{z}_s - \vec{z}_g \|. \qquad (19)$$

where $\vec{z}_s$ and $\vec{z}_g$ are the integrated features for probe sketch and pseudo-sketch from photo.

3.2. Bayesian classifier

Although eigentransformation let the matching be performed in the same modality, the synthesized sketch is still not a perfect estimation to the real one. To further reduce the effect of transformation error at recognition stage we apply the Bayesian classifier [6] in this section. Bayesian face recognition defines the difference (Δ) of two faces as intrapersonal variation Ω_I (caused by transformation error) or extrapersonal variation Ω_E (caused by individual difference). The classification is based on the intrapersonal likehood $P(\Delta | \Omega_I)$.

The transformation error is modeled as Gaussian distribution. First, PCA is applied to the intrapersonal difference set $\{(S_i^r - S_i)\}$, which is the transformation error between synthesized sketch (S_i^r) and the real sketch (S_i), to compute the probabilistic subspace. Projecting Δ onto the probabilistic subspace, $P(\Delta | \Omega_I)$ is estimated as the product of two densities in principle space F and its complementary space $\overline{F}$,

$$\hat{P}(\Delta | \Omega_I) = \left[\frac{\exp\left(-\frac{1}{2} d_F(\Delta)\right)}{(2\pi)^{K/2} \prod_{i=1}^{K} \lambda_i^{1/2}} \right] \left[\frac{\exp\left(-\varepsilon^2(\Delta)/2\rho\right)}{(2\pi\rho)^{(N-K)/2}} \right]. \qquad (20)$$

In Eq. (20), $d_F(\Delta)$ is a Mahalanobis distance in F,

$$d_F(\Delta) = \sum_{i=1}^{K} \frac{y_i^2}{\lambda_i}, \qquad (21)$$

where y_i is the principal component and λ_i is the eigenvalue explicitly describing the energy distribution of transformation error; $\varepsilon^2(\Delta)$ is equivalent to PCA residual error in $\overline{F}$; ρ is the average eigenvalue in $\overline{F}$.

In our algorithm, we separate the face image into shape and texture, and assume that they are independent. The Bayesian classifier is modified to integrate the two kinds of information by,

$$\hat{P}(\Delta | \Omega_I) = \hat{P}((\Delta_I, \Delta_G) | \Omega_I)$$
$$= \hat{P}(\Delta_I | \Omega_I) \cdot \hat{P}(\Delta_G | \Omega_I). \qquad (22)$$

Δ_I and Δ_G are the face difference in texture and shape. where $\hat{P}(\Delta_I | \Omega_I)$ and $\hat{P}(\Delta_G | \Omega_I)$ are the intrapersonal likehoods for texture and shape respectively.

Figure 4 describes the diagram of the sketch recognition framework. The procedure is divided into two stages: training and runtime. There are two training sets. Training set I is for eigentransformation, and training set II is used to compute the probabilistic subspace for Bayesian classifier. At the training stage,

- Use training set I to compute the photo-to-sketch eigentransform coefficients.

- Photos and sketches in training set II are separated into shape and texture.

- Photo texture and shape $\{I_{pi}, G_{pi}\}$ in training set II are transformed to pseudo-sketches $\{I_{gi}, G_{gi}\}$.
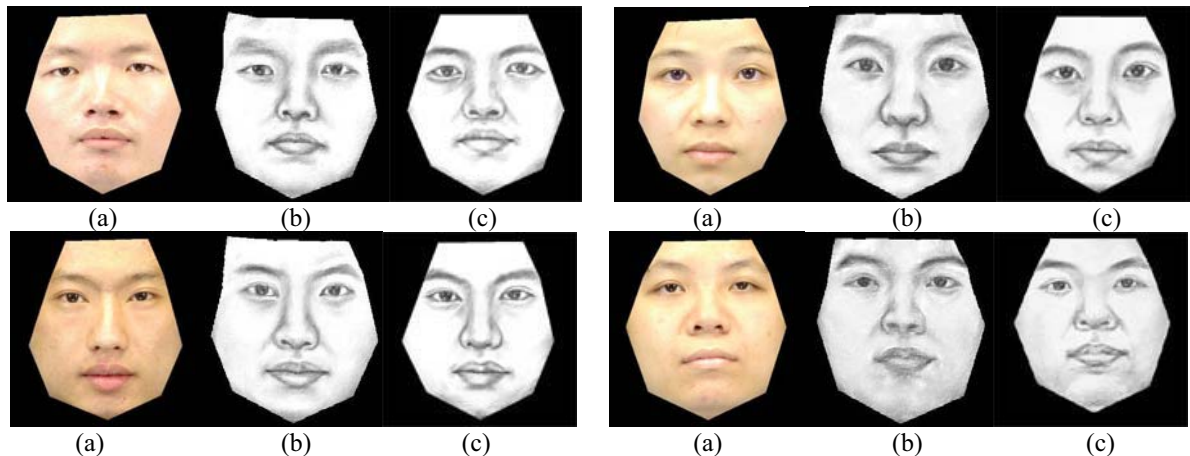
Figure 5. Face sketch synthesis result: (a) face photo; (b) synthesized sketches from photo using separate eigentransformation on texture and shape; (c) real sketch drawn by the artist.
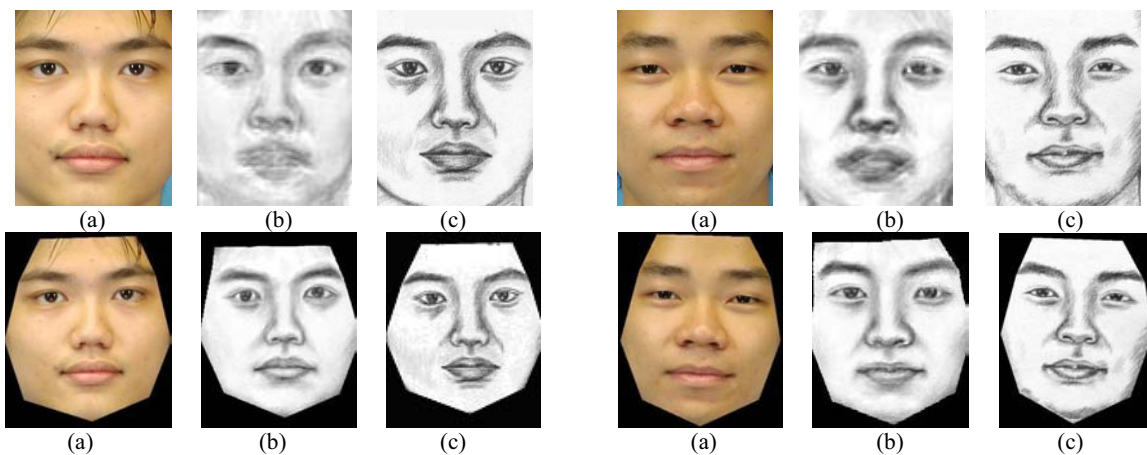

Figure 6. Comparison of the direct eigentransformation on "cropped face" (first row) with separate transformation on texture and shape (second row): (a) the input photo; (b) the synthesized sketch; (c) the sketch drawn by the artist.

- The texture and shape probabilistic subspaces are derived from training sketches $\{I_{si}, G_{si}\}$ and pseudo-sketches $\{I_{gi}, G_{gi}\}$.

At the runtime stage,

- The photo and sketch for matching are separated into texture and shape.
- The photo texture and shape (I_p, G_p) are transformed to pseudo-sketch (I_g, G_g).
- Texture and shape features for sketch and pseudo-sketch are input to Bayesian classifier, and the face sketch is recognized.

4. Experiment

To evaluate the face sketch recognition performance on a large database, we construct a sketch database containing 606 people. For each person, there is a frontal photo face image and a face sketch drawn by an artist. In this experiment, the 606 people are partitioned into three sets. Training set I and II contain 153 photo-sketch pairs each, and the testing set contains 300 photo-sketch pairs. Human hair is discriminative feature for short-term recognition, but it may vary significantly over a long period. We remove most of the hair and background in preprocessing.

4.1. Sketch synthesis performance

Figure 5 shows some results of sketch synthesis using separate eigentransformation on texture and shape. The individuals for the sketches to be synthesized are not in the training set. In the synthesized sketches, the skin color has been transformed to sketch texture, and there are noticeable shadows very similar to the ones drawn by pencil. The distinctive features on the face photo have

Table 1. Recognition accuracies using different features and classifiers (%).

	Cropped face	Texture	Shape	Texture + Shape
PCA	6.3	5.3	30.7	25.0
Eigentransform + PCA	53.7	45.0	35.3	75.0
Eigentransform + Bayes	74.3	56.7	53.0	81.3

Table 2. Cumulative match score for three methods (%).

	1	2	3	4	5	6	7	8	9	10
Eigenface	6.3	8.0	9.0	9.3	11.3	13.3	14.0	14.0	14.3	16.0
EGM	25.3	32.3	40.0	43.0	46.7	48.7	53.0	54.3	56.3	57.7
New method	81.3	91.0	94.7	95.7	96.7	97.0	97.0	97.0	97.0	97.0

been captured and exaggerated in shape. Figure 6 demonstrates the improvement of separating shape and texture when conducting eigentransformation. The first row is the results of directly applying eigentransformation on "cropped face", which are formed by aligning eye centers and cropping the inner face. The synthesized sketches are somewhat blurred and have alias noise because of the non-linear problem. When texture and shape are separated, the results are much clearer and sharper as shown in the second row.

4.2. Sketch recognition performance

Table 1 reports the sketch recognition accuracies using three different classifiers applied on four kinds of features, "copped face", texture, shape and the integration of texture and shape. The direct PCA method treats the probing sketch as a regular photo, and matches photo and sketch in the eigenspace computed from the photo training set. The recognition performance is poor. The low accuracy on shape demonstrates that the reason for photo and sketch look alike is not because of the geometrical similarity of facial components.

The second classifier is PCA based on eigentransformation. The matching is performed between the probing sketch and the synthesized sketch in the eigenspace computed from the training sketches. It achieves significant improvement to the direct PCA method, since the match is performed in the same modality after transformation. The experiment also shows the improvement of transforming face texture and shape separately and integrating them in recognition. It is much better than applying transformation on the "cropped face" without separating texture and shape. In the third classifier, the Bayesian algorithm further reduces the transformation error. The Bayesian classifier based on eigentransformation integrating texture and shape feature has the highest recognition accuracy 81.3%.

In Table 2, we compare our new method with two conventional face recognition method, eigenface [8] and Elastic Graph Matching (EMG) [12], both of which have been successfully applied to face photo recognition. As
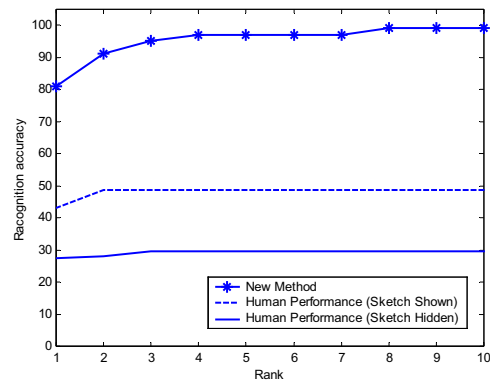

Figure 7. Comparison of accumulative match score between our automatic recognition method and human performance.

discussed in the introduction, both methods [10][4] have also been tested on very small datasets of sketches in previous study. The results in Table 2 clearly demonstrate the superiority of our algorithm over these conventional methods. Using a testing set containing 300 photo-sketch pairs, the first match for conventional methods is no more that 30%, and the tenth rank is no more than 60%. Our algorithm significantly improves the first match to 80%, and the tenth rank to 97%.

4.3. Comparison with human recognition

We conduct two experiments to compare the new method with sketch recognition by human beings. If we can demonstrate that automatic recognition by computers can perform better than human beings, we can then use computers to systematically conduct large-scale search in a large photo-ID database.

We select 100 photo-sketch pairs from the testing set for human recognition. Similar to automatic recognition, the hair is removed in the cropped faces. Thirty candidates are asked to do the test. 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 go on to search the photo database without the sketch reference in front. The candidate can go through the database and are 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 real application scenario. Since, people usually see the sketch of a criminal suspect in a newspaper or on TV briefly, then they have to 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 they search through the photo database. This simulates the case when an eye witness looks though the police database for a suspect.

The encouraging experimental results in Figure 7 shows that a computer can perform better than a human being. The human performance for the first experiment is much lower. 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. Given the good automatic recognition results, 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.

5. Conclusion

In this paper, we develop a framework for face sketch synthesis and recognition. Using a separate eigentransformation algorithm, sketch is synthesized from face photo, and the matching can be effectively performed in the same modality. The Bayesian classifier further reduces the transformation error in recognition. Our algorithm significantly outperforms both conventional methods and human beings. The automatic face sketch recognition system can help law enforcement greatly. In addition, the sketch synthesis approach also can be widely applied to newspaper and magazine drawing, web design, and entertainment industries.

Finally, in this paper, the study focuses on overcoming different modalities between photo and sketches. The photos are frontal face images taken under controlled conditions with not much changes in expression, view, and illumination. It is important to make the system more robust to some non-perfect cases in future study.

References

[1] R. Chellappa, C.L. Wilson, and S. Sirohey, "Human and machine recognition of faces: a survey", 1995.

[2] T. F. Cootes and C. J. Taylor, "Statistical Model of Appearance for Computer Vision," *Technical Report*, University of Manchester, Manchester M13 9PT, U. K., 2001.

[3] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active Appearance Models," *IEEE Trans. on PAMI*, Vol. 23, No. 6, pp. 681-685, June, 2001.

[4] W. Konen, "Comparing Facial Line Drawings with Gray-Level Images: A Case Study on PHANTOMAS," *Proceedings of International Conference on Artifical Neural Networks*, pp. 727-734, 1996.

[5] H. Koshimizu, M. Tominaga, T. Fujiwara, and K. Murakami. "On Kansei Facial Processing for Computerized Facial Caricaturing System PICASSO," *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics,* Vol. 6, pp. 294-299, 1999.

[6] B. Moghaddam, T. Jebara, and A. Pentland, "Bayesian Face Recognition," *Pattern Recognition*, Vol. 33, pp. 1771-1782, 2000.

[7] P. S. Penev and L. Sirovich, "The Global Dimensionality of Face Space", *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition,* pp. 264-270, 2000.

[8] M. Turk and A. Pentland, "Eigenfaces for Recognition", *J. of Cognitive Neuroscience*, Vol. 3, No. 1, pp. 71-86, 1991.

[9] X. Tang, and X. Wang, "Face Photo Recognition Using Sketch," *Proceedings of ICIP*, pp. I-257-I-260, Vol. 1, 2002.

[10] R.G. Uhl and N.d.V. Lobo, "A Framework for Recognizing a Facial Image from A Police Sketch," *Proceedings of CVPR*, pp. 586-593, 1996.

[11] T. Vetter and T. Poggio, "Linear Object Classes and Image Synthesis From a Single Example Image," *IEEE Trans. on PAMI*, Vol. 19, No. 7, July 1997.

[12] L. Wiskott, J.M. Fellous, N. Kruger and C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* Vol. 19, No.7, pp. 775-779, July, 1997.

[13] W. Zhao, R. Chellappa, and P. Phillips. "Face Recognition: A Literature Survey", *Technical Report*, 2002.