Face Image Retrieval Based on Probe Sketch Using SIFT Feature Descriptors

Rakesh S[1] ,Kailash Atal[2], Ashish Arora[2], Pulak Purkait[3],Bhabatosh Chanda[3]

[1]Dept. of Computer Science, National Institute of Technology Karnataka, Surathkal, [2]Department of EEE, IIT Guwahati, Guwahati,[3]ECSU, Indian Statistical Institute, 203,B.T. Road, Kolkata

{rakesh.s.mysore, atalkailash, ashish.arora.iitg }@gmail.com,{chanda,pulak_r}@isical.ac.in

Abstract. This paper presents a feature-based method for matching facial sketch images to face photographs. Earlier approaches calculated descriptors over the whole image and used some transformation and matched them by some classifiers. We present an idea, where descriptors are calculated at selected discrete points (eyes, nose, ears…). This allows us to compare only prominent features. We use SIFT (Scale Invariant Feature Transform) to extract feature descriptors at the annotated points in the sketches and experiment with various methods to retrieve photos. Experimental results demonstrate appreciable matching performances using the presented feature-based methods at a low computational cost.

Keywords: Forensic sketch, image registration, annotated points, SIFT feature descriptors.

1 Introduction

The problem of retrieving facial photos resembling a sketch from gallery of photos has received substantial attention from research community, security agencies, crime investigators etc. An important application of this is to assist law enforcement. During a criminal activity, the photo image of the suspect is generally not available. The criminals are sensitive to not leave behind any trace of their identity in the form of fingerprint or any other biometric. In such situations a recollected description of eyewitnesses is used by forensic artist to draw an estimate sketch of the culprit. The law enforcement agencies need assistance to automatically retrieve photos of potential suspects from the criminal photo database based on available sketch. Since the sketch is not an exact portrayal of the culprit, it becomes difficult to match real-time sketches exactly to their corresponding photos. Additional difficulties are posed due to difference in modalities of sketches and photos. Considering these challenges, criminal investigators are generally interested in the top N retrieved results because of low probability of finding an exact match and relatively higher likelihood of finding a correct match in these retrieved photos. This reduces the burden of investigators to manually search for the exact match of the sketch in the whole database and saves

crucial time. It also helps the witness and artist to modify the sketch drawing of the suspect based on retrieved results.

Our proposed solution to the problem is somewhat in between face recognition and image retrieval. In former case only exact match is considered, while in the latter any object of same category provides an acceptable solution. In this paper we study two interesting and related problems: similar visual feature extraction from sketches and photos, and comparing features of the sketch-photo pairs. In order to solve the problem we use SIFT algorithm to extract visual descriptors at key points on facial sketches and photos such as eyes, nose, ears, lips etc on registered image. We extend our approach by performing experiments using obtained feature vectors to correctly retrieve photos of true subject based on probe sketch. Rest of the paper is organized as follows. Section 2 presents the relevant works of recent past. Proposed method and its performance analysis are given in section 3 and 4 respectively. Finally concluding remarks are placed in section 5.

2 Related Work

Most previous research works on sketch to photo matching has concentrated on linear or non-linear approaches like Eigen-transformation (Tang and Wang [11], [9]) and Markov random field (Wang and Tang [10]). These studies share a common approach of synthesizing sketches from photos and then matching these synthesized sketches with probe sketch or synthesizing a photo from probe sketch (Purkait and Chanda [13]) and the matching the synthesized photo against the gallery of photo database.

Recent works implement SIFT (Karle, Li and Jain [2]) or LBP (Ahonen, Hadid, and Pietikainen [4]) on the whole image which use feature descriptors to create distinct identity of a person. Local descriptors like SIFT [6], LBP [4] and CITE [5] are commonly employed in a number of real time application such as face photo retrieval. These descriptors diminish the effect of difference in modalities of sketch and photo while still maintaining the distinct identity of a person. Image based feature descriptors have shown success in face recognition in the past years [7].

3 Proposed Method

In this section, we give a detailed explanation of our approach to retrieve photos based on probe sketch. This has distinctly two parts : Training and Test. Training starts with manually annotating key-points on the Training set of corresponding sketch-photo pairs. Note that a corresponding pair consists of a sketch and a photo image of same subject. Images are then registered with common shape, called mean shape, using the annotated key points. In case of probe sketch key points are obtained by using active shape model (ASM). This is followed by computation of SIFT descriptors at annotated points. We finally demonstrate the ability to match sketches to photos by directly using Euclidean distance between SIFT feature, Projection Angle method of the SIFT feature and also their combination with the geometric position of the annotated key points. The algorithm may summarily be represented by the following steps:

3.1 Annotating Points

Ordinarily using SIFT on a sketch or a photo of size 500X500 pixels generates descriptors at around 2000 points on the image. Other methods like slicing a photo or sketch in small patches and computing SIFT features in those patches involves computation in very high dimensions. Such methods are often accompanied by use of PCA[1] or LFDA[2]. However points in regions near eyes, lips, nose ears etc. are potential for high distinctness for a person. So we adopted a new approach of calculating SIFT features only at selected points in these regions. During Training we manually marked 41 points on all sketch-photo pairs at identical locations on all images using am_tools [12] however Active Shape Model (ASM) ([3]) can be used to do this automatically for a given probe sketch at the time of testing. ASM is a statistical model of the shape of object in training image which iteratively deform to fit to an object in a new image. It captures the natural variability within a class of shapes. The model is built by learning the patterns of variability of annotated points of a training database.Fig.1 shows 41 key points annotated on sketch-photo pairs.

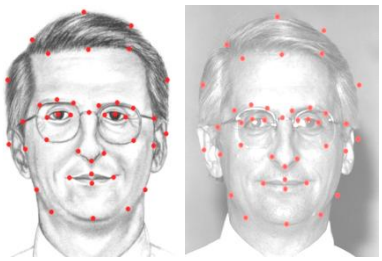

Fig.1 Annotated Points on corresponding sketch-photo pair from FERET database

3.2 Image Registration

The feature at key point based representation requires each sketch and photo image to be registered onto a common platform. This is done by transforming the image suitably such that the mean and standard deviation of the annotated points of the images database become the same.

Firstly the images are normalized by rotating by the angle between the horizontal axis and the line joining mid-point of the eyes. Now for this rotated image I_i calculate the mean (x_{im}, y_{im}), of all new 41 points. We do the same for all the images in the database, thus fetching a set of means $\{(x_{1m}, y_{1m}), (x_{2m}, y_{2m}), ..., (x_{nm}, y_{nm})\}$ of n images $\{I_1, I_2, ..., I_n\}$. Then we compute the global mean (x_{gm}, y_{gm}) of these n means. To make the images mean centered, we translate each image I_i in x direction by x_{gm}- x_{im} and in y direction by y_{gm}- y_{im} to get (x'_{ki}, y'_{ki}).

The final step of registration is scaling. We scale images in such a way that the standard deviation of Euclidean distance of the annotated points about the global mean (x_{gm}, y_{gm}) is same for all. We first compute the local standard deviation of an all the images and scale them by a ratio of mean standard deviation to local standard deviation using the following transformation.

$$\begin{matrix} x''_{ki} \\ y''_{ki} \\ 1 \end{matrix} = \begin{matrix} scale & 0 & 1-scale\ .x_{gm} \\ 0 & scale & 1-scale\ .y_{gm} \\ 0 & 0 & 1 \end{matrix} \cdot \begin{matrix} x'_{ki} \\ y'_{ki} \\ 1 \end{matrix} \quad . \qquad (1)$$

After these three steps – rotation, translation and scaling, all the images would be registered to the same platform.

3.3 Feature Extraction

Visual features are computed at 41 key points. The underlying assumption is that the features extracted from eyes, nose or ears are sufficiently distinct for each. Since we have marked the points in a fixed order, it preserves correspondence while comparing features of a photo and sketch. Suppose $\mathbf{P}=(P_1,P_2,\ldots,P_k)^T$ and $\mathbf{S}=(S_1,S_2,\ldots,S_k)^T$ are feature vectors extracted from photo and sketch respectively. Now

$$\mathbf{S}' = A\mathbf{S} \quad . \qquad (2)$$

where k x k matrix A stands for cross modality transformation which takes sketch features to domain of photo features. Thus to retrieve photo from sketch, we need to match **P** with $\mathbf{S}'$ and not with **S**. Now if A can be diagonalized with transform matrix W, i.e., $A= W^T D W$ then equation (2) reduces to $\mathbf{S}' = W^T DW\mathbf{S}$ or

$$W\mathbf{S}' = DW\mathbf{S} . \qquad (3)$$

Under the assumption that every element of feature vector carry equal amount of information, the diagonal matrix D becomes λI. Then equation (3) may be written as

$$\mathbf{S}' = \lambda\mathbf{S} . \qquad (4)$$

That means in transform domain sketch and photo features are related by a scalar multiplier. SIFT feature, due to its inherent property, satisfies this criterion approximately. This is because SIFT descriptor identifies the scale and dominant orientations at the selected points. The orientation(s), scale and selected locations enables SIFT to construct a canonical view for the point that is invariant to similarity transforms. The reader is referred to [6] for more detailed description on SIFT. Unlike the conventional method we are not using SIFT key point detection, rather the SIFT feature descriptors are computed at predetermined 41 locations. These features are well-suited for sketch-photo matching because they describe the distribution of the direction of edges in the face, which is the information common to both sketches and photos.

SIFT descriptors at j^{th} point in i^{th} photo, $P_i(j)$ is considered as a column vector with 128 elements, where j varies from 1 to 41. P_i j is normalized. Similarly S_k j represents normalized SIFT descriptors at jth point in k^{th} sketch, i.e., $P_i = (P_i\ 1\ , P_i\ 2\ , \ldots, P_i(41))^T$ and $S_k = (S_k\ 1\ ,\ S_k\ 2\ ,\ \ldots,\ S_k(41))^T$.

Suppose $X_{Pk}=(X_P(1), Y_P(1), X_P(2), Y_P(2), \ldots, X_P(41), Y_P(41))$ represents normalized coordinates of key points on photo image. X_{Sk} can be similarly defined for sketch image.

3.4 Recognition and Retrieval Methods

In order to retrieve suitable photos against a probe sketch, S_{probe} , ASM is used to automatically annotate 41 key points on the probe sketch as mentioned in section 3.1 followed by computation of SIFT descriptors at the annotated key points. And all the descriptors of target photos P_i are computed prior to recognition process. Now suppose dissimilarity between X_{Sprobe} and X_{Pi} is measured as $d(X_{Sprobe}, X_{Pi})$ and similarity that between photo and sketch SIFT descriptors S_{probe} and P_i as $d(S_{probe}, P_i)$. Hence, overall dissimilarity between photo and sketch may be measured as

$$E_i = \lambda d\ X_{Sprobe},\ X_{Pi} + 1-\lambda\ d\ S_{probe},\ P_i \quad . \qquad (5)$$

where λ is a parameter that determines the importance of coordinates and SIFT features. The dissimilarity measure is defined next.

3.4.1 Projection Angle Based Dissimilarity Measure

In this approach, SIFT descriptor at each of the 41 points is considered as a column vector with 128 elements as mentioned in section 3.1. Based on these values the dissimilarity may be computed as follows:

i. Find the mean vector P_m for photos and S_m for sketches from training images. This step is computed during training period.

ii. Subtract P_m from all the face-photos and S_m from the probe sketch to get, $P_i' = P_i - P_m$ and $S'_{probe} = S_{probe} - S_m$.

iii. Find the angle $\theta_i(j)$ between $P_i'(j)$ and $S'_{probe}(j)$, by taking dot product.

iv. Find the mean angle, $\theta_{mean,i}$ between S'_{probe} and P_i' :

$$\theta_{mean,i} = \frac{\sum_{j=1}^{41} \theta_i(j)}{41} . \qquad (6)$$

v. $\theta_{mean,i}$ is the measure of dissimilarity between S_{probe} and P_i and equals $d(S_{probe}, P_i)$ for this case.

3.4.2 Euclidean Distance Based Dissimilarity Measure

It is a simple and most trivial approach used for comparing distance between vectors. In this approach, we compute the mean of Euclidean distance between the corresponding SIFT features of the probe sketch S_{probe} and i^{th} photo P_i after converting it to S'_{probe} and P_i' as mentioned in the methodology for Projection Angle Method to

get $d(S_{probe}, P_i)$. In essence, this method is same as Projection Angle method except in step(iii) where Euclidean distance is computed between two vectors instead of dot product.

Finally the dissimilarity between probe sketch and photo is computed using equation (5) for a suitable value of λ. Hence the best match for a probe sketch is a photo for which the dissimilarity measure E_i is minimum. In case of N photos to be retrieved, lowest N dissimilarity values are considered.

4 Experimental Results and Discussion

In this section, we present the performance of our system using 969 photo-sketch pairs from the FERET database[14, 15]. We have divided our database into 2 parts- 569 for training and 400 for testing. We have repeated the experiment with 5 such random splits and the average performance of the various methods is reported in Table 1 below. The results are shown for two cases – (1) using SIFT only and (2) that value of λ for which best average result is obtained.

Table 1. Comparison of accuracy (precision %) for 400 testing sketch-photo pairs

No of photos retrieved	1	5	10	20	30	40	50
Euclidean Distance Method (SIFT only)	74.7	88.35	92.95	95.1	96.65	97.7	98.2
Euclidean Distance Method	77.25	89.7	93.9	96.1	97.3	98.0	98.45
Projection Angle Method (SIFT only)	79.95	91.45	94.6	97.7	98.55	98.9	99.15
Projection Angle Method	83.85	93.65	96.3	98.3	99.1	99.5	99.7

For the case ($\lambda = 0$), the accuracy of both methods for all the ranks is relatively lower when compared against the case ($\lambda \neq 0$). λ was determined experimentally under the constraint to maximize accuracy. This shows that not SIFT features alone, but their weighted combination with the geometric distance between key points on the registered photos and probe sketch is a better measure of dissimilarity between the two. This is shown in Fig.4.

The accuracy of 83.85% for a correct match in first retrieval and that of 96.3% in top ten retrievals using Projection Angle Method is appreciable considering the fact that database of comparison of 400 photo-sketch pairs in our case is reasonably high. The fact that the accuracy of a correct match in top 50 retrievals using this method goes close to 99.7% must be of interest for people developing real time applications.

Our choice of 400 sketch-photo pairs is random. We did not perform the test separately for males and females. Nor was there any distinction on the basis of race or origin as tried upon in the earlier work by A.K. Jain et al.[2]. The FERET database also includes some poor sketch for photo of the same subject making our work even

more difficult. In some cases, a person in photo has spectacles which is not found in the corresponding sketch. Some of these cases are shown in Fig 5. In the background of these complicacies, we may consider our results highly appreciable.

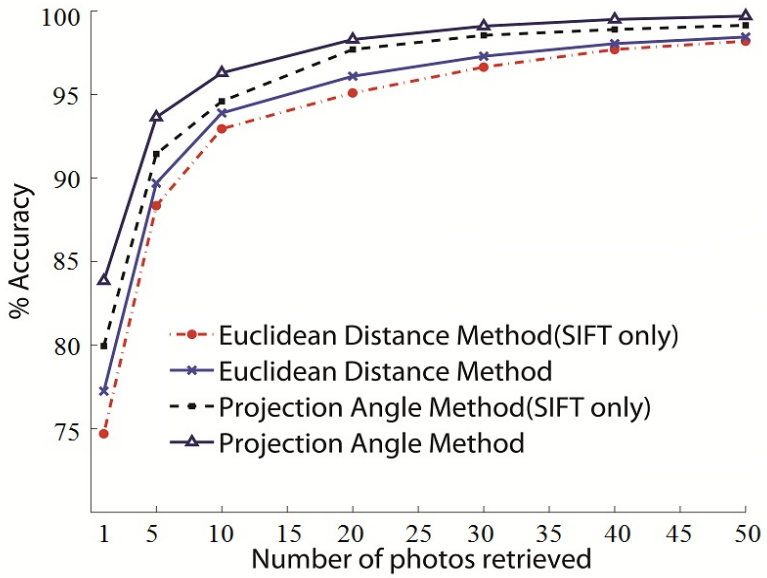

Fig.4 Comparison of Euclidean Distance Method and Projection Angle Method

Figure 5.Poor photo-sketch pairs in FERET

5 Conclusion

Matching or retrieving photos from sketch is a difficult problem. Forensic sketches pose challenges due to inability of a witness to exactly remember the appearance of a suspect which results in inaccurate sketches. Also sketches drawn with pencil has altogether a different modality in comparison with face photos.

Our work presents an alternative approach to retrieve photos from sketch using SIFT descriptors at annotated key points. Our work indirectly highlights the contribution of the information hidden within the geometrical position of eyes, lips, ears etc. in face recognition. Many opportunities for future research stem from the results shown in this work. The proposed approach matches over 100 sketch –photo pairs in less than a minute with a good accuracy. The major contribution of this paper is fusion of linear and non-linear approaches(annotated points and SIFT feature descriptors) which makes real time matching of sketches with photos at low computational cost.

References

1. Yan Ke and Rahul Sukthankar.: PCA-SIFT: A Distinctive Representation for local images descriptors. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04) - Volume 2(2004)
2. Brendan F. Klare, Zhifeng Li, Anil K Jain.: Matching Forensic Sketch to Mug Shot photos. In: IEEE Trans. On Pattern Analysis and Machine Intelligence
3. TF Cootes, CJ Taylor, DH Cooper, J.Graham.:Active Shape Models-their training and application. In: Computer Vision and Machine Understanding Vol. 61, NO 1, January, pp. 38-59. (1995)
4. T.Ahonen, A. Hadid, and M. Pietikainen.: Face description with local binary patterns: Application to face recognition. In: IEEE TPAMI, 28(12):2037. 514,518.(2006)
5. W. Zhang, X. Wang, and X. Tang.: Coupled Information-Theoretic Encoding for Face Photo-Sketch Recognition. In: IEEE Computer Society Conference on Computer Vision and Patter Recognition.
6. D. Lowe.: Distinctive image features from scale-invariant key points. In: IJCV, 60(2):91–110. 514, 518.(2004)
7. K. Mikolajczyk and C. Schmid.: A Performance Evaluation of Local Descriptors," IEEE Trans. Pattern Analysis and Machine Intelligence. In: vol. 27, no. 10, pp. 1615-1630.(Oct. 2005)
8. B. Klare and A. Jain.: Sketch to Photo Matching: A Feature-Based Approach. In: Proc. SPIE Conf. Biometric Technology for Human Identification VII.(2010)
9. X. Tang and X. Wang.: Face Sketch Recognition.In: IEEE Trans. Circuits and Systems for Video Technology, vol. 14, no. 1, pp. 50-57.(Jan. 2004)
10. X. Wang and X. Tang.: Face Photo-Sketch Synthesis and Recognition. In: IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 31, no. 11,pp. 1955-1967. (Nov. 2009)
11. X. Tang and X. Wang.: Face Sketch Synthesis and Recognition. In: Proc. IEEE Int'l Conf. Computer Vision, pp. 687-694.(2003)
12. The University of Manchester, http://personalpages.manchester.ac.uk/staff/timothy.f.cootes/tfc_software.html
13. Pulak Purkait, Bhabatosh Chanda and Shrikant Kulkarni.: A Novel Technique for Sketch to Photo Synthesis. In : 7th International Conference on Computer Vision, Graphics and Image Processing(ICVGIP'10), pp 224-231,Chennai. (December 2010)
14. P. Jonathon Phillips, Harry Wechsler, Jeffrey Huang and Patrick J. Rauss: The FERET database and evaluation procedure for face-recognition algorithms. In: Image Vision Computer. 16(5): 295-306 (1998)
15. The FERET Database http://www.itl.nist.gov/iad/humanid/feret/