Facial Self Similarity for Sketch to Photo Matching

Zohaib Khan, Yiqun Hu and Ajmal Mian
School of Computer Science and Software Engineering
The University of Western Australia
35 Stirling Highway, Perth, 6009
Email: {zohaib,yiqun,ajmal}@csse.uwa.edu.au

***Abstract*—Automatic recognition of suspects from forensic sketches is of considerable interest to the law enforcement agencies. However, this task is complex due to the heterogenous nature of face sketches and photographs. To address this challenge, previous approaches generally learn a transformation of a sketch to photo or a photo to sketch at the image or feature level in order to reduce the modality gap. Such a transformation may be indeterministic and if learned from training data, is likely to over-fit the sketch artist's drawing technique. Instead, we formulate the problem in the context of matching local self similarities which are independently computed from a face sketch and a photo. The proposed Facial Self Similarity (FSS) descriptor is obtained by correlation of a small face patch with its local neighborhood. Thus, our approach avoids the need of a modality transformation, while implicitly reducing the inter-modality gap. The proposed FSS descriptor is evaluated on the CUHK Face Sketch database using sketch-photo pairs of 311 subjects. The FSS descriptor demonstrates high recognition accuracy of 99.53% and outperforms current techniques. We also evaluate the robustness of the descriptor to anomalies such as matching sketches to blurred photographs.**

I. Introduction

A photographic evidence at a crime scene enables the law enforcement agencies to identify criminals by searching the suspect's face in their photographic archives. This task has been made easy by automatic face recognition systems [1]. In case no photographic evidence is retrievable, identification of a suspect may rely solely on the memory of an eye-witness for the production of a forensic sketch. A forensic sketch is usually drawn by an artist on the account of an eye-witness's memory. The forensic sketches are advertised in public so that the suspect may be visually recognized for apprehension.

Recently, there has been an increasing interest in the automation of this task i.e. matching a face sketch to a database of photographs. Although, face photo based recognition has long been investigated and advanced to a significant level, recognition based on matching sketches to photos is a relatively new subject. Matching a sketch to a photograph is a complex task in that the sketch and photo are different modalities and a direct correspondence between the pixel intensity values does not necessarily exist as shown in Fig. 1. This is because a face sketch is usually drawn by repeated line-strokes and shades which differ in texture compared to a photograph. Thus, the major challenge of matching a face sketch to a photograph is to reduce the inter-modality gap. This challenge has recently led to a significant research interest in *Face Sketch Photo Recognition* [2]–[6] wherein a query sketch is matched to a number of photos in a database while reducing the inter modality gap by various methods. The problem can be regarded as a subset of heterogenous face recognition where a probe face is different in modality from that of the gallery face.

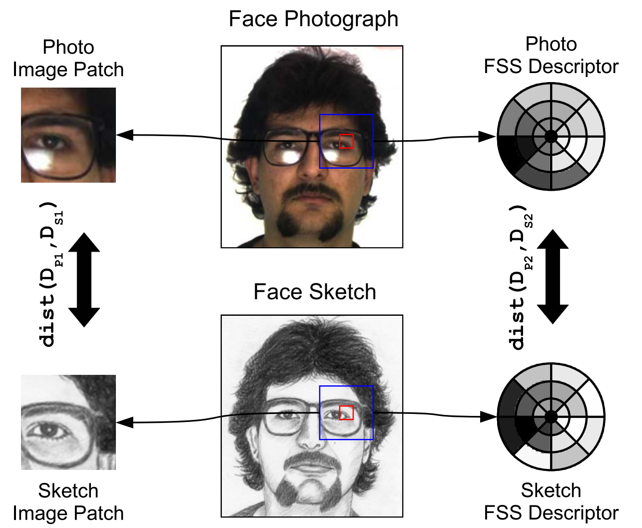

Fig. 1. A face sketch may significantly vary from its actual photo. High photometric variation can be observed between the sketch-photo image pair. Despite the textural variations, the proposed *Facial Self Similarity (FSS)* descriptor is quite similar at corresponding locations.

Previous approaches to face sketch-photo recognition have concentrated on reducing the modality gap by finding a transformation from sketch to photo (pseudo-photo generation) or photo to sketch (pseudo-sketch generation) and subsequently matching with a photo or sketch respectively. Such an ideal transformation from sketch to photo or otherwise may not exist in the first place. This is because the photo and sketch come from two independent sources , i.e. a camera and an artist. Thus, it may not be appropriate to enforce an assumption of the existence of a perfect transformation. Although, an approximate transformation may be learnt from a training set of sketch-photo pairs, but it is likely to over-fit to the intrinsic drawing style of a sketch artist and may not hold good for sketches drawn by other artists.

In our approach, instead of learning a suitable transformation, we extract self similarity based features individually from local regions of sketches and photos. Self similarity features

978-1-4673-2181-5/12/$31.00 ©2012 IEEE

are obtained by correlating a small image patch within its larger neighborhood and therefore, remain relatively invariant to the inter-modality variation. For example, in Fig. 1, the self correlation between the inner and outer patches of the photo is likely to be similar to the correlation between the inner and outer patches in the sketch. Thus, the proposed *Facial Self Similarity (FSS)* descriptor does not require the need of a modality transformation at any stage while it still implicitly reduces the modality gap between the sketch and photo.

II. Related Work

One of the initial noteworthy attempts to address the problem of automatically matching sketches to mug-shot photos was proposed by Uhl and Lobo [7]. They extracted facial features from the sketches followed by a photometric and geometric normalization. Finally, eigen analysis was performed for matching. The algorithm was tested on a small database (of 7 subjects), but provided sufficient evidence to the possibility of automatic retrieval of photos using a query sketch.

Tang and Wang [8] reduced the sketch-photo modality gap by transformation of a sketch into photo (pseudo photo generation) or a photo into sketch (pseudo sketch generation) and then matched with a photo or sketch respectively. Specifically, the photo mean is subtracted from a photo and projected into a face photo eigenspace. The projected photo is then reconstructed using the photo eigenspace, but, adding the sketch means instead of photo means to form a pseudo sketch. Thus a query sketch could be matched to a gallery of pseudo-sketches (generated from photographs). The technique gave better recognition results in comparison to geometrical measures and the traditional eigenface analysis.

Tang and Wang further improved the sketch synthesis process by applying a separate eigen transformation to the face geometry and texture [2]. A synthesized pseudo-sketch was then matched to a query sketch using a Bayesian classifier which also improved slightly over the PCA based classification. Liu et al. [9] synthesized sketches using the idea of Locally Linear Embedding (LLE). Additionally, various classification strategies were investigated and the Nonlinear Kernel Discriminant Analysis (KNDA) was found to be the most effective. In another approach, a photo-sketch model was learnt using a Multiscale Markov Random Field (M-MRF) [4] on local face patches. The learned model was then used for transformation of photos to sketches (pseudo-sketch generation).

Xiao et al. [10] proposed a similar approach based on Embedded Hidden Markov Model (E-HMM) for nonlinear transformation of sketches into photos (pseudo-photo generation). However, the procedure was computationally expensive due to its iterative nature. Bhatt et al. [11] resolved sketches and photos into multi-resolution Laplacian pyramid and applied a variant of LBP descriptor to each pyramid level. A genetic optimization algorithm was used to find the weights of different local regions based on their importance to achieve a better recognition accuracy. Other works which propose different schemes for face sketch synthesis from photos include; sparse representation [12], multi-dictionary sparse representation [13] and local regression models [14].

Yuen et al. [15] proposed a point distribution model for local facial features representation and a geometrical relationship between facial features as a global representation of face sketch and photos. Thus, local and global geometrical features were used to reduce the modality gap. For the identification of a sketch in large photographic databases, relevance feedback was employed by keeping 'Humans in the Loop' to facilitate the recognition of correct identity.

Recent works by Klare and Jain [5], [16] showed that discriminant analysis on local descriptors such as SIFT and LBP is effective for sketch-photo recognition. It is also worth mentioning that for the task of face sketch to photo recognition, a patchwise descriptor computation performs far superior to a descriptor computation at detected keypoints [17]. Zhang et al. [6] proposed a coupled information theoretic encoding based descriptor which aims to encode facial structures in a manner such that similar structures in sketch and photo have the same codes while ensuring that different structures have different codes. The coupled encoding was achieved by an information theoretic projection tree which was extended to a randomized forest for further improving performance.

Nejati and Sim [18] investigated the use of non-artistic face sketches (bearing only the facial component outlines and distinct facial marks) for matching to photographs. Such a crude sketch representation may only take advantage of the geometrical relationship of the facial features and may prove insufficient for discrimination of subjects in large databases. Another important aspect of face sketch-photo recognition is the ability of an artist to bring the sketch as close to the actual photo as possible based upon the witness' description, thereby improving the recognition accuracy [19]. Moreover, there are considerable differences in the way humans recognize a subject from a sketch compared to computer vision systems. This is because humans have the ability to exploit the presence of unique features in a distinct face for recognizing the correct identity [20].

A major focus of face sketch-photo recognition research has been to find a common representation for matching sketch to photo modality. In our approach, we propose the *Facial Self Similarity (FSS)* feature which is invariant to the differences in sketch and photo modalities. It should be noted that our notion of self similarity is different to that of Shectman and Irani's [21] which was mainly used for the task of image retrieval and associated applications. These differences are discussed in Section III-C.

III. Proposed Method

A. Face Normalization

All faces (both sketches and photos) were rotated with respect to the manually located eye coordinates and scaled so that the inter-ocular distance is 25 pixels. A region of 100×100 pixels was then cropped keeping the eyes fixed at row 48 of the image.

B. Preprocessing

A visual observation of the photometric variation between the photos and sketches suggests that the sketches generally comprise of line like texture while the photographs exhibit relatively smooth texture variation. Therefore, we consider it useful to retain the common information from both modalities before the extraction of features. The Difference-of-Gaussian (DoG) filter has been shown to be effective for face photo and sketches [5], [6], reducing the low frequency information and retaining the high frequency information.

The DoG image $D(x, y)$ is obtained by convolving the input image $I(x, y)$ with a DoG filter.

$$D(x, y) = (G(x, y, \sigma_2) - G(x, y, \sigma_1)) * I(x, y) \ , \quad (1)$$

where σ is the standard deviation of the Gaussian filters. We empirically found $(\sigma_1, \sigma_2) = (1, 2)$ to be the best combination. Normalization and preprocessing for a face photo and sketch is shown in Fig. 2

Fig. 2. Face normalization and preprocessing for a photo and sketch. Original Images (top), Normalized Images (center) and Preprocessed Images (bottom).

C. Facial Self Similarity (FSS) Descriptor

Our notion of self similarity in images is inspired by the the *Local Self Similarity (LSS) descriptor* [21]. However, we deviate from the original methodology due to certain limitations of the LSS descriptor in the scenario of face sketch-photo recognition. First, the underlying assumption of LSS descriptor is that there exists a strong statistical co-occurrence of *pixel intensities* across images of similar objects. In contrast, our assumption is based on the co-occurrence of a higher order image representation (Difference of Gaussian image). Second, we propose the usage of `mean` compared to `max` of correlation surface for pooling in histogram bins in order to make it sensitive to the matching part over a small local region instead of a single pixel (as `max` would do). In comparison to the image retrieval problem where a large number of objects with variation in shapes and sizes are retrieved, recognition of faces requires strict spatial correspondence for accurate facial feature matching. Hence, we replace the `log-polar` representation with a `polar` representation to constrain the spatial deformation of face sketches from the original photos. Moreover, we do not eliminate the *uniform* and *salient* descriptor points as they are useful for face discrimination. The *uniform* points come from patches sampled from homogenous regions of a face, whereas, the *salient* points exhibit low self similarity within a local neighborhood. Finally, the descriptors are globally organized and matched using a nearest neighbor classifier as opposed to the computationally expensive *ensemble matching* [21] or *Hough transform based voting* [22].

The procedure to compute the FSS descriptor is composed of two main steps. The first step is to compute a local self similarity surface at a point. The second step is to convert the similarity surface into a polar histogram. The following section gives a detailed description of the proposed FSS descriptor computation.

1) Self Similarity Surface Computation: A uniform triangular grid of N points with a spacing of δ pixels is overlayed on the face image and centralized with respect to the center of eye coordinates. At each grid location (x, y), a square window S_w of side $n_w = 2r_w + 1$ pixels is sampled from the face image $D(x, y)$. Another square patch S_p of side $n_p = 2r_p + 1; r_p < r_w$ pixels is sampled from the center of S_w. The correlation between two images S_p and S_w can be computed by various forms of correlation functions among which the *sum of squared differences* (SSD) can accomplish the task efficiently [21]. The sum of squared differences $\mathbf{S}(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{n_w \times n_w}$ is computed as

$$\mathbf{S}(\mathbf{x}, \mathbf{y}) = \sum_{x=1}^{n_w} \sum_{y=1}^{n_w} (S_p - S_w(x, y))^2 \ , \quad (2)$$

which is a distance surface, however, we are interested in the computation of a similarity surface. Therefore, $\mathbf{S}(\mathbf{x}, \mathbf{y})$ is transformed into a correlation similarity surface $\mathbf{C}(\mathbf{x}, \mathbf{y})$ as [21]

$$\mathbf{C}(\mathbf{x}, \mathbf{y}) = \exp^{-\left(\frac{\mathbf{S}(\mathbf{x}, \mathbf{y})}{\max(\sigma_n, \sigma_a)}\right)} \ , \quad (3)$$

where σ_n is the estimated photometric noise variance and σ_a is the local variance. Due to the texture difference between sketch and photo modalities, the correlations of similar regions may be differently scaled. Thus, we normalize $\mathbf{C}(\mathbf{x}, \mathbf{y})$ by dividing it with its `norm`.

$$\hat{\mathbf{C}}(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{C}(\mathbf{x}, \mathbf{y})}{\|\mathbf{C}(\mathbf{x}, \mathbf{y})\|} \ . \quad (4)$$

2) Polar Histogram Conversion: The similarity surface is pooled into a polar grid of n_θ angular and n_ρ radial intervals as shown in Fig.3. We use the `mean` value of the points falling

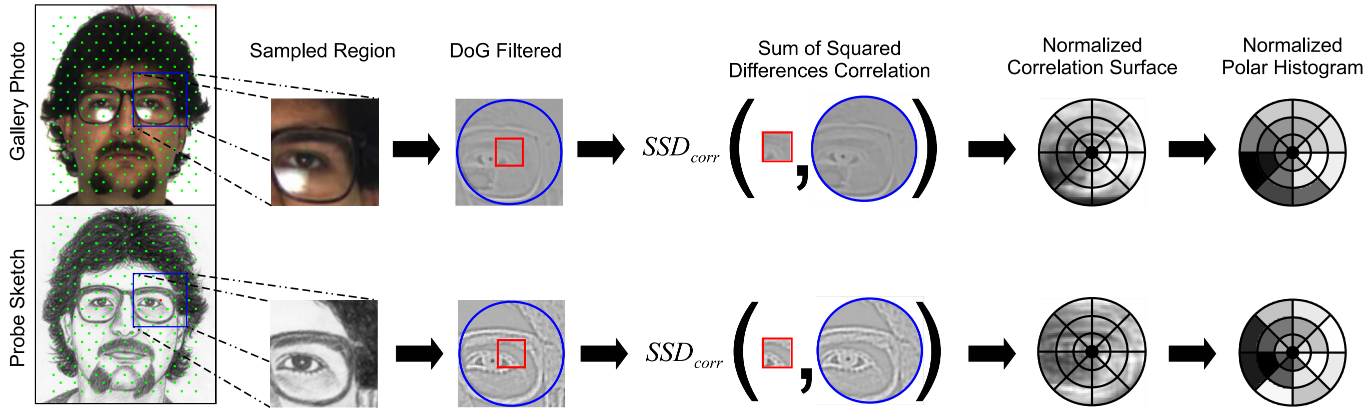

Fig. 3. An example illustration of the FSS descriptor computation for a gallery photo and a probe sketch. Square patches are sampled at uniform spacing in the image (overlayed grid in green). The DoG filter is applied to each patch to retain useful information. An inner patch (red) is correlated with the outer window (blue) in a circular vicinity. The resulting correlation surface is normalized and divided into n_θ angular and n_ρ radial intervals (here, $n_\theta = 8$ and $n_\rho = 3$). Each bin value in the final descriptor is the average of the correlation pixels in that bin. The descriptor is normalized via min-max rule. (best viewed in color)

in a bin b, so as to get a vector $d \in \mathbb{R}^B$, where $B = n_\theta \times n_\rho$

$$d = \frac{1}{n_b} \sum_{\forall (x,y) \in b} \hat{\mathbf{C}}(\mathbf{x}, \mathbf{y}) \ , \tag{5}$$

where n_b is the number of pixels falling in each bin b. The descriptor is then normalized in the range $[0, 1]$ using the `min-max` rule

$$\hat{d} = \frac{d - \min(d)}{\max(d) - \min(d)} \ , \tag{6}$$

where $\hat{d}$ is the final FSS descriptor. Computation of the descriptor at all N sampling points gives $\mathbf{D} = \{\hat{d}_1, \hat{d}_2, ..., \hat{d}_N\}$ which is a global representation of all the local FSS descriptors.

D. FSS Descriptor Matching

Most of the previous works based on self similarity descriptor were used for retrieval of objects in images e.g. cars, hearts etc. Such objects exhibit high intra-class variations in addition to the inter-class variation. In case of sketch to photo recognition all images are frontal faces in a neutral expression. Consequently, the inter class variation is expected to be lower and hence, an accurate spatial correspondence is required between the descriptors sampled from the same locations of a face sketch and a photo. This is accomplished by matching the descriptors using nearest neighbor classification. Given $\mathbf{D}_p, \mathbf{D}_g \in \mathbb{R}^{B \times N}$, the descriptor matrices for a probe sketch and gallery photo of N locally computed descriptors arranged as column vectors, the class y of a probe is determined as the i^{th} gallery photo that corresponds to its nearest neighbor

$$y = arg \min_i \ \mathtt{dist}(\mathbf{D}_p, \mathbf{D}_g^i) \ , \tag{7}$$

where `dist` is the *Euclidean distance* function.

IV. Experimental Results

A. Database

The CUHK Face Sketch (CUFS) database comprises 188 subjects from the CUHK student database [3], 123 subjects from the AR database [23] and 295 subjects from the XM2VTS database [24], making a total of 606 subjects[1]. However, results can only be reported on 311 subjects[2]. Each subject with a face photograph has a corresponding sketch drawn by an artist. All photos are in a frontal pose, uniform lighting and neutral expression. The sketches were drawn by the artists as viewed from the photos. This dataset has frequently been used as a benchmark in face sketch-photo recognition [2]–[6]. A few example images are shown in Fig. 5.

Fig. 5. Photo and sketch image samples from the CUFS database.

[1]CUHK Face Sketch database: http://mmlab.ie.cuhk.edu.hk/facesketch.html

[2]According to the XM2VTS database license agreement, it cannot be merged to form a part of another database. It was confirmed from the database owner that its use in CUFS database is an infringement of the XM2VTS database usage terms. Therefore, we report experimental results on the CUFS database excluding the XM2VTS part.

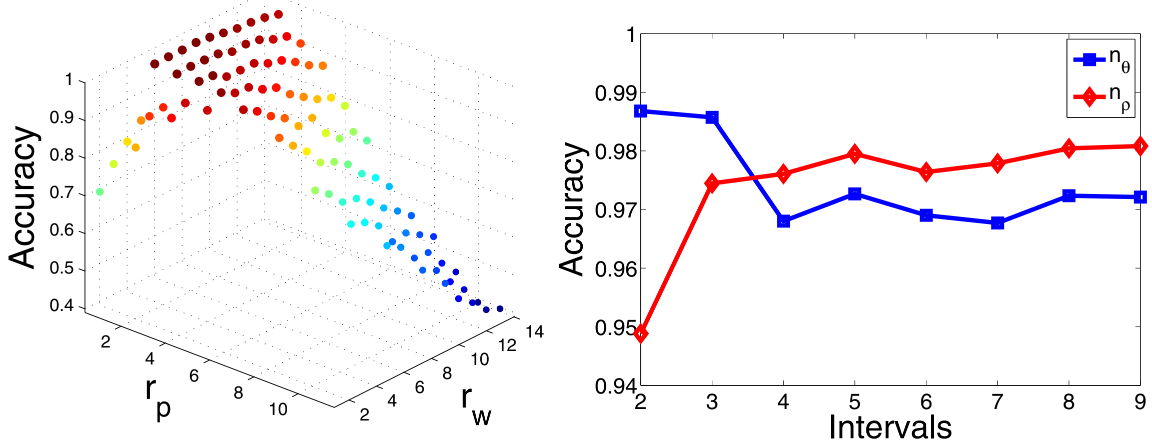

Fig. 4. Variation in accuracy with respect to (a) Descriptor window sizes (r_p and r_w) (b) Polar histogram parameters (n_θ and n_ρ)

First, we find the appropriate values of the descriptor parameters by performance evaluation on a training set. A training set of 161 subjects is randomly selected from the database and the accuracy of the descriptor for a given parameter value is computed in terms of the Rank-1 recognition rate. This process is repeated 10 times resulting in different random selections of subjects in the training set and the average accuracy over the 10 experiments is referred to as the accuracy for that parameter value. The same procedure is followed by repeating the above experiment for a different parameter value.

B. FSS Descriptor Parameters

We evaluate the sensitivity of the FSS descriptor to mainly four free parameters. The self similarity surface dimension, which is dependent on the choice of (r_p, r_w) and the polar histogram dimension, dependent on the choice of (n_θ, n_ρ).

We first find the size of the inner window n_p (determined by r_p) and outer window n_w (determined by r_w), which maximizes the average accuracy on the training set. We bound $r_p = \{1, 2, 3, ..., 10\}$ and $r_w = \{r_p + 1, ...15\}$ for evaluation. As evident from Fig. 4b, the accuracy is more sensitive to the inner patch size r_p compared to the sampled window size n_w. The accuracy drops significantly with the increase in r_p throughout the range. On the other hand, the accuracy first increases with the increase in r_w and then decreases gradually. It should be noted that the sizes of (S_p, S_w) determine the descriptor computation complexity and larger values may affect the overall efficiency. According to these observations we determine $r_p = 1$ and $r_w = 7$ as an affordable choice.

Next, we find appropriate values for the parameters (n_θ, n_ρ) that determine the polar histogram dimensions in the range $(n_\theta, n_\rho) = \{2, 3, ..., 9\}$. It can be seen in Fig. 4a that the accuracy first decreases from $n_\theta = 2$ to 4 and then stabilizes at a certain level. In contrast, the accuracy first increases from $n_\rho = 2$ to 3, and then remains fairly constant. Given these observations and our aim to maximize accuracy, $(n_\rho, n_\theta) = (2, 3)$ seems a suitable choice for determining the descriptor dimensions.

TABLE I
RANK-1 RECOGNITION RATES OF VARIOUS METHODS ON THE CUFS DATABASE

Method	Rank-1 RR(%)
PCA Eigenface [4]	6.3
Locally Linear Embedding + Kernel NDA [9]	87.67
FaceVACS [16]	90.37
Extended Uniform Circular LBP + GA [11]	94.12
Multiscale Markov Random Field + RS-LDA [4]	96.30
SIFT Descriptor Based [16]	97.87
SIFT+MLBP based Discriminant Analysis [5]	99.47
Facial Self Similarity [3]	99.53

C. Recognition Results and Comparison

Using the FSS descriptor parameters found from the training set, we compute the Rank-1 recognition rates averaged over 10 random selections of a test set comprising 150 subjects. The Rank-1 recognition rates for various techniques proposed in the literature are presented in Table I.

It can be observed that the PCA Eigenface algorithm is unable to cope with the heterogenous nature of sketch and photos and hence demonstrates very low recognition rate. The commercial face recognition engine (FaceVACS) which is predominantly expected to perform accurately on photo-photo matching, shows intermediate performance on sketch-photo matching. The Extended Uniform Circular LBP improves at the cost of computation complexity by making use of the genetic algorithm for weighted matching of descriptors. The multiscale MRF based photo-sketch transformation has significant edge over the Locally Linear Embedding based photo-sketch transformation as it is able to handle non-linearity between sketch and photo images to some extent. However, discriminant analysis on local features such as SIFT and SIFT+MLBP show relatively more promising results for sketch to photo matching. The proposed FSS descriptor out-

[3] Other results are reported on a dataset containing additional images from the XM2VTS database. However, we are not permitted by the database owners to report the results on the full CUHK database containing the XM2VTS part.

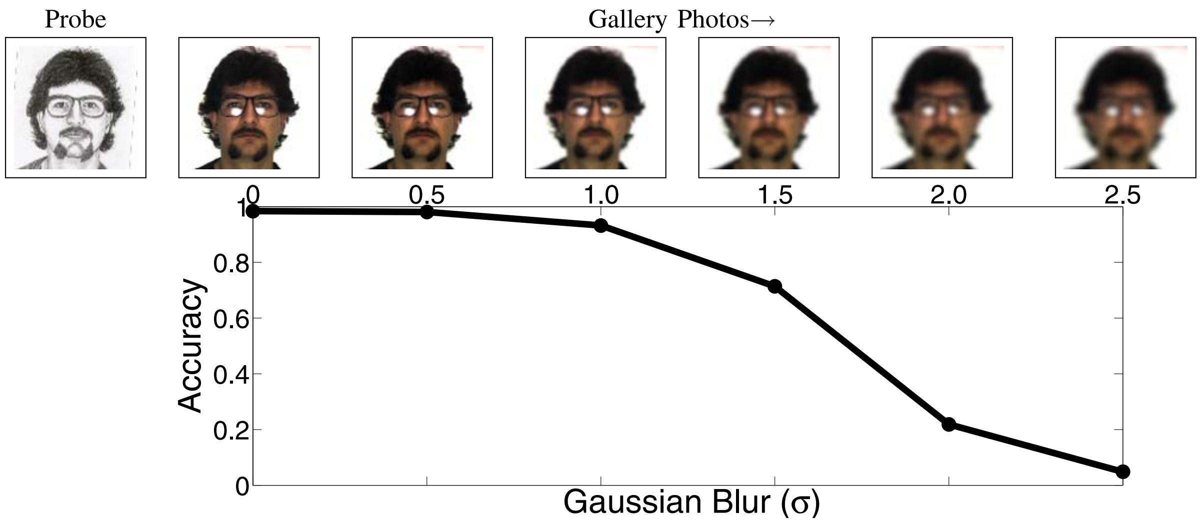

Fig. 6. Effect of blurring on sketch to photo matching using the FSS descriptor

performs all techniques which shows that it is better capable of dealing with the heterogenous nature of sketches and photos.

D. Robustness to Blurring

In practical scenarios e.g. surveillance, the retrieved face photos sometimes come out to be blurred. In this section, we test the robustness of the proposed descriptor to blurring. We sequentially introduce Gaussian blurring in the gallery photos using $\sigma_k = \{0.5, 1.0, ..., 2.5\}$ and repeat the recognition experiments. The results in Fig. 6 indicate only a 7% degradation in accuracy with an increase in the amount of image blur upto $\sigma = 1.0$. Beyond $\sigma = 1.0$, the accuracy drops fairly to a minimum at $\sigma = 2.5$. Thus it can be inferred that the proposed FSS descriptor can successfully match sketches to blurred photographs upto $\sigma = 1.0$.

V. Conclusion

We presented the Facial Self Similarity feature for face sketch to photo recognition and demonstrated that the extraction of self similarities independently from sketches and photos holds more promise compared to image or feature transformations. Experiments on the CUHK Face Sketch database show the efficacy of the proposed approach in matching face sketches to photos with high recognition accuracy. The proposed descriptor can be extended to other heterogenous object recognition scenarios as it is extracted independently from the underlying modalities.

Acknowledgement

This research was supported by ARC Grant DP110102399. The first author was an IPRS scholar in Australia. The authors thank the Chinese University of Hong Kong (CUHK) for providing the CUFS database. The authors would also like to thank Wei Zhang and Brenden Klare for their useful responses.

References

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

[2] X. Tang and X. Wang, “Face sketch synthesis and recognition,” in *Proc. Int. Conf. on Computer Vision (ICCV)*, 2003, pp. 687–694.

[3] ——, “Face sketch recognition,” *IEEE Trans. on Circuits and Systems for Video Technology (CSVT)*, vol. 14, no. 1, pp. 50–57, 2004.

[4] X. Wang and X. Tang, “Face photo-sketch synthesis and recognition,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 1955–1967, 2009.

[5] B. Klare, Z. Li, and A. K. Jain, “Matching forensic sketches to mug shot photos,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, pp. 639–646, 2011.

[6] W. Zhang, X. Wang, and X. Tang, “Coupled information-theoretic encoding for face photo-sketch recognition,” in *Proc. Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 513–520.

[7] R. G. U. Jr. and N. da Vitoria Lobo, “A framework for recognizing a facial image from a police sketch,” in *Proc. Computer Vision and Pattern Recognition (CVPR)*, jun 1996, pp. 586–593.

[8] X. Tang and X. Wang, “Face photo recognition using sketch,” in *Proc. Int. Conf. on Image Processing (ICIP)*, 2002, pp. I–257–I–260.

[9] Q. Liu, X. Tang, H. Jin, H. Lu, and S. Ma, “A nonlinear approach for face sketch synthesis and recognition,” in *Proc. Computer Vision and Pattern Recognition (CVPR)*, 2005, pp. 1005–1010.

[10] B. Xiao, X. Gao, D. Tao, and X. Li, “A new approach for face recognition by sketches in photos,” *Signal Processing*, vol. 89, no. 8, pp. 1576–1588, 2009.

[11] H. S. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa, “On matching sketches with digital face images,” in *Proc. Biometrics: Theory Applications and Systems (BTAS)*, 2010, pp. 1–7.

[12] L. Chang, M. Zhou, Y. Han, and X. Deng, “Face sketch synthesis via sparse representation,” in *Proc. International Conference on Pattern Recognition (ICPR)*, 2010, pp. 2146–2149.

[13] N. Wang, X. Gao, D. Tao, and X. Li, “Face sketch-photo synthesis under multi-dictionary sparse representation framework,” in *International Conference on Image and Graphics (ICIG)*, 2011, pp. 82–87.

[14] N. Ji, X. Chai, S. Shan, and X. Chen, “Local regression model for automatic face sketch generation,” in *Int. Conf. on Image and Graphics (ICIG)*, 2011, pp. 412 –417.

[15] P. Yuen and C. Man, “Human face image searching system using sketches,” *IEEE Transactions on SMC Part A: Systems and Humans*, vol. 37, no. 4, pp. 493 –504, 2007.

[16] B. Klare and A. K. Jain, “Sketch-to-photo matching: a feature-based approach,” *SPIE, Biometric Technology for Human Identification VII*, vol. 7667, 2010.

[17] A. Gupta and K. A. Pati, “Sketch to photo matching,” Indian Institute of Technology, Kanpur, Tech. Rep., nov. 2010.

[18] H. Nejati and T. Sim, “A study on recognizing non-artistic face sketches,” in *Proc. IEEE Workshop on Applications of Computer Vision (WACV)*, 2011, pp. 240–247.

[19] J. Zhong, X. Gao, and C. Tian, “Face sketch synthesis using E-HMM and selective ensemble,” in *Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2007, pp. I–485–I–488.

[20] Y. Zhang, S. Ellyson, A. Zone, P. Gangam, J. Sullins, C. McCullough, S. Canavan, and L. Yin, “Recognizing face sketches by a large number of human subjects: A perception-based study for facial distinctiveness,” in *IEEE International Conference on Automatic Face Gesture Recognition and Workshops*, march 2011, pp. 707–712.

[21] E. Shechtman and M. Irani, “Matching local self-similarities across images and videos,” in *Proc. Computer Vision and Pattern Recognition (CVPR)*, 2007, pp. 1–8.

[22] K. Chatfield, J. Philbin, and A. Zisserman, “Efficient retrieval of deformable shape classes using local self-similarities,” in *Int. Conference on Computer Vision Workshops*, 2009, pp. 264 –271.

[23] A. Martinez and R. Benavente, “The AR face database,” Computer Vision Center (CVC), The Ohio State University, Tech. Rep. 24, 1998.

[24] K. Messer, J. Matas, J. Kittler, and K. Jonsson, “XM2VTSDB: The Extended M2VTS Database,” in *International Conference on Audio and Video-based Biometric Person Authentication*, 1999, pp. 72–77.