Sketch to Photo Matching: A Feature-based Approach

Article *in* Proceedings of SPIE - The International Society for Optical Engineering · April 2010

DOI: 10.1117/12.849821

CITATIONS
73

READS
830

2 authors, including:

Brendan Klare
Rank One Computing
38 PUBLICATIONS **3,231** CITATIONS

SEE PROFILE

Sketch to Photo Matching: A Feature-based Approach

Brendan Klare[a] and Anil K Jain[a,b]

[a]Department of Computer Science and Engineering
Michigan State University
East Lansing, MI, U.S.A
[b] Department of Brain and Cognitive Engineering
Korea University
Seoul, Korea

ABSTRACT

This paper presents a local feature-based method for matching facial sketch images to face photographs, which is the first known feature-based method for performing such matching. Starting with a training set of sketch to photo correspondences (i.e. a set of sketch and photo images of the same subjects), we demonstrate the ability to match sketches to photos: (1) directly using SIFT feature descriptors, (2) in a "common representation" that measures the similarity between a sketch and photo by their distance from the training set of sketch/photo pairs, and (3) by fusing the previous two methods. For both matching methods, the first step is to sample SIFT feature descriptors uniformly across all the sketch and photo images. In direct matching, we simply measure the distance of the SIFT descriptors between sketches and photos. In common representation matching, the distance between the descriptor vectors of the probe sketches and gallery photos at each local sample point is measured. This results in a vector of distances across the sketch or photo image to each member of the training basis. Further recognition improvements are shown by score level fusion of the two sketch matchers. Compared with published sketch to photo matching algorithms, experimental results demonstrate improved matching performances using the presented feature-based methods.

Keywords: face recognition, feature-based, sketch, fusion, SIFT

1. INTRODUCTION

The true identity of an individual is invaluable information. While the average person has no qualms with their identity being known, a collection of individuals would prefer to keep such knowledge hidden despite the negative impact it may cause on the population at large. Typically, the sole motivation for an individual to hide his identity is to evade detection by law enforcement agencies for some type of criminal activity. Ongoing progress in biometric recognition has offered a crucial method to help ascertain who a person truly is.

The three most popular biometric traits in use are the fingerprint, face, and iris. Though fingerprint and iris are generally considered more mature and accurate biometric technologies, face recognition is now receiving a significant amount of interest in the research community. The two main reasons for a growing interest in face biometrics are: (i) unlike fingerprint and iris, faces can be captured in a covert way, so it is an extremely valuable biometric for surveillance applications. With the rapidly growing number of digital cameras capturing data in public areas, having a robust and accurate face recognition method is critical to apprehend suspects and prevent crimes. (ii) Solving unconstrained face recognition requires a significant amount of research in face modeling, feature extraction and matching. The past two decades have witnessed a tremendous progress in face

Further author information:
B.K.: E-mail: klarebre@msu.edu,
A.K.J.: E-mail: jain@cse.msu.edu

The authors would like to thank Xiaoou Tang and Xiaogang Wang for providing the sketch images used in the paper, and Zhifeng Li for his valuable comments.
Anil Jain's research was partially supported by WCU (World Class University) program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology(R31-2008-000-10008-0).

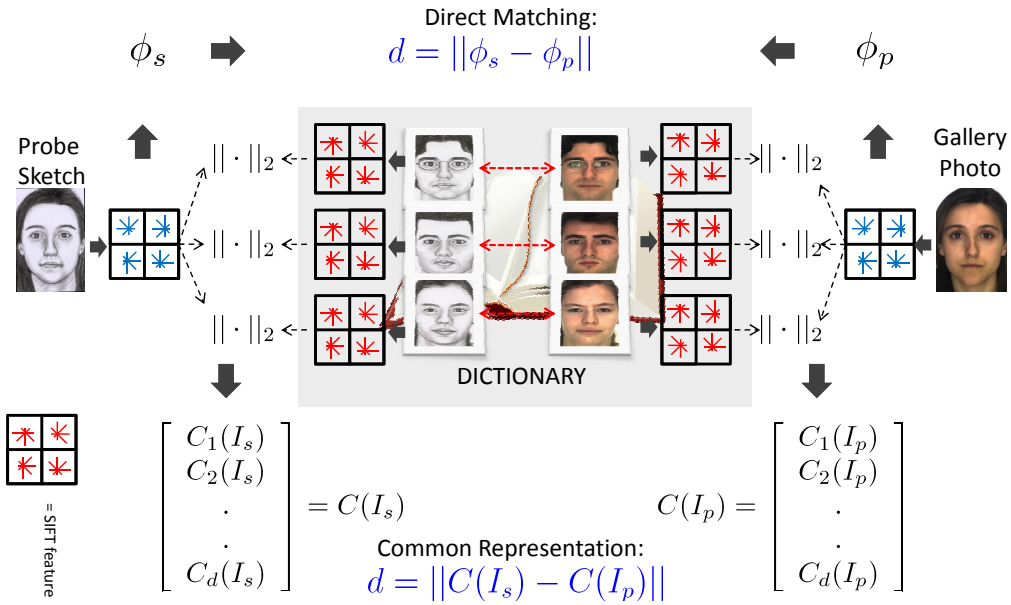

Figure 1. The process of comparing a face sketch to a face photo using both the common representation and the direct matching method is illustrated here. The first step is to compute the SIFT representation for each image (Section 3.1). Direct matching (top) proceeds by computing the distance of the SIFT representation between the sketch and photo. For the common representation (bottom), we describe the probe sketch image as a d-dimensional vector, where d is the product of the number of subjects in the dictionary (n) and the number of patches sampled to generate the SIFT features (p). The d vector components are the L_2 distances from the p sampled SIFT descriptors of the sketch to the same descriptors for each of the n sketches in the training set. The same process is then applied to the gallery photos, this time comparing them to the training set photos. The sketches and photos can then be directly compared in this common representation.

recognition algorithms. Turk and Pentland's holistic Eigenface matching algorithm[1] served as the precedent for modern face recognition engines. Since the introduction of the Eigenface algorithm almost 20 years back, face recognition accuracy has increased by orders of magnitude,[2] to the point where the face recognition rates under controlled imaging conditions (good ambient lighting, frontal pose, neutral expression and uniform background) are comparable to fingerprint and iris matching rates. Unfortunately, real world face recognition scenarios do not satisfy such controlled conditions. This has prompted researchers to turn the focus of their research in face recognition to more difficult problems such as varying illumination,[3] non-frontal pose[3] and occlusion.[4]

A new problem of interest in face recognition that has emerged deals with matching a sketch image of an individual to his photograph. The consequence of being able to develop a robust algorithm for matching sketches to photos is of great value to law enforcement agencies. When a witness sees a crime being committed, in many instances, the *verbal description* of the suspect provided by the witness is used by a police artist to create a sketch. Many criminals have been apprehended when they are identified by citizens based on such sketches. State of the art face recognition engines are not able to successfully match a sketch to the photographs stored in law enforcement databases (gallery images). Note that *sketch to photo matching capability implies a camera is not always necessary to capture the face biometric.*

In this paper, we present a new algorithm for matching a sketch to a photo. The proposed method differs significantly from published approaches[5–8] in that we use a local feature-based representation to compare sketches and photos. Previous approaches only performed holistic matching on sketches that were transformed to photographs (or vice-versa) using either a linear transformation directly on the intensity image[5,6] or by generating a synthetic photograph.[7,8] Because the proposed matching algorithm is a local feature-based matcher, it can be used in conjunction with the aforementioned holistic algorithm for hybrid matching.[9]

In order to compare the similarity between a sketch and a photo, we first represent each image using a

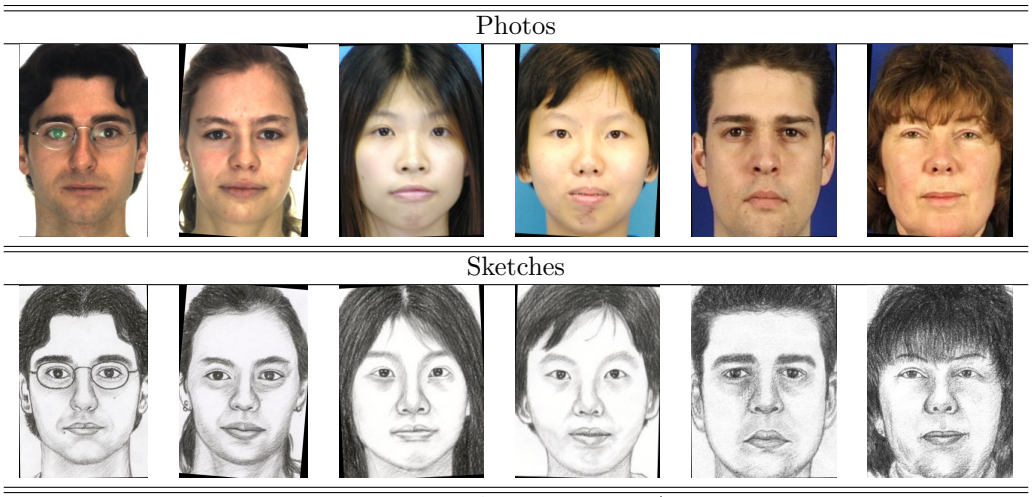

Figure 2. Examples of corresponding sketch photo pairs. A set of n sketch/photo pairs are used in our matching algorithm to build a common representation for both probe sketches and gallery photos. These n pairs are referred to as the training set.

SIFT-based feature descriptor at uniformly sampled patches across the face. Once the sketch and photo images are described using these descriptors we have two separate ways to match them. The first method is to measure the normed distance directly between the set of SIFT descriptors describing each sketch and photo. The second method seeks to describe the sketch and photos in a common representation by first measuring the distance between the sketch and photo from a training set of known sketch/photo correspondences. Using this distance vector, where each component denotes the descriptor distance at a particular patch from every training image in their corresponding domain (sketch or photo), we can then directly compare a sketch and photo. An overview of the common representation approach is shown in Figure 1. Both of these methods are substantially different from previous sketch/photo matching algorithms, yet offer improved matching performance.

The remainder of the paper is organized as follows. In Section 2 we briefly review the previous research that has been conducted in sketch/photo matching. Our method for sketch/photo matching is then detailed in Section 3. Matcher fusion is discussed in Section 4. The results of our experiments using this method can be found in Section 5, followed by concluding remarks in Section 6.

2. PRIOR WORK

Much of the existing work in sketch/photo matching has been performed by Tang et al.[5–8] Early approaches by Tang and Wang[5,6] used a global linear transformation to convert a sketch to a photograph. This conversion was performed on the entire image by projecting the eigenspace representation of the image pixels in the sketch domain to the eigenspace of photo domain.[5] Later Tang and Wang[6] extended this method by separating the shape and texture of the image. Transformations were then applied to the texture and shape of the face separately.

Liu et al.[7] generated a synthetic photograph from a sketch by first breaking the sketch into a set of overlapping patches. For each sketch patch, the k nearest sketch patches (based on Euclidean pixel distance) from a training set of sketch/photo correspondences are selected as candidates. Each patch is then converted from the sketch domain to the photo domain by first solving the set of weights for the k candidate patches that minimize the reconstruction error from the sketch patch being converted. Next, these weights are applied to the photo patch counterparts of the sketch candidate to construct a photo estimate of the sketch patch.

Wang and Tang[8] improved the photo synthesis algorithm of Liu et al.[7] by modeling the faces with a Markov random field (MRF). The sketch to be converted into a photo was broken into a set of overlapping patches. These patches were then modeled as the observed nodes in the MRF, and each node is connected to a hidden photo node counterpart to be estimated. The possible states for the hidden photo nodes were the photo counterparts of the k nearest training sketch patches of the observed sketch patch. The solution to the MRF was estimated using the belief propagation algorithm,[10] and solution patches were stitched together to yield a synthetic photograph

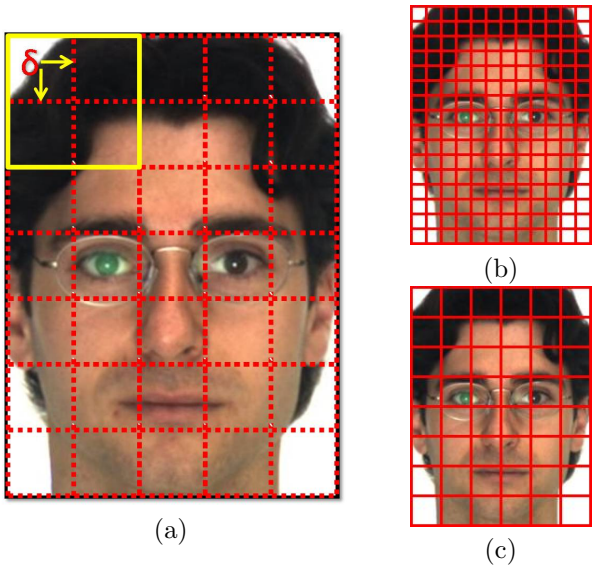

Figure 3. The SIFT sampling scheme. (a) The solid window is the initial placement of the window used to compute the SIFT feature descriptor. This window is then moved over the image in a raster scan fashion, being displaced by δ pixels each time. In this case, δ is half the size of the window. (b) Sampling the face with window size $s = 16$. (c) Sampling the face with window size $s = 32$. In (b) and (c), $\delta = s$ for display purposes, but in our experiments we used the value $\delta = s/2$.

representation. Global face matching algorithms were used to match the synthetic photograph to a gallery of photos.

Li et al.[11] matched sketches and photos using a method similar to the eigentransform presented by Tang and Wang,[5] focusing on converting sketches to photos (as opposed to converting photos to sketches). Zhang et al.[12] compared the performance of human sketch recognition to automated sketch recognition and showed the benefit of using multiple sketches drawn by different artists.

Though the ability to match sketches has consistently improved through the methods presented above, limitations in sketch matching still exist. One such limitation is that no existing algorithm has been shown to work using local feature matching methods. By limiting the sketch matching solutions to global matching algorithms certain discriminating information in the sketches may be discarded. A second drawback to the prior methods is that as the performance of the algorithms have increased so has the complexity of both the runtime and the implementation. The feature-based sketch matching algorithm we present offers both a simpler and faster design, and utilizes localized sketch information in the matching process.

3. FEATURE-BASED SKETCH/PHOTO MATCHING

Given a gallery containing one photograph per subject in a population, this section details two separate feature-based methods for matching a query sketch to the photograph of the true identity of the subject.

The chief requirement for our second matching algorithm (Section 3.3) is a dictionary, or training set, of sketch/photo correspondences. That is, we require a set of n dictionary entries that contain both a sketch and photograph of the same individual. An example of such correspondences can be seen in Figure 2.

3.1 SIFT Representation

For each sketch query, gallery photograph, and each sketch/photo correspondence in our dictionary, we compute a SIFT feature representation. SIFT based object matching is a popular method for finding correspondences between images. Introduced by Lowe,[13] SIFT object matching consists of both a scale invariant interest point detector as well as a feature-based similarity measure. Our method is not concerned with the interest point

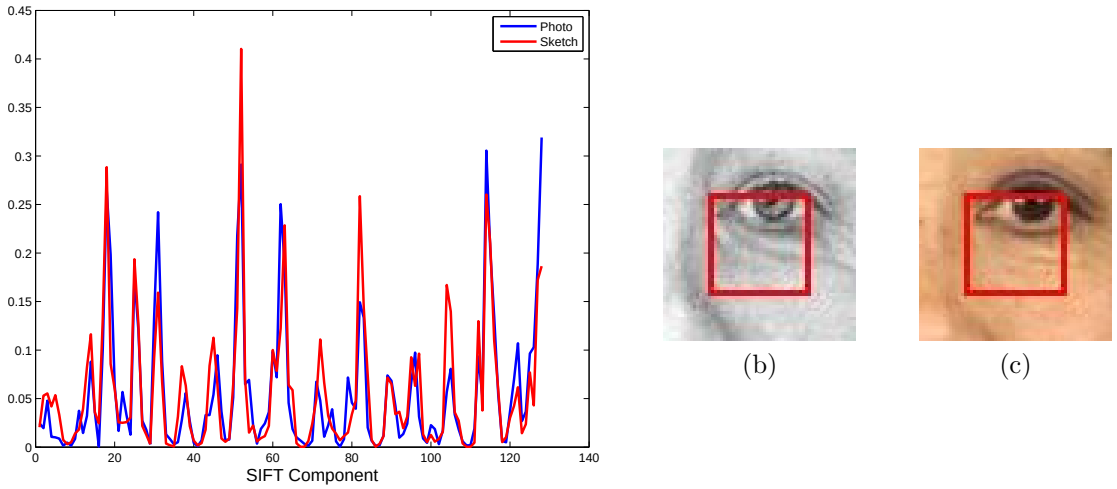

Figure 4. The similarity between sketch and photo image patches of the same person ($s = 32$). (a) Plots of the SIFT descriptor computed of the sketch image (b) and the photo image (c). The SIFT descriptor was computed within the solid box of the sketch and photo. The two descriptors exhibit high levels of similarity despite being computed in different image domains.

detection aspect of the SIFT framework, but instead utilizes only the gradient-based feature descriptors (known as SIFT-features).

A SIFT image feature is a compact vector representation of an image patch based on the magnitude, orientation, and spatial vicinity of the image gradients. For an s x s sized patch of image pixels, the SIFT feature vector is computed as follows. First, the intensity image is used to compute the gradient image, which is weighted by a Gaussian kernel using $\sigma = s/2$. The spatial coordinates in the gradient image are then coarsely quantized into m x n values (generally such that $m = n$). With each gradient image pixel containing a gradient orientation ranging from $[0, \pi)$, the values are then quantized into one of k orientations. At each of the $m \cdot n$ spatial coordinates, the sum of the Gaussian weighted gradient magnitude values for each of the k orientations is computed. This yields a $(m \cdot n \cdot k)$-dimensional feature descriptor, where each component contains the sum of weighted gradient magnitudes at the given location and orientation. The final step is to normalize the feature vector to unit length. A second normalization step is performed by suppressing any component larger than 0.2 down to 0.2 and re-normalizing the vector to unit length. Typical parameters used in this process are $m = 4$, $n = 4$, and $k = 8$, which results in a 128-dimensional vector. These are also the parameter values used in our algorithm.

The process of computing SIFT feature descriptors described above is used for a single image patch. Because we are dealing with high resolution face images (with respect to typical SIFT descriptor patch sizes), we are able to compute many such features from a single image. To do so, we sample SIFT feature vectors from the face image uniformly. Starting in the upper left corner of the image, we slide an s x s sized window across the image in a raster scan fashion, displacing the location of the window by δ pixels each time (see Figure 3). Given a W x H image, an $M \cdot N$ SIFT descriptor set is then computed at each sampling point, where $W = s + \delta(M - 1)$ and $H = s + \delta(N - 1)$. The final result is a set of p 128-dimensional vectors for each face image ($p = M \cdot N$). For an image I, we will refer to the 128-dimensional SIFT descriptor for the jth patch ($1 \leq j \leq p$) as $\tau_j(I)$.

3.2 Sketch/Photo Direct Matching

We initially believed that direct matching between sketches and photos using the SIFT descriptors would not be successful because the gradient images generated from each image domain are not the same. This initial (and incorrect) belief motivated our development of the common representation vector described in Section 3.3. However, further investigation demonstrated that directly matching sketches and photos described by SIFT descriptors was highly successful (see Figure 4).

For direct matching, the distance between a sketch and photo is $d = ||\phi_s - \phi_p||_2$, where ϕ is a $(128 \cdot M \cdot N)$-dimensional vector of all the $M \cdot N$ SIFT descriptors sampled across the face (Figure 3) concatenated together. The subscript s is for the sketch image and p for the photo.

3.3 Sketch/Photo Common Representation Matching

In this section we detail a common representation feature vector that describes a sketch and photo in a common feature space. This common representation is based on the assumption that if a given query sketch image contains similar feature values to a sketch image in the training set, then a photo of the same query subject will also contain similar values to the corresponding photo in the dictionary. Contrarily, sketches that have very different feature values will have photo counterparts that also be dissimilar.

After computing the SIFT features for each sampled patch of the dictionary, gallery, and probe sketches, the common representation $C(I_s)$ for sketch image I_s, and $C(I_p)$ for photo I_p is

$$C_{j,k}(I_s) = ||\tau_k(I_s) - \tau_k(T_j^s)||_2 \quad (1)$$

$$C_{j,k}(I_p) = ||\tau_k(I_p) - \tau_k(T_j^p)||_2 \quad (2)$$

where T_j^s and T_j^p are the jth sketch and photo images in the training set, respectively. At this point we have the Euclidean distances of the SIFT descriptor vectors at each patch of an image (from either the sketch or photo domain) to each dictionary image from the same image domain. Thus, we define the image I in a domain-independent, d-dimensional vector ($d = M \cdot N \cdot n$). To compare the similarity between photo I_p and sketch I_s, we can now directly compare the distance between the two vectors $C(I_p)$ and $C(I_s)$. As a post-processing step, we normalize the lengths of the vector $C(\cdot)$ to unit length. This is an important step so that the vectors represent their relation to the dictionary basis accurately.

The common representation presented here is a generic method for matching across image domains. In the case of sketch and photo domains, it turned out that the SIFT descriptors were very similar across the two domains. However, other problems involving matching across image domains may not have an image descriptor that is largely invariant to the change of the domains. Given a set of known correspondences across the two domains (our training set), the common representation method allows for a direct comparison between the modalities.

4. MATCHER FUSION

We are able to further improve the ability to match sketches to photos using biometric fusion. Specifically, we use sum of score fusion[14] for multiple matchers.

Score level fusion is performed after computing the distance between all probe and all gallery subjects.[14] Given a gallery population size of n and m probe sketches, let the distances of probe i to all gallery members, using matcher j, be the n-dimensional vector D_i^j. For all m probe subjects we have an m x n score matrix D^j. If the distances between matchers being fused are not computed in the same feature space then distances must be normalized, which is the case between our two matchers. We performed min-max normalization on each score matrix (Equation 3), where $\hat{D}^j$ is the normalized scores for matcher j.

$$\hat{D}^j = (D^J - \min\{D^j\})/(max\{D^j\} - min\{D^j\}) \quad (3)$$

Let the superscript d denote the direct matcher (Section 3.2), and superscript c denote the common representation matcher (Section 3.3). Using sum of score fusion, the fused match score given the two matchers is

$$D^{\text{fuse}} = \hat{D}^d + \hat{D}^c \quad (4)$$

Table 1. A comparison of Rank-1 matching performance of various sketch/photo matching algorithms from the first experiment with 300 test sketches. Our results are the average performance over 5 random splits of 100 training pairs and 300 test sketches.

Matching Algorithm	Rank 1 Accuracy (%)
FaceVACS	90.37
Eigen-transform[5]	90.00
BP Synthesis[8]	96.30
Feature-based, Direct	97.00
Feature-based, Common	96.47
Feature-based, Fusion	**97.87**
Feature-based, Fusion + FaceVACS	99.73

4.1 Multi-scale Representation

Using different SIFT window sizes s represents the face image at different scales. Empirically, we found the highest success using the values $s = 16$ and $s = 32$ (with $\delta = s/2$). However, the different scales can be combined in a multi-scale representation to further improve the matching performance. We observed that both feature level fusion[14] and score level fusion between the two scales offered similar improvements. Using score level fusion, the multi-scale match scores for matcher j are: $D^j = D^{j_{16}} + D^{j_{32}}$, where j_{16} and j_{32} are the match scores using $s = 16$ and $s = 32$. Incorporating the multi-scale representation with the fusion of our two matchers (Equation 5) yielded the highest sketch matching performance (Section 5).

$$D^{\text{fuse + multi-scale}} = \hat{D}^{d_{16}} + \hat{D}^{d_{32}} + \hat{D}^{c_{16}} + \hat{D}^{c_{32}} \tag{5}$$

It is important to clarify that while the entire SIFT framework is scale invariant, the SIFT feature descriptor itself is not. When used in conjunction with entire SIFT framework, the scale of the SIFT feature descriptor is determined from the feature point detection stage. This scale is then used to set the window size s. In our case, we are not detecting feature points, but instead uniformly sampling the descriptors. Therefor, the scale of the feature descriptor must be set manually, which we do in a multi-scale fashion with window sizes $s = 16$ and $s = 32$.

5. EXPERIMENTAL RESULTS

We tested the performance of the proposed matching methods using the same collection of sketch/photo pairs used by Wang and Tang.[8] The photos in this dataset are 123 images from the AR database,[15] 295 images from the XM2VTS database,[16] and 188 images from the CUHK database, yielding a total of 606 sketch/photo pairs. Two examples from each of these datasets can be seen in Figure 2. Using the eye location of each subject, we rotated the face images so that the angle between the two eyes was 0°, and scaled the images to a 75 interocular pixel distance (IPD). Each face image was then cropped to a size of 196 by 258 pixels. This image preprocessing is important to align all of the face images together; otherwise the computed differences between two images could be due to misalignment and not due to the fact that images appear different.

We conducted two separate experiments to measure the performance of our feature-based sketch matching algorithm. The first experiment was designed to compare our algorithm to other published sketch matching algorithms. In this first experiment our dictionary was populated by randomly selecting 100 sketch/photo pairs from the available 606 sketch/photo pairs for training. We then randomly selected another 300 sketch/photo pairs to test our algorithm, which is the same number of sketch/photo pairs used from this set by Wang and Tang[8] and Tang and Wang.[5] Though we did not know the exact membership of the training and test sets used by Wang and Tang[8] and Tang and Wang,[5] this experiment allowed a rough comparison of the feature-based matching performance against the synthesis approach and the linear transformation method. We report the average match scores and the standard deviation across 5 random splits.

In the second experiment we performed five-fold cross validation on the entire dataset by using 121 images for training, and 485 images for testing. This experiment is intended to offer a more thorough analysis of the performance of our method.

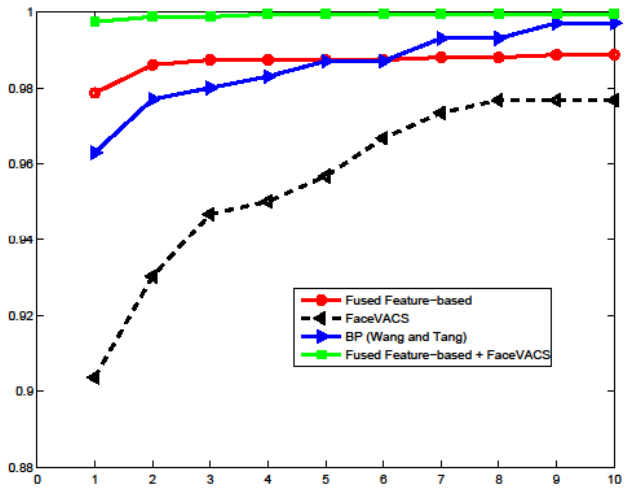

Figure 5. CMC plot (rank versus accuracy) of the average recognition performance of sketch/photo matching algorithms in experiment 1

In addition to comparing the proposed method against published sketch/photo algorithms, we also performed sketch/photo matching using Cognitec's FaceVACS, one of the best performing commercial face recognition engines.[17] The performance of matching sketches and photos using a commercial face matcher more accurately defines the baseline performance. Note that the baseline performance in previous studies[5,6,8] was based on a PCA matcher.[1]

Table 1 lists the Rank-1 performance of sketch/photo matching algorithms from the first experiment, and Figure 5 shows the CMC plots for select matching algorithms. Both of our sketch/photo matching algorithms perform slightly better than the belief propagation synthesis matching framework presented by Wang and Tang.[8] Fusing the two algorithms boosts the Rank-1 performance about a percentage point higher.

The results from the second experiment (cross validation) can be seen in Table 2, where (despite the test population increasing from 300 to 485) the average accuracy remained very similar to the first experiment with a very small standard deviation. In both experiments the matching scores between the direct feature matching and the common representation are quite similar, despite the fact that they are rather different in implementation.

It should be noted that we consider our method for matching sketches and photos to be far simpler to implement than the photo synthesis method;[8] the Markov random field is highly sensitive to parameter selection and coding a belief propagation engine takes much care. Additionally, the total time to match two images takes less than one second using our method, while a successful convergence of belief propagation can take several minutes per image using the photo synthesis method. The advantage of the photo synthesis method is that it generates a synthetic photograph of a sketch, which can be used for improved human recognition of the sketches.

Another important distinction between our approach and earlier approaches is the choice of baseline. In previous studies the baseline performance of matching sketch and photos was pessimistically biased since it was based on PCA matcher. In our first experiment the performance of a state of the art matcher, FaceVACS, on matching sketches and photos has a rank-one accuracy of 90.37%, which is orders of magnitude higher than the baseline of 6.3% previously reported.[8] This fact means that while the method presented in this paper and by Tang and Wang offer improvements in sketch/photo matching over a state of the art face matcher, the performance gain over a state of the art matcher is not as substantial as originally implied. A score level fusion of FaceVACS and the proposed feature-based method achieved an average Rank-1 accuracy of 99.73%.

6. CONCLUSION

We have proposed an effective method for matching facial sketch images to face photographs. Our method uses local image features to describe both sketch and photo images in a common representation framework. Many opportunities for future research stem from the results shown in this work. Of immediate interest is to conduct

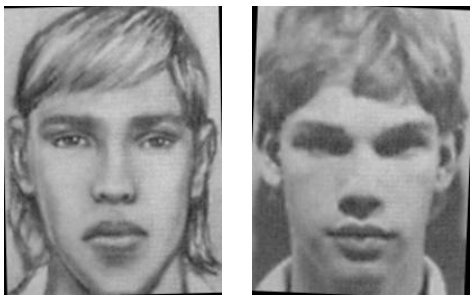

Figure 6. Example of a description-based sketch[18] (left). The true identity depicted in the photograph (right) was unknown at the time the sketch was drawn. Future research will focusing on matching these more difficult sketches that are drawn by an artist that can only use descriptions from an eye witness.

fusion between our local feature-based method and Wang and Tang's global matching framework.[8] We believe that such a method of hybrid sketch/photo matching should improve recognition accuracy even further due to the complementary nature of the two approaches (i.e. one method harnesses local differences between two faces, while the other considers global differences). The use of alternate image features is also a fruitful direction of research. In our experimentation we have observed that simple image features such as image intensity, Haar features, and Gabor images do not yield successful matching results, though there likely exists other descriptors with the same (or better) discriminative capabilities as the SIFT descriptor. Finally, the effectiveness of our matching algorithm across other image domains (e.g. NIR and visible light images) should be investigated.

The next phase in sketch to photo matching is to begin using description-based sketches. While sketch matching has already been shown to be a difficult problem, the method presented in this paper and in prior publications[5–8,11] have only dealt with sketches that were drawn by an artist who viewed each person's photograph. However, real world uses of sketch matching are with forensic sketches that are only drawn from eye witness description. Figure 6 shows a description-based sketch drawn by famed forensic sketch artist Lois Gibson.[18] Larger discrepancies between the sketch and photo are observed in the description based sketch than sketches shown in Figure 2. Future sketch matching will need to account for not only the difference between sketches and photos, but also the further appearance changes introduced by only a description being used to draw the sketch.

REFERENCES

[1] Turk, M. and Pentland, A., "Eigenfaces for recognition," *Journal of Cognitive Neuroscience* **3**(1), 71–86 (1991).

[2] Phillips, P., Scruggs, W., OToole, A., Flynn, P., Bowyer, K., Schott, C., and Sharpe, M., "Frvt 2006 and ice 2006 large-scale results," in [*NISTIR 7408*], (2007).

[3] Gross, R., Baker, S., Matthews, I., and Kanade, T., "Face recognition across pose and illumination," in [*Handbook of Face Recognition*], Li, S. Z. and Jain, A. K., eds., Springer-Verlag (2004).

[4] Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., and Ma, Y., "Robust face recognition via sparse representation," *IEEE Trans. Pattern Analysis & Machine Intelligence* **31**(2), 210–227 (2009).

[5] Tang, X. and Wang, X., "Face sketch synthesis and recognition," in [*Proc. of IEEE International Conference on Computer Vision*], 687–694 (2003).

[6] Tang, X. and Wang, X., "Face sketch recognition," *IEEE Trans. Circuits and Systems for Video Technology* **14**(1), 50–57 (2004).

Table 2. Sketch matching accuracy using 5-fold cross validation (121 training images and 485 test images).

Matching Algorithm	Mean Accuracy (%)	Std Dev (%)
FaceVACS	87.55	0.70
Feature-based, Direct	96.25	0.40
Feature-based, Common	95.59	0.96
Feature-based, Fusion	**97.57**	0.36

[7] Liu, Q., Tang, X., Jin, H., Lu, H., and Ma, S., "A nonlinear approach for face sketch synthesis and recognition," in [*Proc. of IEEE Conference on Computer Vision & Pattern Recognition*], 1005–1010 (2005).

[8] Wang, X. and Tang, X., "Face photo-sketch synthesis and recognition," *IEEE Trans. Pattern Analysis & Machine Intelligence* **31**, 1955–1967 (Nov. 2009).

[9] Zhao, W., Chellappa, R., Phillips, P. J., and Rosenfeld, A., "Face recognition: A literature survey," *ACM Comput. Surv.* **35**(4), 399–458 (2003).

[10] Freeman, W. T., Pasztor, E. C., and Carmichael, O. T., "Learning low-level vision," *International Journal of Computer Vision* **40**, 25–47 (October 2000).

[11] Li, Y., Savvides, M., and Bhagavatula, V., "Illumination tolerant face recognition using a novel face from sketch synthesis approach and advanced correlation filters," in [*Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*], **2**, II–II (May 2006).

[12] Zhang, Y., McCullough, C., Sullins, J., and Ross, C., "Human and computer evaluations of face sketches with implications for forensic investigations," in [*Biometrics: Theory, Applications and Systems, 2nd IEEE International Conference on*], 1–7 (29 2008-Oct. 1 2008).

[13] Lowe, D., "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision* **60**(2), 91–110 (2004).

[14] Ross, A. and Jain, A., "Information fusion in biometrics," *Pattern Recognition Letters* **24**(13), 2115–2125 (2003).

[15] Martinez, A. and Benavente, R., "The AR face database," in [*CVC Technical Report 24*], (1998).

[16] Messer, K., Matas, J., Kittler, J., and Jonsson, K., "XM2VTSDB: The extended M2VTS database," in [*Proc. of Audio and Video-based Biometric Person Authentication*], (1999).

[17] FaceVACS Software Developer Kit, Cognitec Systems GmbH, http://www.cognitec-systems.de.

[18] Gibson, L., [*Forensic Art Essentials*], Elsevier (2008).