Data Insufficiency in Sketch Versus Photo Face Recognition

Jonghyun Choi, Abhishek Sharma, David W. Jacobs and Larry S. Davis

Institute for Advanced Computer Studies

University of Maryland, College Park

{jhchoi,bhokaal,djacobs,lsd}@umiacs.umd.edu

Abstract

Computerized sketch-face recognition is a crucial element for law enforcement and has received considerable attention in the recent literature. Sketches of the suspect are hand-drawn or computer-rendered based on a verbal description of the suspect. However, the most popular and the only publicly available dataset, i.e. the CUFS face-sketch dataset, is far from realistic because the sketches are hand-drawn with the artist looking at the photographs to be matched later. After years of effort, researchers are producing nearly perfect results. However, we show that this is not because the problem is solved, but because of flaws in the dataset. In this paper, we empirically show that an off-the-shelf face recognition system for photo-sketch and sketch-photo matching with simple shape and edge features outperforms more sophisticated state-of-the-art approaches even without using training data. We additionally show that just using the hair region gives a 85.22% recognition rate. Based on the empirical evidences we argue that the current dataset available for face-sketch matching purposes is not appropriate and needs to be replaced by a more realistic one for advancement of this field.

1. Introduction

A realistic face recognition system requires robustness to variations during acquisition. Possible variations are pose, lighting condition, expression and image modality. Image modality difference is caused due to different acquisition schemes that result in different representations of face images. Examples are 3D face scan, infrared image, document photo and a sketch drawn by an artist. Each modality presents a different set of visual characteristics which makes cross-modal recognition challenging.

Photo to sketch recognition is one of the popular cross-modal situations and has a notable application in finding criminals from description-based-hand-drawn sketches,
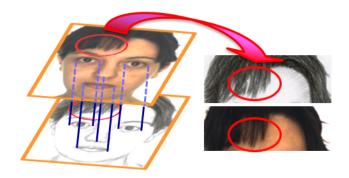

Figure 1. An example of an image pair of photo and sketch images in CUFS dataset. They are edge-wise well aligned even in the details of hair style.

called *forensic sketches*[1]. Owing to its importance, it has attracted the attention of several researchers. The problem involves finding the best match for a given sketch in a database of face photos and vice-versa. Due to representational differences of facial structure in gallery and probe images, direct matching is ineffective for recognition. There have been many approaches developed to address this problem [2, 4, 7, 9, 11–13, 18, 21, 24–26, 29–32, 32, 33].

We are motivated by the recent work by Torralba and Efros [27] that questions the validity of current object recognition datasets as motivation and benchmark for current technology. The sole purpose of our work is beautifully captured in the following excerpt from [27] :

"Datasets are an integral part of contemporary object recognition research. They have been the chief reason for the considerable progress in the field, not just as source of large amounts of training data, but also as means of measuring and comparing performance of competing algorithms. At the same time, datasets have often been blamed for narrowing the focus of object recognition research, reducing it to a single benchmark performance number. Indeed, some datasets, that started out as data capture efforts aimed at representing the visual world, have become closed

---

[1]The term is coined in [9]

worlds unto themselves (e.g. the Corel world, the Caltech-101 world, the PASCAL VOC world). With the focus on beating the latest benchmark numbers on the latest dataset, have we perhaps lost sight of the original purpose?"

Unarguably, the effectiveness of an approach in a real-world scenario is reflected in its performance on standard datasets that *capture real-world situations* [27]. The most popular and publicly available face-sketch dataset is CUFS dataset from CUHK [26], in which sketches were hand-drawn by an artist while looking at the face images to be matched later. The first version of the dataset had 188 photo-sketch pairs and was extended to 606 photo-sketch pairs owing to its popularity in the community. Recently, it has been expanded again by adding 1,194 more sketch-photo pairs from the FERET dataset (called CUFSF dataset). Sketch images in the CUFSF are also hand-drawn in the same way as the CUFS dataset so there are now 1,800 sketch-photo pairs in the combined set of CUFS and CUFSF. There are many approaches that have been evaluated on the CUFS dataset including [4, 8, 9, 13, 21, 24–26, 30–32]. Recent works have shown close to 100% recognition rate on the CUFS dataset [9, 11, 32] which implies that the problem is almost solved. We, however, argue that the dataset used for evaluation is neither challenging nor representative of real world conditions. So, the performance of various approaches is illusory and a more realistic dataset is required for a true assessment of the current approaches for sketch-image matching.

In this paper, we substantiate our claim by using an off-the-shelf face recognition system to outperform all previous state-of-the-art results. We train models using one modality data only, *i.e.* for sketch to photo matching, training requires only photo images and vice-versa, without the help of an additional training set to reduce the modality gap. Our two main contributions are 1) A systematic study of the CUFS dataset which shows good alignment makes this data *too easy* and unrealistic for evaluation of computerized photo-sketch matching algorithms. For the study, we analyzed the discrimination utility of various sub-regions of the face and found that using just the hair region can achieve very high accuracy for the entire 1,800 subjects due to its simplified data preparation scheme which involves *drawing sketches while looking at the photographs*. 2) Guidelines for future data preparation that are based on the studies of the most successful forensic sketch artist Lois Gibson [5] and requirements of a computerized photo-sketch matching system.

The paper is organized as follows; Section 2 discusses some related works, Section 3 presents our analysis on the CUFS face dataset, Section 4 shows experimental results on it, and Section 5 discussed future guidelines for face-sketch data preparation and concludes the paper in the following Section 6.

2. Related Work

Most methods developed for sketch-face matching on *viewed sketch*[2] datasets are compared on the publicly available CUFS face-sketch dataset [26]. Beginning with a simple baseline based on the Eigenface method [28], which had poor performance on this dataset [12], many approaches have been proposed and evaluated on this dataset. We categorize them into two groups; synthesis based approaches and modality gap modeling approaches.

Synthesis based approaches The underlying principle of these approaches is to bring the gallery and probe in a single modality by converting one modality to another and use a standard face recognition system to match the probe and gallery. In this way, one can take advantage of the state-of-the-art face recognition algorithm to directly solve this problem. It can be done by either synthesizing sketches from photos or vice-versa. Many works have been proposed in this category including [4, 12, 13, 24–26, 30, 31, 35]. Some of these methods report very good accuracy for sketch-photo recognition on CUFS dataset with appropriate parameters tuned.

Modality gap modeling based approaches The basis of this group of methods is to find a pair of mappings from sketch and photo modalities to a common representation and carry out matching in the common representation, which is not necessarily either of the two modalities, *e.g.* an intermediate space. Recent approaches bridge the modality gap by metric learning [32] or by maximizing covariance [21]. Approaches in this category are simple and perform well.

Efforts to tackle more realistic dataset There are more realistic face-sketch datasets other than CUFS datasets that have been constructed by having different artists to create sketches from photos, but none of them is publicly available due to privacy issues [18, 34].

Notably [9] evaluated their algorithm both on a *viewed sketch* dataset, *i.e.* CUFS dataset, as well as a *forensic sketch* dataset, which is hand-drawn from verbal descriptions. To solve the problem, they built a large set of discriminative models using LDA on two feature descriptors of SIFT and MLBP extracted from finely partitioned overlapping regions with an additional training set. Although it achieved very good recognition accuracy in *viewed sketch* dataset, it performed poorly on the *forensic sketch* dataset that is not publicly available.

---

[2]The term is coined in [9]

3. CUFS Dataset Analysis

The CUFS face-sketch dataset [30] is the most popular and the only publicly available benchmark for sketch-face recognition. It was first released with the CUHK student datasets (88 training pairs, 100 testing pairs) along with the AR sketch datasets (123 pairs) drawn from a subset of the AR dataset [14] and the XM2GTS dataset (295 pairs) drawn from a subset of the XM2VTS dataset [15]. Recently, new data from the FERET dataset [19] (1,194 pairs, named CUFSF dataset) have been added to the CUFS dataset [1]. The combined dataset of CUFS and CUFSF, which we call CUFS for simplicity, has 1,800 pairs of photos and sketches in the same frontal pose and neutral expression.

3.1. Feature transform

In [9], low level features that capture common information in photos and sketches are used to bridge the modality gap for photo-sketch matching. The common information includes face shape and edges. We use Gabor wavelet response [22] to capture the overall shape information and the CCS-POP feature [3] to capture subtle edge information.

Gabor Gabor wavelet responses are known to perform well for face recognition applications [22]. They are obtained by convolving multiple anisotropic Gaussian kernels with different scales and orientations centered at a pixel location to capture shape information including fiducial components' configuration information in expression insensitive manner. Gabor feature performs well under good alignment.

CCS-POP Circular Center Symmetric-Pairs of Pixel feature or CCS-POP was recently proposed to capture subtle edge information similar to local binary pattern (LBP) but using edge magnitude without histogramming [3]. It extracts rich information about edges and is useful for well-aligned images.

3.2. Cropping Regions

Although it is important to specify the cropping region for small gains in recognition accuracy, it is typical not to specify the cropping region. We explore the effect of various cropping protocols for sketch-face recognition using three different types of cropping 1) Tight cropping, 2) Medium cropping and 3) Loose cropping (see Fig. 2). The tightness of cropping presents a trade-offs between more information and noisy information. Vertical and horizontal strips and fiducial component-wise cropping can simulate the imperfectly aligned face-sketch datasets drawn by many artists [18, 34].

3.2.1 Tight or Loose Cropping in Face Region

The three most popular cropping regions in the face recognition literature are shown in Fig. 2. We evaluate our meth-

ods on all three different cropping regions; tight face region (tight), medium tight face region including the center of hair and chin (medium), and a loose face region including overall face outline (loose). We discuss the trade-offs of tight or loose cropping in Section 4.4.1.
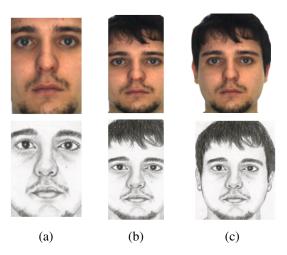

(a)   (b)   (c)

Figure 2. Examples of cropped regions of tight and loose. (a) Tight cropping. (b) Medium cropping. (c) Loose cropping

3.2.2 Horizontal/Vertical Strip

We also crop the loosely cropped face images with horizontal or vertical strips with equal width or equal height to simulate the situation that the partial region is edge-wise aligned. We also cropped the horizontal and vertical regions cumulatively to see the effect of having gradually more regions horizontally or vertically on identification rate. We denote cumulatively cropped regions by postfix `c`, *e.g.* `H2c`. Fig. 3 shows the horizontal and vertical strips of loosely cropped images. An analysis of face identification rate with each region is presented in Section 4.4.2.

3.2.3 Fiducial Components

Although the configuration of fiducial components is one of the most important cues in face recognition, each fiducial component can be used individually for face recognition [23]. To examine which component is useful in face-sketch recognition under a component-wise aligned situation, we investigate the discriminability of each fiducial component by cropping each region only as shown in Fig. 4. Since the dataset provides aligned face images in a canonical position, we can simply crop the fiducial components at the canonical location with an appropriately sized window without automatically detecting them. An analysis of the results is presented in Section 4.4.3.
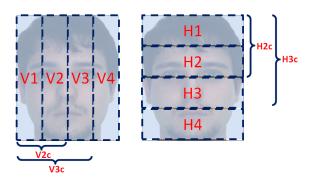
Figure 3. Examples of cropped regions horizontally and vertically. Postfix c denotes cumulative region. V4c and H4c are full face.
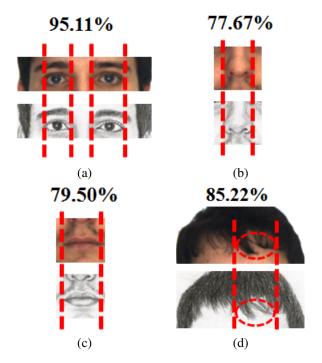

Figure 4. Examples of cropped regions for each fiducial component (including hair). (a) Ocular region. (b) Nose region. (c) Mouth region. (d) Hair region. Dotted lines indicate the images are edge-wise aligned. Note that, as the dotted circle in (d) shows, the hair region in the sketch image contains subtle shape information that is hardly found in real *forensic sketch*. The numbers above each fiducial component indicate rank-1 sketch-face identification rate with each fiducial component only. Refer to Section 4.4.3 for detailed experimental results.

4. Experimental Results

We perform comprehensive experiments with different cropping regions on the big CUFS dataset, which contains 1,800 pairs. We use a discriminant weighting scheme for uni-modal face recognition based on PLS linear regression [20] with shape and edge feature descriptors. We vi-

sualize the discriminative weights obtained by the PLS regression by a heat map of the feature weighting on shape (Gabor) and edge (CCS-POP). Since the method we use does not require an extra training set to reduce the modality gap but only uses images in the gallery set to build PLS weightings, we use the entire set of 1,800 face-sketch pairs for testing in all our experiments. As the photo partition of the CUFS dataset has negligible illumination variations, we do not apply any photometric pre-processing scheme.

Although the real world application only involves matching from sketch images to photos, we also investigate matching from photos to sketch images for possible applications such as matching a photo from surveillance video clips to several candidate montages. Since our low level analysis is focused on the common information in both modalities, it is expected to perform similarly on both matching sketch to photos and vice versa. In contrast, for synthesis-based approaches, matching a sketch to a photo is difficult since generating a photo from a sketch is nearly impossible.

4.1. Building discriminative PLS regression models

We use a recently proposed uni-modal face identification method based on PLS regression, which learns discriminative feature weighting using a one against all scheme with only the images in the gallery set [20]. For the parameters of the algorithm, we use four different numbers of PLS factors (the number of dimensions in latent space); 15,17,19 and 21. We choose the best performing one to obtain the rank-1 identification rate. Note that these numbers of latent factors are very small compared to the original feature dimensions of 144,000 obtained by Gabor and CCS-POP concatenated.

4.2. Parameters of features

For the Gabor features, scales and number of orientation are set to 5 and 8 respectively and we skip every 3 pixels horizontally and vertically to avoid redundancy. In CCS-POP, radius, angle and threshold value are set to $\{1,3\}$, $45°$ and 5 respectively and we skip every other pixels to avoid information redundancy.

4.3. Analysis of Weighted Regions

Fig. 5 and Fig. 6 show heatmaps of the feature weights obtained in the 'photo gallery' for sketch to photo matching and the feature weights obtained in the 'sketch gallery' for photo to sketch matching, respectively. In the heatmaps, the overall shape information centered at the nose tip and the edge information near the eyes are the most discriminative for this subject. We can observe edge-wise alignment in both modalities.

Our experiments indicate that most of the discriminative information is captured by Gabor features. The micro-edge information by CCS-POP marginally (0.05∼1.5%) improves the sketch(probe)-to-photo(gallery) matching accu-
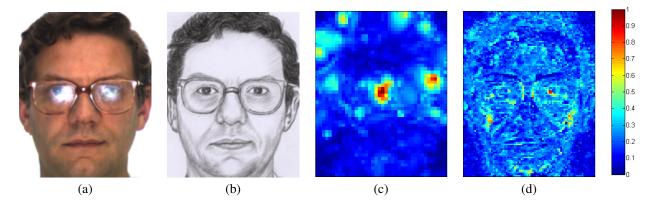
Figure 5. An example of a heatmap of PLS weighting obtained with a gallery set of photos. (a) Photo image. (b) Sketch image. (c) Weighting of Gabor feature. (d) Weighting of CCS-POP feature. Note that even with the presence of illumination highlights on the glasses, other salient regions have more weights.

racy but improves by a more significant amount (1.5~3%) in a photo-to-sketch matching with small sub regions.

Our experiments also indicate that matching sketches to photos always slightly outperforms the matching photos to sketches. There are two reasons for this. First, the PLS feature weights for photos can contain unnoticed salient features that artists can miss when they draw the sketches. Second, the PLS feature weights for sketches can give high weight to artificial lines drawn by the artist to represent shadows which appear as smooth illumination variations in the original images. For example, a small dot on the right side of the bottom of the glasses shown in Fig. 5-(a) is not reproduced by the artist in (b) while the heatmaps for the photos in (c) and (d) show high weights on it and the artificial lines for shadows in the forehead region observed in Fig. 5-(b) are given high weights as shown in the right forehead region in Fig. 6.
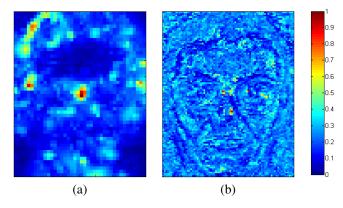

Figure 6. An example of a heatmap of PLS weighting obtained with a gallery set of sketches of the subject of Fig. 5. (a) Weighting of Gabor feature. (b) Weighting of CCS-POP feature. Note that the weights obtained by 'sketch gallery' are slightly more scattered than Fig. 5-(c) and (d).

4.4. Different Cropping Region

We show experimental results based on different cropping regions. We compare the rank-1 face identification rate of tightly, medium and loosely cropped images, horizontally and vertically cropped images, and the cropped images centered at each fiducial component.

4.4.1 Tight and Loose Crop

Table 1 shows the rank-1 identification rates of tight-, medium- and loose-cropped images. Despite the trade-offs of noisy and more information, loose cropping performs better, which is expected because the PLS discriminant weighting suppresses the noisy information and gives weights to the useful information only.

Table 1. Comparison of rank 1 face identification rate (%) according to components of a face. G+C denotes a combined feature of Gabor and CCS-POP. "S to P" denotes Sketch to Photo and "P to S" denotes Photo to Sketch.

Feature	Gabor		G+C	
Regions	P to S	S to P	P to S	S to P
Tight	95.78	96.33	97.2	97.6
Medium	96.90	97.80	98.16	98.8
Loose	98.06	99.50	99.39	**99.94**

4.4.2 Horizontal and Vertical Strip

Table 2 shows rank-1 face identification rates with different horizontal and vertical strips shown in Fig. 3. Even using only a small portion of the face image, which seems not to be discriminative to humans, *e.g.* V1 region, we obtain a quite good recognition accuracy of 93.94%.

Table 2. Comparison of rank 1 face identification rate (%) according to horizontal and vertical regions with different feature. G+C denotes a combined feature of Gabor and CCS-POP. "S to P" denotes Sketch to Photo and "P to S" denotes Photo to Sketch.

Feature	Gabor		G+C	
Region ID	P to S	S to P	P to S	S to P
H1	76.28	78.00	75.28	78.83
H2	90.94	92.56	93.00	94.50
H3	90.06	92.06	91.89	94.39
H4	86.28	89.17	86.44	90.00
V1	89.22	91.00	89.83	93.94
V2	90.61	90.83	92.50	94.22
V3	89.89	91.33	92.06	94.28
V4	87.33	89.83	87.22	90.44
H1	76.28	78.00	75.28	78.83
H2c	97.5	98.67	99.06	99.72
H3c	97.61	99.22	99.50	99.89
full(H4c)	98.06	99.50	99.39	**99.94**
V1	89.22	91.00	89.83	93.94
V2c	98.33	99.11	99.27	99.67
V3c	98.22	99.17	99.38	99.72
full(V4c)	98.06	99.50	99.39	**99.94**

4.4.3 Fiducial Component

Table 3 shows rank-1 face identification rates for each fiducial component (including hair). Notably, the hair region, which is hardly ever aligned in a real world situation, is a hand drawn snap shot of the photo and results in a good rank-1 identification accuracy of 85.22%.

Table 3. Comparison of rank 1 face identification rate (%) according to components of a face. G+C denotes a combined feature of Gabor and CCS-POP. "S to P" denotes Sketch to Photo and "P to S" denotes Photo to Sketch.

Feature	Gabor		G+C	
Regions	P to S	S to P	P to S	S to P
ocular	93.50	**93.83**	93.89	**95.11**
nose	74.11	75.00	76.06	77.67
mouth	77.11	77.39	76.94	79.50
hair	79.33	81.78	81.67	85.22

4.5. Comparison to Previous Work

We compare rank-1 photo-sketch and sketch-photo identification rate of discriminative PLS model based approach to previous approaches [4, 9, 13, 21, 24–26, 30–32, 35] in Table 4.

As the table shows, our discriminative shape and edge analysis based method gives consistently 100% with five random partitions of the original CUFS dataset of 300 face-sketch pairs out of 608 full pairs (denoted by **Ours**). Note

that most of the performance comes from Gabor features (99.80% ± 0.44). The edge feature (CCS-POP) alone also shows a descent performance of 95.53 ± 0.90. On the full sized dataset of 1,800 pairs, we obtain 99.94% rank-1 accuracy, which is even better than the state-of-the-art performance of 99.87% (on 300 sized subset of 1,800 sized set) that is obtained by learning modality gap with photometrically pre-processed images [32].

Table 4. Comparison of rank 1 face identification rate (%) to the previous approaches. #Train and #Test denote number of training and testing samples respectively. MS MRF refers to Multi-scale MRF model and W.PCA refers to a whitened PCA. Ours, referring to the discriminative shape and edge analysis method with PLS regression, belongs to neither of two categories: synthesis based methods and modality gap modeling based methods. 'Ours (Gabor only)' refers to our method without the edge feature (CCS-POP), 'Ours (CCS-POP only)' refers to our method only with the edge feature.

Approach	#Train	#Test	Rate (%)
Sketch Synthesis			
Tang and Wang [24, 26]	88	100	71
Tang and Wang [25]	306	300	81.3
Nonlinear [13]	306	300	87.67
E-HMM [4, 35]	306	300	95.24
MS MRF+LDA [30]	306	300	96.3
MS MRF+LDA (from [31])	88	100	96
MS MRF+W.PCA [31]	88	100	99
Modelling Modality Gap			
PLS-subspace [21]	88	100	93.6
Klare *et al.* [9]	306	300	99.47
CITP [32]	306	300	99.87
Ours (Gabor only)	0	300	99.80 ± 0.44
Ours (CCS-POP only)	0	300	95.53 ± 0.90
Ours	0	300	**100**
Ours (Gabor only)	0	1,800	99.50
Ours (CCS-POP only)	0	1,800	96.28
Ours	0	1,800	**99.94**

Note that Klare *et al.* evaluated their method with a realistic dataset (called Forensic Sketch dataset) in [9]. They report a significant performance drop in the realistic dataset (the rank-1 identification rate: 99.47% to 17% in good sketch, 0% in poor sketch). Since the dataset is not publicly available due to privacy issue, we could not evaluate our method on it. We also expect a serious performance drop in the realistic dataset.

5. Guideline for New Datasets

Although the CUFS dataset has encouraged active research and competition for the face-sketch recognition problem, the dataset has a back door to a near-perfect solution with a proper shape and edge analysis armed with a fea-

ture weighting scheme such as partial least squares. For the sketch-face recognition research to be more practical, the sketches should be generated by verbal descriptions and released publicly. The current protocol for generating sketch images by allowing artists to draw while looking at the real photos, even with number of artists involved, does not reflect the real world situation for matching forensic sketches to images in a public watch-list.

Based on the guidelines put forth by Gibson in her book [5] and the fact that generated sketches are to be matched automatically to photos of suspects in the database, we propose some guidelines for a realistic sketch-photo dataset and contrast it with the present CUFS dataset.

1. **Sketches drawn by verbal description** - The sketches should be drawn using verbal descriptions only to reflect real-life scenarios. The CUFS database is not drawn accordingly as shown in Fig. 8.

2. **"Filled in" sketches** - The sketches should be filled in to give an impression of 3D structure of the face rather than a stroke based sketch as shown in the Fig. 7. Such rendering makes the sketches more akin to a photo and also they are found to be much better for recognition than edge based sketches [5].

   (a) **Matching illumination conditions** - Given the suspect database of photos taken under a standard frontal lighting condition such as mug shots, sketches should be drawn to match the same frontal illumination condition to facilitate better matching.

   (b) **Skin color** - Skin color is a major factor which can be helpful in automatic matching of the images and also human based verification once the suspect is caught. However, sketches in CUFS do not capture this information.

3. **Side information** - A realistic and easily recognizable forensic sketch *must* use side information for rendering [5]. For example, aged men have a broad chin while young men typically have a narrow chin [5], it was shown in [5] that adjusting the width of the chin made a good amount of visual difference in the appearance of the person in the sketch. The side information can also be used to narrow down the search scope resulting in faster and more accurate matching. The CUFS dataset does not include side information.

We believe that a dataset based on these guidelines would reflect real-life constraints and at the same time contain enough information for automatic sketch-photo matching.
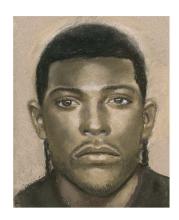

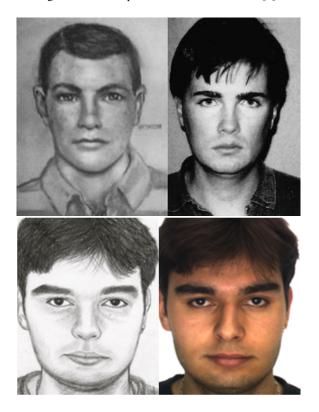Figure 7. An example of a "filled in" sketch in [5].

Figure 8. Comparison between the forensic sketch (hand drawn by verbal descriptions) and photo pair from [5, 9] (top) with the viewed sketch and photo pair from CUFS dataset (bottom). Note the difference between the hair style in forensic photo-sketch pair which is absent in the CUFS pair and the difference in the eyebrow shape which is present in the forensic photo-sketch pair.

6. Conclusion and Future Work

We empirically showed that the most popular publicly available photo-sketch CUFS dataset is far from the realistic conditions faced when matching sketches to photos.

The dataset has near perfect pixel-wise alignment between the sketch and corresponding photos, which makes it possible to use a pixel-centered shape feature and a simple edge-based feature with an appropriate classifier to achieve nearly 100% accuracy. To prove this empirically, we have used an off-the-shelf face recognition system with a Gabor filter and an edge feature and outperformed all previous approaches evaluated on CUFS dataset even without using any training images. Finally, we have suggested some guidelines to prepare a future dataset that would be realistic and useful for assessment of current progress in photo-sketch matching.

As implied in the recent works by Ng and Hinton [6, 10, 16, 17] on deep learning that is inspired by the human visual pathway, they could reconstruct images of faces and objects from only edge patches. It implies that the multiple layers of neural network is the reason that human can easily match sketch to photo and vice versa. In contrast, many vision algorithms use a single layer classifier. For a promising future work, we could leverage a deep belief network (DBN) to match the sketch to photo face images with diverse sources of variations such as illumination, expression and pose.

References

[1] http://mmlab.ie.cuhk.edu.hk/cufsf/. 3

[2] H. S. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa. On matching sketches with digital face images. In *BTAS*, pages 1–7, 2010. 1

[3] J. Choi, W. R. Schwartz, H. Guo, and L. S. Davis. A Complementary Local Feature Descriptor for Face Identification. In *WACV*, 2012. 3

[4] X. Gao, J. Zhong, J. Li, and C. Tian. Face Sketch Synthesis Algorithm Based on E-HMM and Selective Ensemble. *Circuits and Systems for Video Technology, IEEE Transactions on*, 18(4):487–496, 2008. 1, 2, 6

[5] L. Gibson. *Forensic Art Essentials: A Manual for Law Enforcement Artists*. 1 edition, 2007. 2, 7

[6] G. E. Hinton, S. Osindero, and Y. Teh. A fast learning algorithm for deep belief nets. *Neural Computation*, 18:1527–1554, 2006. 8

[7] A. K. Jain and B. Klare. Matching Forensic Sketches and Mug Shots to Apprehend Criminals. *Computer*, 44(5):94–96, 2011. 1

[8] B. Klare and A. Jain. Sketch to Photo Matching: A Feature-Based Approach. In *Proc. SPIE Conf. Biometric Technology for Human Identification*, 2010. 2

[9] B. Klare, Z. Li, and A. K. Jain. Matching Forensic Sketches to Mug Shot Photos. *IEEE T. PAMI*, 33(3):639–646, 2011. 1, 2, 3, 6, 7

[10] Q. V. Le, J. Ngiam, A. Coates, A. Lahiri, B. Prochnow, and A. Y. Ng. On optimization methods for deep learning. In *ICML*, 2011. 8

[11] Z. Lei and S. Z. Li. Coupled Spectral Regression for matching heterogeneous faces. In *CVPR*, 2009. 1, 2

[12] D. Lin and X. Tang. Inter-modality Face Recognition. In *ECCV (4)*, pages 13–26, 2006. 1, 2

[13] Q. Liu, X. Tang, H. Jin, H. Lu, and S. Ma. A nonlinear approach for face sketch synthesis and recognition. In *CVPR*, volume 1, 2005. 1, 2, 6

[14] A. M. Martinez and R. Benavente. The AR Face Database. Technical report, 1998. 3

[15] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre. XM2VTSDB: the Extended of M2VTS Database. In *International Conference on Audio- and Video-Based Person Authentication*, pages 72–77, 1999. 3

[16] J. Ngiam, Z. Chen, P. Koh, and A. Y. Ng. Learning deep energy models. In *ICML*, 2011. 8

[17] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng. Multimodal deep learning. In *ICML*, 2011. 8

[18] H. Nizami, J. P. Adkins-Hill, Y. Zhang, J. R. Sullins, C. McCullough, S. Canavan, and L. Yin. A biometric database with rotating head videos and hand-drawn face sketches. In *BTAS*, pages 1–6, 2009. 1, 2, 3

[19] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The FERET Evaluation Methodology for Face-Recognition Algorithms. *IEEE T. PAMI*, 22:1090–1104, 2000. 3

[20] W. R. Schwartz, H. Guo, and L. S. Davis. A Robust and Scalable Approach to Face Identification. In *ECCV*, 2010. 4

[21] A. Sharma and D. Jacobs. Bypassing Synthesis: PLS for Face Recognition with Pose, Low-Resolution and Sketch. In *CVPR*, pages 593–600, 2011. 1, 2, 6

[22] L. Shen and L. Bai. A review on Gabor wavelets for face recognition. *Pattern Anal. Appl.*, 9:273–292, 2006. 3

[23] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell. Face Recognition by Humans: Nineteen Results All Computer Vision Researchers Should Know About. *Proceedings of the IEEE*, 94(11):1948–1962, 2006. 3

[24] X. Tang and X. Wang. Face photo recognition using sketch. In *ICIP*, volume 1, 2002. 1, 2, 6

[25] X. Tang and X. Wang. Face sketch synthesis and recognition. In *ICCV*, 2003. 1, 2, 6

[26] X. Tang and X. Wang. Face sketch recognition. *Circuits and Systems for Video Technology, IEEE Transactions on*, 14(1):50–57, 2004. 1, 2, 6

[27] A. Torralba and A. A. Efros. Unbiased Look at Dataset Bias. In *CVPR*, 2011. 1, 2

[28] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *CVPR*, pages 586–591, 1991. 2

[29] Uhl and N. da Vitoria Lobo. A framework for recognizing a facial image from a police sketch. In *CVPR*, pages 586–593, 1996. 1

[30] X. Wang and X. Tang. Face Photo-Sketch Synthesis and Recognition. *IEEE T. PAMI*, 31(11):1955–1967, 2009. 1, 2, 3, 6

[31] W. Zhang, X. Wang, and X. Tang. Lighting and Pose Robust Face Sketch Synthesis. 2010. 1, 2, 6

[32] W. Zhang, X. Wang, and X. Tang. Coupled Information-Theoretic Encoding for Face Photo-Sketch Recognition. In *CVPR*, 2011. 1, 2, 6

[33] 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 *FG*, pages 707–712, 2011. 1

[34] Y. Zhang, C. McCullough, J. R. Sullins, and C. R. Ross. Hand-Drawn Face Sketch Recognition by Humans and a PCA-Based Algorithm for Forensic Applications. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 40(3):475–485, 2010. 2, 3

[35] J. Zhong, X. Gao, and C. Tian. Face Sketch Synthesis using E-HMM and Selective Ensemble. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, volume 1, 2007. 2, 6