Recognizing Composite Sketches with Digital Face Images via SSD Dictionary

Paritosh Mittal*, Aishwarya Jain*, Gaurav Goswami, Richa Singh, and Mayank Vatsa
IIIT-Delhi, India
{paritosh10059,aishwarya10007,gauravgs,rsingh,mayank}@iiitd.ac.in

Abstract

Sketch recognition has important law enforcement applications in detecting and apprehending suspects. Compared to hand drawn sketches, software generated composite sketches are faster to create and require lesser skill sets as well as bring consistency in sketch generation. While sketch generation is one side of the problem, recognizing composite sketches with digital images is another side. This paper presents an algorithm to address the second problem, i.e. matching composite sketches with digital images. The proposed algorithm utilizes a SSD based dictionary generated via 50,000 images from the CMU Multi-PIE database. The gallery-probe feature vectors created using SSD dictionary are matched using GentleBoostKO classifier. The results on extended PRIP composite sketch database show the effectiveness of the proposed algorithm.

1. Introduction

Sketches are an integral part of the process followed by law enforcement agencies for apprehending suspects. Traditionally, sketches were prepared by artists, however, with advancements in technology, law enforcement agencies are switching to the use of softwares for drawing face composites according to the description of eyewitnesses. These sketches or composites are matched with existing records and/or to create public awareness in case a match is not available in the database. Composite sketches have the advantage of being relatively independent of the artist and consistent across different agencies and states. Figure 1 shows a digital image with the corresponding hand-drawn (viewed) and composite sketch. With increasing size of mugshot databases that law enforcement agencies have, it is imperative to have an automatic sketch to digital face matching algorithm. It can enable the law enforcement agencies to undertake swift action by reducing the matching time and minimizing human efforts.

Figure 1. Hand-drawn sketch and composite sketch with corresponding digital image.

As it is evident from the composite digital image pair, a composite is not always very close to the digital image. Therefore, there are several challenges in creating an algorithm to match the two representations. Despite their many advantages, composite sketches also suffer from the problems prevalent in hand drawn sketches. They are bereft of texture information and minute facial features as compared to a digital image of the subject. Since the witness observes the suspect's facial characteristics very briefly, the description provided by him/her is not accurate and fuzzy at best Therefore, composite sketches generally capture the basic shape of the face and some uniquely identifiable features such as scars, moles, and prominent marks. Due to this, traditional feature extraction and matching algorithms may not perform well in the context of composite sketches. Moreover, the accuracy of existing state-of-the-art algorithms is not very high and there is large scope for improvement [4, 9, 10, 13, 18]. In this research, an efficient composite sketch matching algorithm is proposed. The key contributions of this research are as follows:

- A dictionary based algorithm to match composite sketches with digital images is proposed. Bag-of-words features are extracted at three Gaussian smoothing levels using Self Similarity Descriptor (SSD). Feature histograms are computed and then compared using χ^2 distance. Finally, GentleboostKO classifier provides the matched ranked list for identification.
- A new set of composite sketches is created by an Indian user in accordance with the PRIP composite sketch database. The extended dataset is termed as the e-PRIP dataset.

*Equal contributions by the authors.

2. Literature Review

Bhatt *et al.* [10] classified existing sketch recognition algorithms into generative and discriminative approaches. Among discriminative approaches, Uhl and Lobo [16] proposed the use of Eigen analysis for matching sketches with digital images while Zhang *et al.* [23] suggested Principal Component Analysis for matching. Klare and Jain [4] proposed Local Feature Discriminative Analysis. Bhatt *et al.* [5, 10] proposed Multiscale Circular Weber Local Descriptor (MCWLD) in conjunction with an evolutionary algorithm for matching hand-drawn sketches with digital images.

Among existing generative approaches, Tang *et al.* [14] proposed an algorithm which separates texture and shape information from images in combination with a Bayesian classifier for matching. They also proposed Eigen-analysis to first convert digital images to sketches [21]. Gao *et al.* [20] devised an approach that involves embedded Hidden Markov Models for synthesizing pseudo-sketches. Markov Random Fields can also be utilized to generate sketches from digital images as proposed by Zhang *et al.* [22].

In composite sketch recognition literature, Han *et al.*[9] proposed a component based framework for matching. The face is divided into parts based on facial keypoints. Each part is encoded using local binary patterns. The key feature of the algorithm is that the approach does not involve any training. Recently, Mittal *et al.* [13] demonstrated that learning based approach can achieve higher accuracy. They proposed feature extraction using DAISY descriptor from circular patches of both digital image and composite sketch. The extracted features are combined using Boosting [15]. Klum *et al.* [11] in their analysis, observed that composite sketches were more effective compared to hand-drawn forensic sketches. Chugh *et al.* [18] proposed an algorithm for matching composite sketches with digital images across large age variations.

3. Proposed Algorithm

As discussed before, composite sketches lack detailed texture information and only contain the basic shape of face and prominent features. Based on these observations, the proposed algorithm utilizes Gaussian smoothing in order to focus on global features. Recent research in face recognition has demonstrated that matching images based on a dictionary generally yields better results than direct comparison [19]. Therefore, this research proposes a dictionary based approach to perform composite sketch to digital image matching. Figure 2 shows the steps involved in the proposed algorithm.

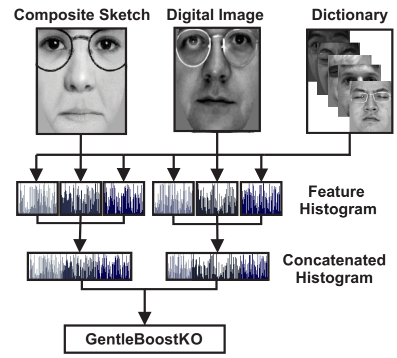


Figure 2. An overview of the proposed algorithm.

Preprocessing: Digital and composite sketch images differ in both intrinsic and extrinsic properties. Composite sketches can have software logos embedded onto the sketch while digital images can have some background information. Since such elements are non-informative for recognition, it is essential to filter them and retain only the most discriminative regions of the image. Therefore, CSU face toolbox [7] is utilized to detect the facial region and affine registration is performed on the basis of eye coordinates. After preprocessing, all the images are converted to grayscale and resized to 130×150.

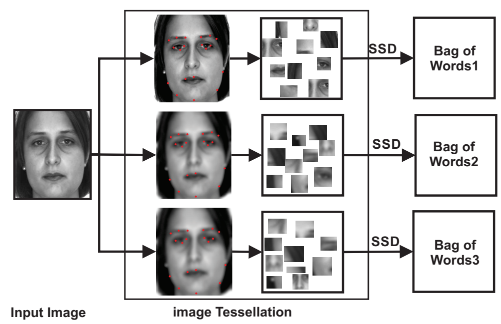


Figure 3. Feature extraction steps involved in the proposed algorithm.

Creating a dictionary: A dictionary of 50,000 digital faces is created from the Multi-PIE dataset [17] (Frontal pose images only). After preprocessing, features are extracted from each face image in the dictionary. The first step in feature extraction is to identify feature-rich keypoints. Mean face I_{avg} of the dictionary images are obtained and key points are identified from I_{avg} by finding the corner points in the image using the Harris corner detector [6]. Since the number of keypoints is large, only top K points are selected for feature extraction and matching. The optimal value of K is chosen by performing a grid search with identification accuracy as the optimization criteria. For every keypoint patches of size $p \times p$ are extracted. Since the patches have to be of size $p \times p$, corner points that are less than $\frac{p}{2}$ pixels away from any border of the image are discarded.

As discussed earlier, a digital image and sketch contain different kinds of information, therefore, it is our assertion that extracting features at multiple granularities can provide better matching. In this research, three Gaussian smoothing levels have been utilized: the first level being the original image, the second level with variance = 15, and the third level with variance = 21. Let I be the face image and G_i be the Gaussian filter at i^{th} smoothing level. A patch is extracted around each keypoint from each F_i according to Equations (1) and (2):

$$F_i = I * G_i \tag{1}$$

$$P_{ij} = F_i(x \in [x_{k_j} \pm p],\, y \in [y_{k_j} \pm p]),\, j \in [1, K] \tag{2}$$

Here, P_{ij} denotes the patch extracted from the i^{th} Gaussian smoothing level and j^{th} keypoint. (x_{k_j}, y_{k_j}) denotes the coordinates of the j^{th} keypoint.

Finally, features are extracted from the patches corresponding to the three Gaussian smoothing levels. Self Similarity Descriptor (SSD) [8], S_{ij}, is calculated for each patch P_{ij}. SSD extracts local feature information from the patch and encodes the internal geometric layouts, colors, edges, and patterns. A collection of bag-of-words is created for the i^{th} Gaussian smoothing level by encoding the K patches extracted from it. Each SSD feature of size 80×40 is treated as one bag-of-word where each word is a histogram of size 40. These features are extracted and stored for each dictionary face image. Figure 3 summarizes the steps involved in dictionary generation.

Matching composite sketches with images (Training): The training data consists of a set of sketches and the corresponding ground truth (digital face images), which are labeled such that a unique number is associated with the sketches and digital images belonging to the same subject in the database. First, the sketches and digital faces are preprocessed and features are extracted from all the training images as described previously during creation of the dictionary. The SSD features of an image S_{ij} are matched with the SSD features of the n^{th} dictionary image, denoted by S^n_{ij}, according to Equation (3):

$$\vec{d_{in}} = \sqcup_{j=1}^{K} \chi^2(S_{ij},\, S^n_{ij}), \tag{3}$$

Here, $\vec{d_{in}}$ denotes the match score histogram for the n^{th} dictionary image and the i^{th} Gaussian smoothing level. χ^2 represents the χ^2 distance metric commonly used to match two histograms and $\sqcup$ denotes the histogram operator. Since the size of the SSD feature is 80×40, matching two SSD features provides 80 distance scores. Therefore, by matching K SSD features, $K \times 80$ distances are obtained. After obtaining $K \times 80$ scores via χ^2 distance metric, a histogram of these scores is computed. These scores are quantized to 100 distinct levels and each real value is mapped uniquely to one of these 100 levels. $\sqcup$ operator therefore generates a histogram of size 100. This process is repeated for all the Gaussian smoothing levels and individual match score histograms obtained are concatenated together according to Equation (4):

$$\vec{d_n} = \oplus_{i=1}^{3} \vec{d_{in}} \tag{4}$$

Here, $\oplus$ denotes the horizontal concatenation operator and $\vec{d_n}$ denotes the concatenated match score histogram for the n^{th} dictionary image. As discussed previously, the image is matched to each dictionary image and concatenated histograms are obtained for each using Equations (3) and (4). These concatenated histograms are combined over the entire dictionary as per Equation (5):

$$\vec{d} = \sum_{i=1}^{N_d} \vec{d_i} \tag{5}$$

The final match score histogram $\vec{d}$ is therefore the bin-wise summation of all the individual concatenated match score histograms. Once the match score histograms are computed for both sketch and all the digital faces from the gallery, a difference histogram is computed according to Equation (6):

$$\vec{d_h} = \frac{\left(\vec{d}_{sketch} - \vec{d}_{digital}\right)^2}{\vec{d}_{sketch} + \vec{d}_{digital}} \tag{6}$$

Here, $\vec{d_h}$ denotes the difference histogram between the sketch and digital image. $\vec{d}_{sketch}$ and $\vec{d}_{digital}$ denote the concatenated match score histogram of the sketch and the digital image respectively. Since the training data is scarce, GentleBoostKO is trained to perform *identification-in-verification* mode. For each training sketch image, we can obtain one genuine digital pair from the training gallery and the remaining pairs act as impostor. The difference histograms of these pairs are input to GentleBoostKO and it is trained for two-class classification.

Identification: To identify a given probe sketch, SSD features are extracted from the given probe sketch and matched with dictionary faces according to Equation (3). The difference histogram between the probe and all the gallery images are obtained using Equation (6). These difference histograms are provided to trained GentleBoostKO classifier which provides match scores for the probe sketch with each subject in the gallery in a pair-wise manner. The subjects that are classified as genuine are then sorted in descending order of match score in order to obtain the ranked list of matches for the probe sketch.

Since the patches extracted using Equation (2) may contain some redundant information across the three smoothing levels, identifying and removing the unnecessary features based on data characteristics can help improve the performance. By using the Knock Out feature, GentleBoostKO [12] classifier reduces this redundancy and helps in efficient classification. Moreover, the weak classifiers obtained by the individual difference histograms are combined to perform accurate classification by utilizing a boosting approach.

4. Experimental Results

The performance of the proposed algorithm is evaluated on the PRIP database. The PRIP database contains composite sketches of 123 subjects (70 male, 63 female) from the AR database [3]. The dataset is created by users from two different ethnicities, Caucasian and Asian, using two software tools, FACES [1] and IdentiKit [2]. A set of composite sketches is generated by both users using FACES toolkit. In addition, the Asian user has also created another dataset using Identikit. Thus, total of three sets of composite sketches are present in the PRIP database. Since the database contains composites drawn from two different ethnicities, another set is created by a user belonging to Indian ethnicity using the software FACES. This set along with the existing PRIP database is termed as the Extended PRIP Database (e-PRIP). Figure 4 shows sample images from all three users. For performance evaluation, the e-PRIP dataset is divided into 40% training (48 subjects) and 60% (75 subjects) testing partitions. Five-times random cross validation is performed and average accuracies along with the Cumulative Match Characteristic (CMC) curves are reported. Two different sets of experiments are performed to evaluate the proposed algorithm:

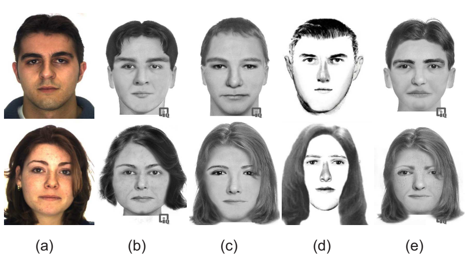


Figure 4. Sample images from the e-PRIP database [9]. (a) digital images, (b) Caucasian user (FACES), (c) Asian user (FACES), (d) Asian user (Identi-Kit), and (e) Indian user (FACES).

- Experiment 1: In this experiment, the gallery size is equal to the probe size of 75 subjects. The experiments are performed to compare the performance of the proposed algorithm with existing algorithms as well as evaluate the impact of different components of the proposed algorithm on the matching performance.
- Experiment 2: The gallery size is extended to 1500 subjects (images obtained from law enforcement agencies) while the probe size remains 75 subjects. This experiment is performed to evaluate the robustness of the algorithm with larger gallery size, as is the case in law enforcement applications.

4.1. Experiment 1: Baseline Experiment

Baseline experiments are performed on the e-PRIP database and the performance of the proposed algorithm is compared with two existing sketch recognition algorithms, MCWLD [10] and Mittal *et al.* [13], along with a commercial-off-the-shelf (COTS) face recognition algorithm. The results are shown in Figure 5 and summarized in Table 1. Some of the key observations are discussed below:

- At rank 10, the proposed algorithm outperforms existing algorithms with a significant performance margin. For sketches created by Caucasian user, the proposed algorithm achieves a rank 10 identification accuracy of 51.9 % and outperforms the next best performing algorithm (Mittal *et al.* [13]) by 19.5%. Similar results are observed on the other sketch sets as well.
- COTS yields poor results for composite sketch recognition, achieving the best performance of only 11.3% rank 10 accuracy. Similarly, lower performance is observed for MCWLD [10] algorithm. This is primarily because COTS and MCWLD [10] are not optimized for matching composite sketches with digital images.
- The performance of individual bag-of-words at different Gaussian smoothing levels is presented in Figure 5. It is observed that the combination of three bags improves the results and performs better than any other combination of a different number of bags. This demonstrates that fusion of bag-of-words information is aiding recognition performance.
- As illustrated in Figure 5, GentleBoostKO outperforms all other classifiers. This can be attributed to the fact that GentleBoostKO is a Knock Out feature boosting algorithm and it efficiently handles the redundant and outlier features. Such features may be present due to overlapping patches and some extent of common features present across different gaussian smoothing levels.

4.2. Extended Gallery Experiments

The two best performing datasets (made by the Caucasian and Indian users) in baseline experiments are selected for extended gallery experiments in order to evaluate the performance of the proposed algorithm on large databases. The results are presented in Figure 6 and summarized in Table 2. At rank 50, the proposed algorithm outperforms MCWLD and COTS with a large margin of atleast 16% on the extended gallery. As shown in Figure 6, all the algorithms have low identification accuracies at rank 1. However, difference in accuracies can be observed at higher ranks and the proposed algorithm achieves the best

Table 1. Rank-10 identification accuracy (%) obtained on the e-PRIP composite sketch dataset.

Algorithm	**Faces (Ca)**	**Faces (Indian)**	**Faces (As)**	**IdentiKit (As)**
COTS	11.3 ± 2.1	9.1 ± 1.9	7.2 ± 2.2	8.1 ± 2.1
Bag1	45.4 ± 2.6	39.2 ± 2.3	34.6 ± 3.1	37.4 ± 3.5
Bag2	29.3 ± 2.9	29.5 ± 3.4	28.0 ± 3.7	21.3 ± 4.1
Bag3	32.0 ± 3.2	25.3 ± 3.1	24.0 ± 3.4	26.6 ± 3.6
Equal Weighted Sum	33.3 ± 4.3	37.2 ± 3.7	31.3 ± 2.8	33.6 ± 4.3
Weighted Sum	48.0 ± 1.2	41.3 ± 1.4	37.3 ± 0.8	40.0 ± 1.2
AdaBoost	46.6 ± 1.6	45.3 ± 0.9	35.9 ± 0.7	38.6 ± 1.1
MCWLD[10]	23.2 ± 3.2	24.0 ± 3.4	15.7 ± 3.0	15.4 ± 3.1
Mittal *et al.* [13]	32.4 ± 2.4	30.3 ± 1.7	21.3 ± 2.1	27.6 ± 1.8
Proposed	$\mathbf{51.9 \pm 1.2}$	$\mathbf{53.3 \pm 1.4}$	$\mathbf{42.6 \pm 1.2}$	$\mathbf{45.3 \pm 1.5}$

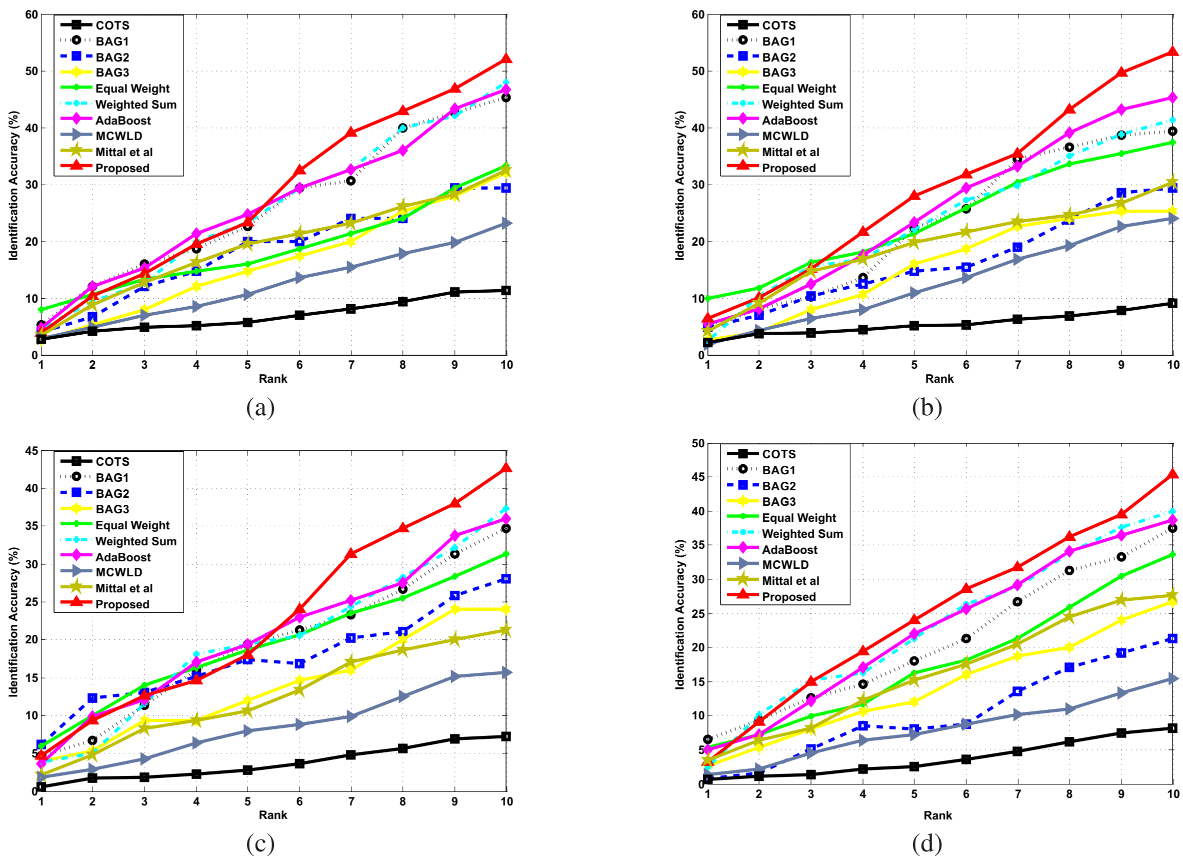


Figure 5. CMC curves to compare existing algorithms with the proposed algorithm on composite sketch database (a) Faces (caucasian user), (b) Faces (Indian user), (c) Faces (asian user), and (d) IdentiKit (asian user).

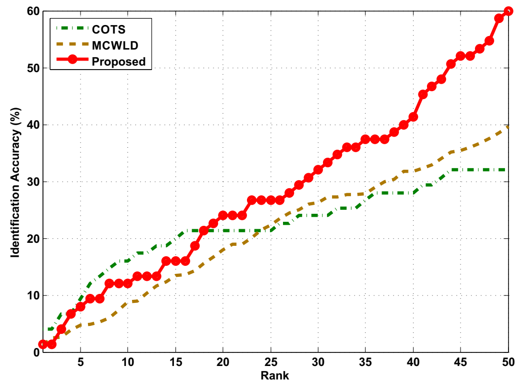
performance at rank 50. The proposed algorithm requires 220 seconds to return the rank 50 match list from the extended gallery of 1500 digital images under MATLAB programming environment.

5. Conclusion

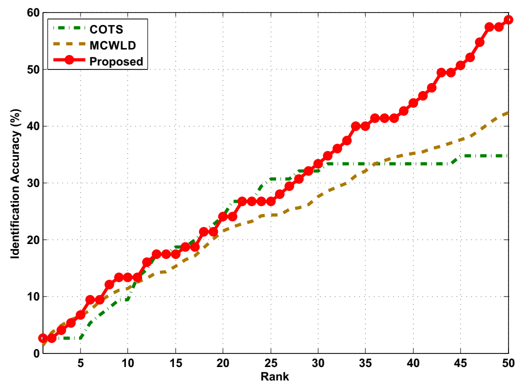
This research focuses on matching composite sketches with digital face images. The contributions of the paper are two fold: (1) developing a SSD dictionary based feature

Table 2. Rank-50 identification accuracy (%) obtained on the extended gallery on the dataset Faces(Caucasian), Faces(Indian).

Algorithm	**Faces (Ca)**	**Faces (Indian)**
COTS	32.4	34.6
MCWLD [10]	39.6	42.5
Proposed	**60.0**	**58.6**



(a)



(b)

Figure 6. CMC curves to evaluate the performance of the proposed algorithm on composite sketch database with extended gallery. The sketches used as probe are generated using the Faces software by (a) caucasian user and (b) Indian user.

extraction and GentleBoostKO classifier based composite sketch to digital face image matching algorithm and (2) extending the PRIP sketch database by including a set of composite sketches prepared by an artist of Indian ethnicity. The results on the e-PRIP database show that the proposed algorithm yields the best performance. Experiments are also performed on an extended gallery of 1500 subjects and the proposed algorithm yields rank 50 accuracy of atleast 58%.

Acknowledgements

The authors thank PRIP lab for providing access to the PRIP composite sketch database.

References

[1] Faces 4.0, iq biometrix. http://www.iqbiometrix.com, 2011.
[2] Identi-kit solutions. http://www.identikit.net/, 2011.
[3] A.R. Martinez and R. Benavente. The AR face database. Technical report, 1998. Computer Vision Center.
[4] B. F. Klare, L. Zhifeng, and A. K. Jain. Matching forensic sketches to mug shot photos. *IEEE TPAMI*, 33(3):639–646, 2011.
[5] H. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa. On matching sketches with digital face images. In *IEEE International Conference on Biometrics: Theory Applications and Systems*, pages 1–7, 2010.
[6] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey vision conference*, volume 15, page 50, 1988.
[7] D. S. Bolme, J. R. Beveridge, M. Teixeira, and A. B. Draper. The CSU face identification evaluation system: its purpose, features, and structure. In *Computer Vision Systems*, pages 304–313. 2003.
[8] E. Shechtman and M. Irani. Matching local self-similarities across images and videos. In *CVPR*, pages 1–8, 2007.
[9] H. Han, B. F. Klare, K. Bonnen, and A. K. Jain. Matching Composite Sketches to Face Photos: A Component-Based Approach. *IEEE TIFS*, 8(1), 2013.
[10] H. S. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa. Memetically optimized MCWLD for matching sketches with digital face images. *IEEE TIFS*, 7(5):1522–1535, 2012.
[11] S. Klum, H. Han, A. Jain, and B. Klare. Sketch based face recognition: Forensic vs. composite sketches. In *International Conference on Biometrics*, pages 1–8, 2013.
[12] L. Wolf and I. Martin. Robust boosting for learning from few examples. In *CVPR*, volume 1, pages 359 – 364, 2005.
[13] P. Mittal, A. Jain, R. Singh, and M. Vatsa. Boosting local descriptors for matching composite and digital face images. In *ICIP*, pages 2797–2801, 2013.
[14] Q. Liu, X. Tang, J. Hongliang, H. Lu, and S. Ma. A nonlinear approach for face sketch synthesis and recognition. In *CVPR*, volume 1, pages 1005–1010, 2005.
[15] R. E. Schapire. The boosting approach to machine learning: An overview. *Lecture Notes in Statistics*, pages 149–172, 2003.
[16] R. G. Uhl Jr. and N. da Vitoria Lobo. A framework for recognizing a facial image from a police sketch. In *CVPR*, pages 586–593, 1996.
[17] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-PIE. *Image and Vision Computing*, 28(5):807–813, 2010.
[18] T. Chugh, H. S. Bhatt, R. Singh, and M. Vatsa. Matching age separated composite sketches and digital face images. In *BTAS*, pages 1–6, 2013.
[19] V. M. Patel, W. Tao, S. Biswas, P. J. Phillips, and R. Chellappa. Dictionary-based face recognition under variable lighting and pose. *IEEE TIFS*, 7(3):954–965, 2012.
[20] X. Gao, J. Zhong, J. Li, and C. Tian. Face sketch synthesis algorithm based on E-HMM and selective ensemble. *TCSVT*, 18(4):487–496, 2008.
[21] X. Tang and X. Wang. Face photo recognition using sketch. In *ICIP*, volume 1, pages I–257–I–260, 2002.
[22] Y. Wang, L. Zhang, Z. Liu, G. Hua, Z. Wen, Z. Zhang, and D. Samaras. Face relighting from a single image under arbitrary unknown lighting conditions. *IEEE TPAMI*, 31(11):1968–1984, 2009.
[23] Y. Zhang, C. McCullough, J. R. Sullins, and C. R. Ross. Human and computer evaluations of face sketches with implications for forensic investigations. In *BTAS*, pages 1–7, 2008.