IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 20, 2019, accepted January 24, 2019, date of publication February 5, 2019, date of current version March 13, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2897599

Cascaded Static and Dynamic Local Feature Extractions for Face Sketch to Photo Matching

SAMSUL SETUMIN[1,2] AND SHAHREL AZMIN SUANDI[1], (Senior Member, IEEE)

[1]Intelligent Biometric Group, School of Electrical and Electronics Engineering, Universiti Sains Malaysia, Engineering Campus, Nibong Tebal 14300, Malaysia
[2]Faculty of Electrical Engineering, Universiti Teknologi MARA Pulau Pinang, Permatang Pauh 13500, Malaysia

Corresponding author: Shahrel Azmin Suandi (shahrel@usm.my)

This work was supported in part by the Universiti Sains Malaysia Research University Individual (RUI) Research under Grant 1001/PELECT/814208 and Grant 1001/PELECT/8014056, and in part by the Universiti Teknologi MARA.

ABSTRACT The automatic identification of a corresponding photo from a face sketch can assist in criminal investigations. The face sketch is rendered based on the descriptions elicited by the eyewitness. This may cause the face sketch to have some degrees of shape exaggeration that make some parts of the face geometrically misaligned. In this paper, we attempt to address the effect of these influences by a cascaded static and dynamic local feature extraction method so that the constructed feature vectors are built based on the correct patches. In the proposed method, the feature vectors from the local static extraction on a sketch and photo are matched using the nearest neighbors. Then, some n most similar photos are shortlisted based on the nearest neighbors. These photos are eventually re-matched using feature vectors from the local dynamic extraction method. The feature vectors are matched using the L_1-distance measure. The experimental results for The Chinese University of Hong Kong (CUHK) Face Sketch Database (CUFS) and CUHK Face Sketch FERET Database (CUFSF) datasets indicate that the proposed method outperforms the state-of-the-art methods.

INDEX TERMS Cascaded feature extraction, identity of interest, local feature extraction, sketch to photo, static and dynamic, deep learning.

I. INTRODUCTION

In a criminal investigation, identifying the Identity of Interest (IoI) automatically from large mugshots merely based on a facial sketch can help to speed up the process of suspect apprehension. The facial sketch of the suspect is generated when there is no other evidence at hand except the descriptions elicited from the victim or eyewitness. The forensic artist draws a sketch based on the descriptions given. This obviously causes the resulting face sketch to be less accurate and prone to shape exaggeration. Therefore, retrieving a photo from its corresponding face sketch is an extremely challenging task. Due to the modality difference in the image production, there are two main approaches used by researchers to reduce this modality gap. In the first approach, researchers [1]–[16] attempt to close this gap by generating a pseudo-image such that both images are in the same modality. It is followed by a feature extraction and matching process in the same modality. This approach is called the intra-modality approach. In the second approach, researchers [17]–[23] describe the image using features that are invariant to modality difference. The feature extraction is performed in different modalities, and the matching process is based on these features. This approach is called the inter-modality approach.

In the former approach, a synthetic photo or sketch (i.e., pseudo-photo or pseudo-sketch, respectively) is generated using an advance synthesizing algorithm. This is usually computationally complex. In addition to its complexity, the transformation algorithms try their best to transform the image from one modality to another. This conversion is intuitively naive in which it preserves the shape of the image being transformed. In the case where the generated face sketch contains shape exaggeration, matching these images using insensitive shape exaggeration descriptors may result in a low matching rate (i.e., although the matching procedure is executed in the same modality). As for the latter approach, researchers mostly focus on seeking modality-invariant

The associate editor coordinating the review of this manuscript and approving it for publication was Michele Nappi.

features to represent the image. In general, there are two main approaches to extract features: handcrafted feature [18]–[22] and deep learning feature [24]–[28]. Here, we attempt to contribute to the handcrafted feature approach. In this approach, shape exaggeration effects are generally not considered. Furthermore, the local feature extraction approach is the most popular method to extract features. However, the local features are usually extracted from static patches (i.e., the image is divided into some equal size of overlapping patches). Consequently, the extracted feature from a patch with exaggerated shape may be inaccurate and hence the similarity measure is made based on improper feature vectors. This may eventually degrade the retrieval rate.

The Difference of Gaussian Oriented Gradient Histogram (DoGOGH) [29] has been demonstrated to be effective for face sketch to photo matching with illumination effects. However, in the proposed method, the shape exaggeration effect is not treated well. To overcome the limitations mentioned above, we propose that the feature vector is extracted in a cascaded fashion by combining static and dynamic local feature extraction. By doing this, the feature vector of the photo can be reconstructed (i.e., using dynamic local feature extraction) according to the local features from the face sketch. If the feature vector construction is constructed based on the appropriate image patches, then the retrieval rate can be increased due to the fact that the patch comparison is made of the appropriate pairs. Two baseline datasets (i.e., Chinese University of Hong Kong (CUHK) Face Sketch Database (CUFS) and CUHK Face Sketch FERET Database (CUFSF)) are studied in the experiment to demonstrate the efficacy of the proposed method. This is because the sketches in the CUFS dataset have slight shape exaggeration while the sketches in the CUFSF dataset have more shape exaggeration and thus are closer to real forensic sketches.

The contributions of this paper are twofold. First, we introduce a dynamic local feature extraction method. To the best of our knowledge, no other local feature extraction method in the literature uses dynamic extraction. Secondly, we propose cascaded local feature extraction involving a static and dynamic extraction method. The rest of this paper is organized as follows. Section II and Section III discuss and explain the related work and the proposed method, respectively. Section IV elaborates the experimental setup and discusses the results obtained. A conclusion is drawn in Section V.

II. RELATED WORK

Traditionally, the process of searching potential suspects is performed manually. A large number of photographs need to be browsed by an eyewitness before selecting a few selected candidates. This process is very time-consuming and may not be accurate due to the fact that the environment may interfere with the eyewitness' focus, or they may experience fatigue while browsing the photographs. Assisting law enforcement to narrow down the criminal suspects is among the applications of interest. This is done by automatic matching of a sketch at hand (i.e., when there is no other evidence) to photos in the mugshot database. One of the techniques used to create a criminal face sketch is by sketching it on paper using a pencil. Lois Gibson and Karen Taylor are well-known forensic artists involved in this kind of sketching [30], [31]. With the aid of eyewitness descriptions, the artists visualize the face in their mind and translate it into a sketch by obeying a specific procedure as in [32]. The sketch is eventually digitized using an electronic scanner before the matching algorithm is employed.

To find the match automatically, Uhl and Lobo [33] started to use Eigenface and Principle Component Analysis (PCA) to match forensic sketches to photos. The proposed method uses geometric alignment for image normalization and patch level matching. As for research advancement and the fact that forensic sketches are often confidential, Tang and Wang took the initiative to create a public dataset called CUFS [1]. This is a clean dataset because the sketches have only a small degree of shape exaggeration. Then, to make the sketches closer to real forensic sketches, Zhang et al. introduced another dataset named CUFSF [34]. The sketches have more shape exaggeration with the corresponding photo exposed to lighting variations. Based on these initiatives, many researchers continue proposing the state-of-the-art methods to obtain better recognition accuracy. From the literature, it is noted that the proposed methods can be divided into intra-modality and inter-modality approaches.

In the intra-modality approach, to match the images, the image from one modality is transformed to another modality (as a synthetic image called pseudo-image) at the preprocessing stage. Then, the matching algorithm is applied to these images. This approach has been pioneered by Tang and Wang [1], [2], [5] and their following researchers [3], [6]. The approach was expanded by Gao *et al.* [4] and succeeding researchers [7], [8], [10]–[12], [16], [35]. This approach has been surveyed comprehensively by Wang *et al.* [36]. Tang and Wang [2] proposed the Eigensketch transformation algorithm to transform a photo into a sketch prior to matching. Liu *et al.* [3] proposed a synthesizing technique that employs Kernel-based Nonlinear Discriminant Analysis (KNDA) by preserving the local geometry. Gao *et al.* [4] synthesized the sketches using Embedded Hidden Markov Models (E-HMM) that have the capability to model the nonlinear relationship between a sketch and photo. Later, Wang and Tang [5] proposed a synthesizing model based on Markov Random Fields (MRF) to synthesize sketch to photo or vice versa. Zhang et al. improved this model to work under pose and lighting variations. Gao *et al.* [37] proposed Sparse Neighbor Selection (SNS) to render the initial pseudo-image and then used Sparse Representation-based Enhancement (SRE) to improve the quality of the synthesized image. To minimize the empirical loss for training samples, Wang *et al.* [8] introduced a probability graphic model and transductive learning. Peng *et al.* [9] utilized MRF to learn multiple representations and alternating optimization strategy and then proposed a Superpixel-based synthesis method [10]. Recently, Wang et al. proposed several

frameworks for face sketch synthesis that achieved better performance than the state-of-the-art methods [12]–[15]. This was then followed by Cao *et al.* [35] who proposed Asymmetric Joint Learning (AJL) that attempts to cater for image discrepancies due to the modality difference. Other researchers have explored a deep learning approach to synthesize the image [11], [38], [39].

In the inter-modality approach, to match the images, the modality-invariant features are extracted from the images prior to the similarity computation. This approach skips the transformation or synthesizing procedure at the preprocessing stage. The extracted features are usually discriminative and invariant across modalities [18]–[22]. Generally, there are two main approaches to extract features: handcrafted feature [18]–[22] and deep learning feature [24]–[28]. Here, we attempt to contribute to the handcrafted feature approach. Klare and Jain [17] proposed a local feature extraction approach using a Scale Invariant Feature Transform (SIFT) descriptor. To improve the accuracy, Klare *et al.* [18] extended their approach by fusing Multiscale Local Binary Pattern (MLBP) and SIFT with Local Feature Discriminant Analysis (LFDA). Zhang *et al.* [34] proposed a new face descriptor based on Coupled Information-Theoretic Encoding (CITE). Recently, Roy and Bhattacharjee [40] proposed a Local Gradient Fuzzy Pattern (LGFP) for sketch to photo matching. This was then followed by Peng *et al.* [41], [42] who proposed a method that takes into account the facial spatial structure while extracting the features for matching.

To describe the image, most of the researchers in the inter-modality approach (i.e., for handcrafted feature) utilize local feature extraction as in [18]. The image is divided into patches of the same size, and features are extracted locally from each patch (i.e., patch by patch). Then, to represent the image, these features are concatenated to make up the feature vector. Our proposed method follows this approach. We extend the local extraction method from purely static patch to a cascaded static and dynamic patch based on its nearest neighbor similarity distance. This is to ensure that the extracted features are immune to slight shape exaggeration.

III. PROPOSED METHOD

Matching a face sketch to a photo using local feature extraction has shown promising accuracy [18]. It extracts the feature from local patches throughout the image, and the extracted features are concatenated to make up a full feature vector that represents the image. In this paper, the proposed method attempts to address the effect of shape exaggeration by cascaded static and dynamic local feature extraction methods so that the constructed feature vectors are built based on the correct patches. Note that before the feature extraction process, all the face images are aligned such that the fiducial points are positioned at the predetermined reference points (this is explained in Section III-A). In this work, the static local feature is defined as the extracted local feature from a fixed patch without considering the features of its neighboring patches while the dynamic local feature is defined as the extracted local feature from a selected patch within a specified neighboring patch distance. First, the feature vectors that are locally extracted using the static extraction method (refer to Section III-B) for a sketch and photo are matched using nearest neighbors. Then, some n most similar photos are shortlisted based on the nearest neighbors (using L_1-distance). These photos are eventually re-matched using the local feature vectors extracted using the dynamic extraction method (refer to Section III-C). The following subsections elaborate more on each process.

A. FACE ALIGNMENT

Matching face images without proper alignment may result in a poor recognition rate. Aligning faces with respect to some pre-defined points in common across the images is the solution. Most researchers normally perform 2D transformation (i.e., translation, rotation and scaling) of faces with reference to the centers of the eyes [17], [18]. However, according to Klare *et al.* [18], for forensic sketches, the inner face regions like eyes are less salient than the outer face regions. This is because the outer regions carry salient features and therefore they are more discriminative than the inner regions. Based on these findings, here, the face images are aligned using three fiducial points from the outer regions as proposed in [29].

B. STATIC LOCAL FEATURE EXTRACTION

After the image is aligned properly, local feature vectors are extracted across the image. Here, the Difference of Gaussian Oriented Gradient Histogram (DoGOGH) [29] is used to extract the features. A detailed description of the extraction method can be obtained from the respective literature. Algorithm 1 revisits the DoGOGH algorithm that is used in our proposed method.

C. DYNAMIC LOCAL FEATURE EXTRACTION

A rendered sketch has typically some degree of shape exaggeration (especially a viewed sketch in the case where the face is detected automatically, or a forensic sketch) that makes some parts of the face geometrically misaligned (as illustrated in Fig. 1). This effect can be observed in both datasets used in this work. It may result in a low recognition rate. If the feature vector construction is built based on appropriate patches, it may increase the recognition rate due to the fact that the patch comparison is made of the correct pairs. An example is illustrated in Fig. 2. From the figure, it can be clearly seen that some of the neighboring patches may have a higher similarity score (i.e., smaller L_1-distance) compared to the patch at the origin. This is due to some degree of patch misalignment. Based on this observation, the proposed method attempts to cater for this problem by extracting the local features dynamically.

Dynamic local feature extraction extracts feature vectors within a specified distance from the patch of interest (on a photo) dynamically based on a reference feature vector (extracted from a sketch) at the same patch position.

Algorithm 1 : Static DoGOGH

Input: Grayscale image, $I(x, y)$.

***Step* 1**: *Intensity Correction*. To cater for lighting variation, lighten the dark regions by using (1).

$$\hat{I}(x, y) = \log(I(x, y)). \tag{1}$$

***Step* 2**: *Image Transformation*. Transform the image in Step 1, $\hat{I}(x, y)$ into the Difference of Gaussian (DoG) image, $\hat{I}_{dog}(x, y)$ using (2) and (3). Note that two different sigma are used here.

$$G_\sigma(x, y) = \frac{1}{\sigma\sqrt{2\pi}} \exp^{\frac{x^2+y^2}{2\sigma^2}} \tag{2}$$

$$\hat{I}_{dog}(x, y) = \hat{I}(x, y) * (G_{\sigma_1}(x, y) - G_{\sigma_2}(x, y)) \tag{3}$$

***Step* 3**: *Histogram of Oriented Gradient (HOG)*. On the $\hat{I}_{dog}(x, y)$ image, compute the HOG features by binning the pixel magnitude from (4) according to the orientation from (5).

$$|\hat{I}_{dog}(x, y)| = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \tag{4}$$

$$\theta\hat{I}_{dog}(x, y) = tan^{-1}\frac{G_y(x, y)}{G_x(x, y)} \quad \in [-180, 180] \tag{5}$$

where $G_x(x, y)$ and $G_y(x, y)$ are computed using (6) and (7), respectively.

$$G_x(x, y) = \frac{\partial\hat{I}_{dog}(x, y)}{\partial x} \tag{6}$$

$$G_y(x, y) = \frac{\partial\hat{I}_{dog}(x, y)}{\partial y} \tag{7}$$

***Step* 4**: *Local Feature Extraction*. The image, $\hat{I}_{dog}(x, y)$ is divided into 50 percent overlapping patches of the same size, $N \times N$. The HOG feature is extracted locally from each patch (i.e., patch by patch) as in Step 3. Let $P = [p_a, \ldots, p_M]$ and f_a be the patches and the HOG feature of a patch, respectively. Here, $a = 1, 2, \ldots, M$ and M is the total number of patches. Each HOG feature from each patch is normalized using (8).

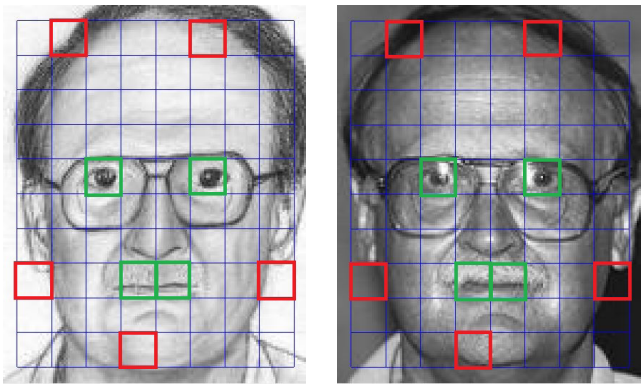
$$\hat{f}_a = \frac{f_a}{\|f_a\|_1 + \epsilon} \tag{8}$$

where ϵ is a small constant.

***Step* 5**: *Feature Vector Construction*. To represent the image, the features from all patches in Step 4 are concatenated to make up the feature vector, $F = [\hat{f}_a, \ldots, \hat{f}_M]$.

Output: F.

The process of obtaining a DoG image is similar to the static feature extraction method (refer to Algorithm 1). Let $\hat{I}^S_{dog}(x, y)$ and $\hat{I}^P_{dog}(x, y)$ be the DoG image for a sketch and photo, respectively. Also, the DoG image is then divided equally into a set of M small overlapping patches. Let $F^S = [f^S_a, \ldots, f^S_M]$ and $F^P = [f^P_a, \ldots, f^P_M]$ denote the feature vector extracted from image $\hat{I}^S_{dog}(x, y)$ and $\hat{I}^P_{dog}(x, y)$, respectively. Here, a is the patch number and $a = 1, 2, \ldots, M$. To dynamically extract a feature vector f^P_a, let's consider $\hat{F}^P = [f^P_{ab}, \ldots, f^P_{ML}]$. L is the total number of patches (i.e., the target and its neighboring patches) and $b = 1, 2, \ldots, L$. The extraction method is illustrated in Fig. 3. First, the center pixel of each patch is assigned as a target point. Then, for the sketch, HOG feature F^S is extracted on these points while for the photo, HOG feature $\hat{F}^P$ is extracted on these points and its neighboring points. By doing so, at every single patch, one feature vector f^S_a (as reference) and L number of feature vectors f^P_{ab} (coverage depends on the maximum pixel d_p distance) are extracted. Based on these feature vectors, the distances between f^S_a and f^P_{ab} are computed using nearest neighbours (i.e., L_1-distance). The feature vector from f^P_{ab} that has the smallest L_1-distance against f^S_a is chosen to represent the current patch feature vector f^P_a. This process is reiterated for all patches within the image to construct F^P. Algorithm 2 shows the extraction details.



(a) (b)

FIGURE 1. Example image pair with the sketch (a), has shape exaggeration as compared to its corresponding photo (b). The patches at the inner regions (the examples are highlighted in green) indicate that the patches are properly aligned while the patches at the outer regions (the examples are highlighted in red) indicate that the patches are not really well aligned. This image pair is from the CUHK Face Sketch FERET Database (CUFSF).

D. CASCADED STATIC AND DYNAMIC LOCAL FEATURE EXTRACTION

Matching face sketches to photos using static local features may not yield good accuracy because of the shape exaggeration effects. Extracting local features dynamically may result in better accuracy but requires an extremely long extraction time. To address this, we propose to combine the static feature extraction method and the dynamic feature extraction method in a cascaded fashion. The static feature is used to shortlist n nearest candidates so that the dynamic feature extraction only extracts features on a few strong candidates. Fig. 4 shows the proposed method.

E. SIMILARITY MEASURE

In order to match the features, we use nearest neighbors. The L_1-distance metric is used in this work. The matching

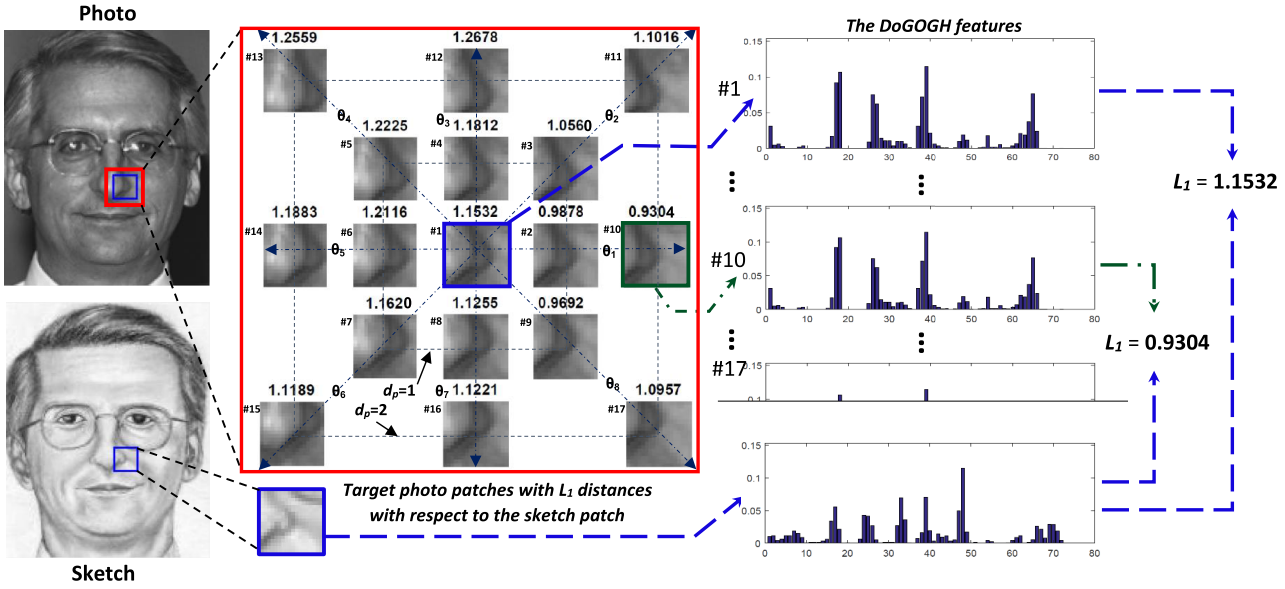


FIGURE 2. An example illustration of L_1-distance measured between a target patch from a face sketch and its corresponding patch from a photo. The distance of the corresponding patch from the photo (i.e., patch #1) is larger than some of its neighboring patches. The neighboring patches are built from the patch that is shifted a few pixels away from its origin in eight different directions (i.e., $\theta_1, \theta_2, \ldots, \theta_8$) within a predetermined maximum pixel d_p distance. Patch #10 gives the smallest L_1-distance to indicate that it has a higher similarity score to the current sketch patch, hence becoming the correct candidate to extract the feature from.

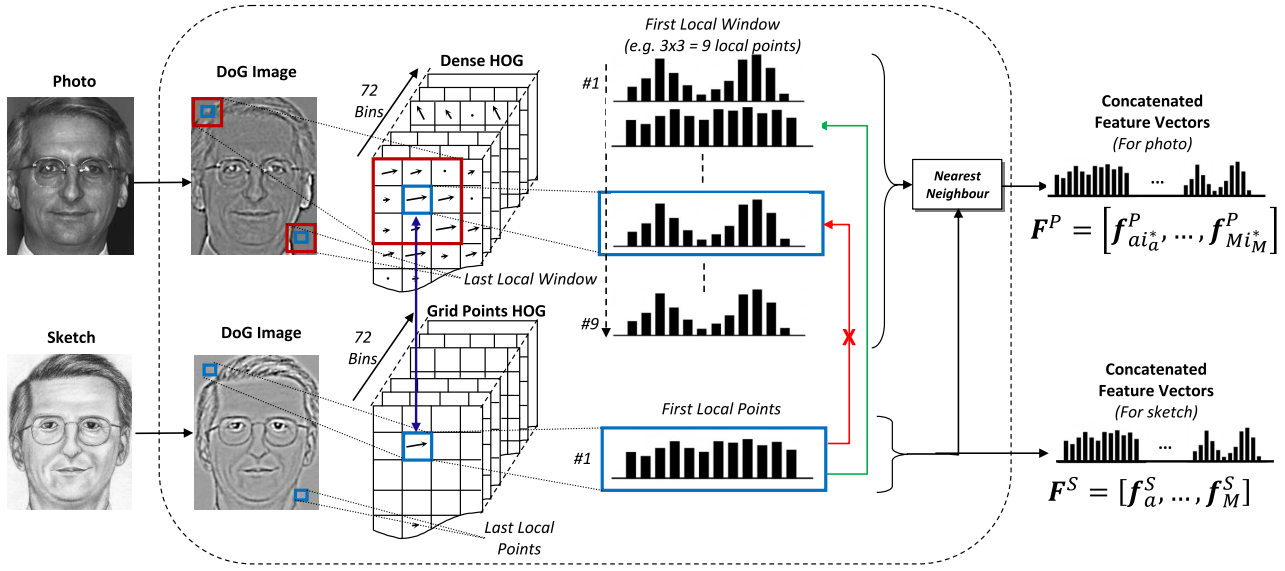


FIGURE 3. The proposed dynamic local feature extraction for the face sketch-to-photo matching system. This is to address problems with regard to shape exaggeration. The region of interest (red box) shows 3x3 pixels that correspond to 8 neighbors of one pixel distance d_p from the center (blue box). The arrow with cross sign (red line) shows static feature extraction while the other arrow (green line) indicates dynamic feature extraction (i.e., selection is based on the nearest neighbor). Only eight neighbors are considered for each pixel distance d_p to cater for eight different exaggerated directions. Example: if $d_p = 1$, we will have 9 (1+8) local points; if $d_p = 2$, we will have 17 (1+8+8) local points; if $d_p = 3$, we will have 25 (1+8+8+8) local points; and so on.

algorithm is computed as in Algorithm 3. Note that G is the total number of photos in the gallery during the first stage (static) while in the second stage, $G = n$ where n is the shortlisted photos from the gallery given from the previous stage (static).

IV. EXPERIMENTS

Two datasets are used to evaluate the effectiveness of the proposed method: CUHK Face Sketch Database (CUFS) and CUHK Face Sketch FERET Database (CUFSF). This is because the sketches in the CUFS dataset have a slight degree of shape exaggeration while the sketches in the CUFSF dataset have a higher degree of shape exaggeration and thus are closer to real forensic sketches. These datasets are from the *Viewed Sketch* category. Note that a *Viewed Sketch* is defined as a sketch that is rendered while the forensic artist is viewing the photograph of the subject or the real subject itself.

Algorithm 2 : Dynamic DoGOGH

Input: The DoG image, $\hat{I}^S_{dog}(x, y)$ for a sketch, and the DoG image, $\hat{I}^P_{dog}(x, y)$ for a photo. Target points (i.e., patch centroids) $D = [D_a, \ldots, D_M]^T = [(x_a, y_a), \ldots, (x_M, y_M)]^T$.

***Step* 1**: *Add Neighboring Points*. Let $J_l(x, y) = [(x+l, y), (x+l, y+l), (x, y+l), (x-l, y+l), (x-l, y), (x-l, y-l), (x, y-l), (x+l, y-l)]$ where l is the number of pixels off from the center and $l = 1, 2, \ldots, d_p$.

$$\hat{D}_a = [D_a, J_l(x_a, y_a), \ldots, J_{d_v}(x_a, y_a)]^T] \tag{9}$$

***Step* 2**: *Extract Features*. Extract sketch features f^S_a from $\hat{I}^S_{dog}(x, y)$ at D_a and photo features f^P_{ab} from $\hat{I}^P_{dog}(x, y)$ at $\hat{D}_a$ where b is the index of a neighboring patch.

***Step* 3**: *Features Reconstruction*. Find the smallest L_1-distance, i^*_a between f^S_a and f^P_{ab} as in (10).

$$i^*_a = arg\min_b \|f^S_a - f^P_{ab}\|_1 \tag{10}$$

and hence perform reconstruction by concatenation:

$$F^P = [f^P_{ai^*_a}, \ldots, f^P_{Mi^*_M}] \tag{11}$$

Output: F^P.

Algorithm 3 : Similarity Measure

Input: Feature vector for sketch, F^S, and the feature vector for photo, F^P_g. Here, $g = 1, 2, \ldots, G$ where G is the total number of photos to be matched.

***Step* 1**: Calculate the L_1-distance d_g between F^S and F^P_g as follows:

$$d_g = \|F^S - F^P_g\|_1 \tag{12}$$

***Step* 2**: The L_1-distance, d_g in Step 1 is sorted in ascending order and is d_{g_s} where g_s is the sorted indexes.

Output: g_s.

A. DATABASES

Two *Viewed Sketch* datasets are elaborated here: CUFS and CUFSF. The CUFS dataset [1], [5] contains 606 image pairs from the CUHK student dataset [43], AR dataset [44] and XM2VTS dataset [45] (i.e., 188, 123 and 295 image pairs, respectively). All images were the frontal view. The photographs were taken without lighting variation and with a neutral expression. Due to the XM2VTS dataset not being freely available, it was not included in this study. Overall, for testing, only 311 image pairs from CUHK and AR datasets were used. Fig. 5 (a) and (b) show the example image pair from this dataset. The CUFSF dataset [5], [34] was prepared based on 1,194 photographs from the FERET database [46]. All images were the frontal view. The sketches were rendered with shape exaggeration and the photographs were mostly exposed to lighting variation. Fig. 5 (c) shows the example image pair from this dataset. To evaluate the proposed method, all available samples were used as there is no training required.

B. EXPERIMENTAL SETUP

The images were affine transformed to a fixed reference point $\boldsymbol{r} = [r_1, r_2, r_3] = [(15, 80), (126, 80), (71, 161)]$. Then, the images were cropped (center-based) using a window size of 175 × 140. For the image transformation as in Algorithm 1 Step 2, the two different widths σ_1 and σ_2 were set to 1 and 2, respectively. A 50 percent overlapping patch of size 16 × 16 was used in this experiment. With this setting, the total patches M per image was 320. To extract the HOG feature, the number of allocated orientation bins α was 18 and each patch was divided into 4 cells to yield a $72M$ concatenated feature vector. To evaluate the feasibility of some popular local descriptors that may outperform the DoGOGH, the evaluation was extended to the descriptors: Histograms of Oriented Gradients (HOG) [47], Speeded Up Robust Features (SURF) [48], Scale-Invariant Feature Transform (SIFT) [49], and Multiscale Local Binary Patterns (MLBP) [18]. Similarly, each local descriptor was extracted from the same number of patches M based on the

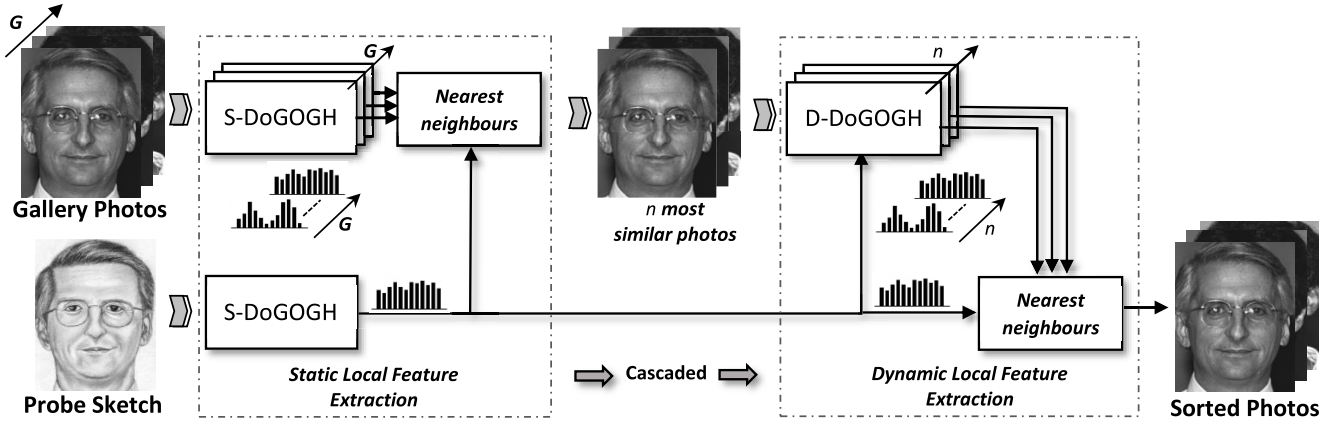


FIGURE 4. The proposed cascaded static and dynamic local feature extraction for the face sketch-to-photo matching system.

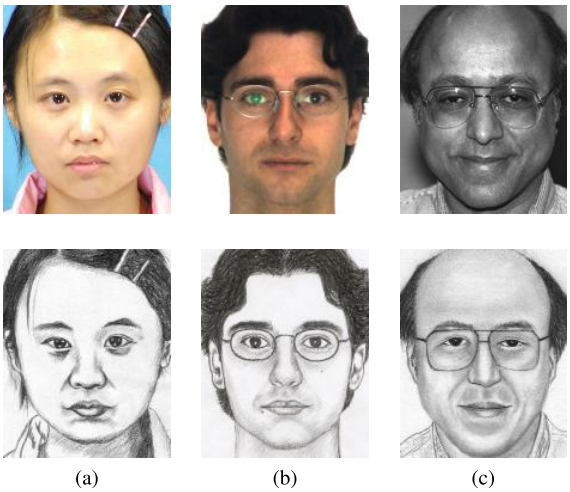


FIGURE 5. Examples of facial viewed sketch pairs; (a) CUHK, (b) AR, (c) FERET. (a) and (b) are CUFS and (c) is CUFSF.

50 percent overlapping patch of size 16 × 16. The MLBP, SIFT, SURF and HOG descriptors yielded $236M$, $128M$, $64M$, and $72M$ concatenated feature descriptors, respectively. For the HOG and SURF descriptors, the embedded function in the MATLAB toolbox was employed in this implementation, whereas for the SIFT descriptor, the feature vector was extracted using the function from an open source library [50]. The following sub-section elaborates the additional settings or parameters (if any) required in any particular experiment. The experiments were conducted using MATLAB R2016b under Windows 10 Pro 64 with a 3.6GHz quad-core processor and 16GB RAM.

C. RESULTS

The performance of the proposed method was evaluated using a Cumulative Match Curve (CMC). This is a popular evaluation method applied by most researchers in this field [9], [17], [18], [21], [40], [51]–[56]. It accumulates the rate of correct identity across the ranks. As an example, let's consider a classification problem with nearest neighbors using L_1-distance. The output is an array of distances between the probe and gallery images. Next, the match is computed based on its nearest neighbor (i.e., the smallest distance between the probe and gallery images). Then the percentage of correct identity is accumulated across the ranks. From the ranks, rank-1 accuracy indicates the percentage that the correct match can be retrieved merely based on the smallest distance (similar to that recognition rate), whereas the rank-10 accuracy gives the retrieval rate such that the correct match can be retrieved within the first ten smallest distances. Based on this fact, if the rank-1 percentage is at 100%, it demonstrates that the method is capable of identifying the subject without error arising. Similarly, if the accumulated percentage progressively increases to achieve 100% at rank-10, it means that the correct match can be retrieved within the top 10 matches.

Table 1 shows the rank-1 accuracies across n and d_p ranges. The matching accuracy of the cascaded static and dynamic local feature extraction relies on the number of n and d_p (coverage pixels away from its center pixel). Here, the n-range is limited to 10 and the d_p-range is limited to 8 pixels with a step of 2. Considering the dataset with almost no shape exaggeration, i.e., CUFS, the results suggest that the best d_p is 2 regardless of the number of n. With these settings, the accuracy achieved 100%. For a higher degree of shape exaggeration dataset, i.e., CUFSF, when the d_p was set to 4, it demonstrated optimal accuracy across n in comparison with the other values of d_p. The results also suggest that the best d_p and n are 4 and 7, respectively. With these settings, the accuracy achieved 89.03%. Furthermore, regardless of the number of n, the results indicate that the accuracy improves when the maximum pixels d_p is slightly increased by a few pixels and begins to degrade when d_p is increased further (the patch is scanned too far from its origin). In terms of the number of n, theoretically, rank-1 matching accuracy using the static extraction method can still be improved up to rank-n matching accuracy but not beyond it. However, n cannot be too large as it suffers from an extremely slow extraction rate as well as losing discriminative features (due to a higher chance of obtaining too many similar patches from a large sample).

A performance comparison of the proposed method with the state-of-the-art method is tabulated in Table 2.

TABLE 1. Rank-1 accuracy comparison across n and d_p ranges on CUFS and CUFSF datasets. The results in bold indicate the highest accuracy for each n.

Maximum Pixels (d_p)	n								
	2	3	4	5	6	7	8	9	10
CUHK Face Sketch Database (CUFS)									
2	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
4	99.68	99.36	99.36	99.36	99.36	99.36	99.36	99.36	99.36
6	99.04	98.71	98.71	98.71	98.71	98.71	98.71	98.71	98.71
8	99.36	99.04	99.04	99.04	99.04	98.71	98.71	98.71	98.71
CUHK Face Sketch FERET Database (CUFSF)									
2	86.77	87.60	87.77	87.94	88.02	88.02	87.94	87.94	87.94
4	**87.19**	**87.77**	**88.36**	**88.78**	**88.94**	**89.03**	**88.69**	**88.69**	**88.78**
6	86.93	87.19	88.27	88.44	88.27	88.53	88.11	88.19	88.19
8	86.26	86.26	87.19	87.02	86.60	86.77	86.60	86.68	86.68

TABLE 2. Rank-1 CMC accuracy (%) of state-of-the-art methods on CUFS and CUFSF datasets. The accuracies are taken from the respective literature except for the deep learning pre-trained model accuracies.

State-of-the-art methods	No. training samples	No. testing samples	Rank-1 accuracy (%)
CUHK Face Sketch Database (CUFS)			
Intra-Modality:			
MMRF + RS-LDA [5]	306	300	96.30
SNS-SRE [37]	306	300	96.50
TFSP + RS-LDA [8]	306	300	97.70
MrFSPS + RS-LDA [9]	306	300	97.70
S-FSPS + LDA [10]	306	300	99.10
Fully Convolutional Network [39]	88	100	100
Inter-Modality:			
CITE [34]	306	300	99.87
SIFT + MLBP [18]	306	300	99.47
*VGG Face Descriptor [26]	0	311	55.95
*Light CNN [27]	0	311	89.71
*ArcFace [28]	0	311	13.83
DoGOGH [29]	0	311	100
C-DoGOGH	0	311	**100**
CUHK Face Sketch FERET Database (CUFSF)			
Intra-Modality:			
TFSP + RS-LDA [8]	500	694	72.62
MrFSPS + RS-LDA [9]	500	694	75.36
S-FSPS + LDA [10]	500	694	72.19
Inter-Modality:			
CITE [34]	500	694	89.54
*VGG Face Descriptor [26]	0	1194	36.01
*Light CNN [27]	0	1194	38.27
*ArcFace [28]	0	1194	8.21
DoGOGH [29]	0	1194	83.75
C-DoGOGH	0	1194	**89.03**

* The deep learning pre-trained models were used to extract the features.

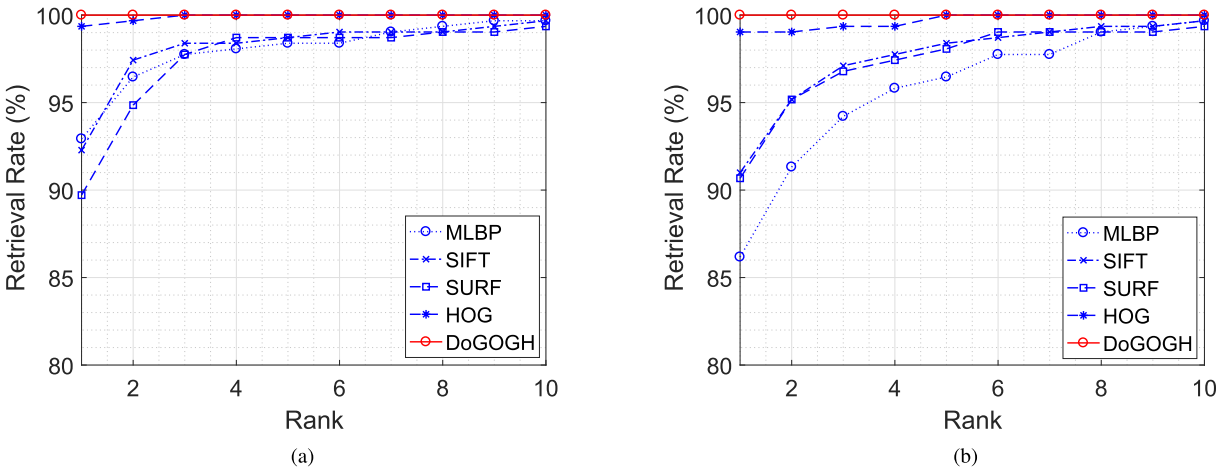


FIGURE 6. Retrieval rate (CMC) comparison of the DoGOGH and several popular local descriptors evaluated on the CUFS dataset using (a) static, and (b) cascaded static and dynamic local feature extraction.

From the table, the proposed method performs better than other methods (i.e., based on the reported accuracy from the respective publications). Additionally, the proposed method does not require any training or synthesizing process (i.e., to avoid the influence of synthetic image artifacts) and is thus suitable for real-time application. As the proposed method is an inter-modality approach, a comparison was also made of inter-modality approaches that include Coupled Information-Theoretic Encoding (CITE) and Scale-Invariant Feature Transform (SIFT) + Multiscale Local Binary Pattern (MLBP). Interestingly, our testing sample was larger than for the other methods.

In order to ascertain the infeasibility of the other local descriptors replacing the DoGOGH in the proposed cascaded method, the evaluation is extended such that the DoGOGH is substituted in turn by four popular local descriptors (i.e., MLBP, SIFT, SURF, and HOG). For this experiment, n was set to 10 and d_p was set to 2 and 4 for the CUFS and CUFSF datasets, respectively. Fig. 6 and Fig. 7 show the results obtained when the proposed cascaded method

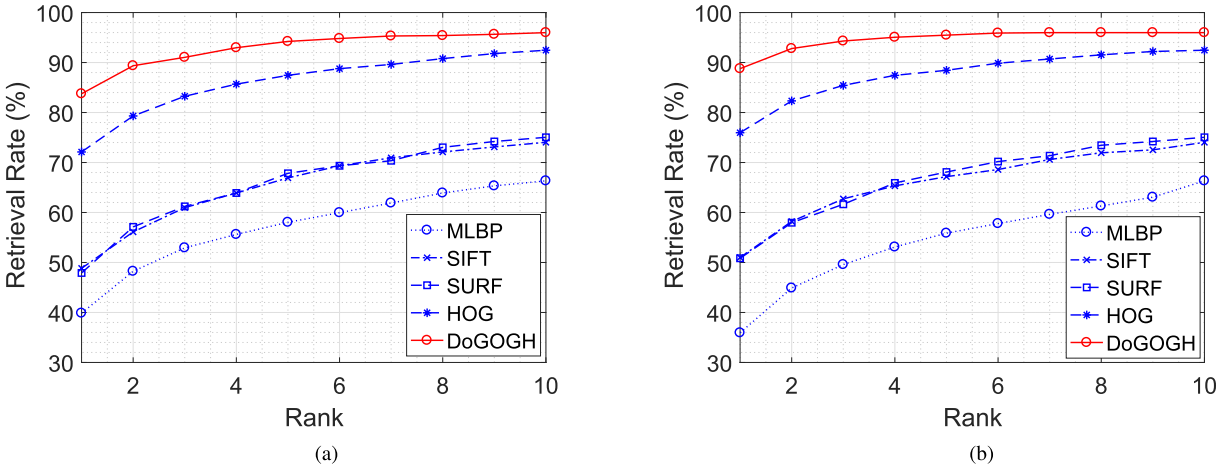


FIGURE 7. Retrieval rate (CMC) comparison of the DoGOGH and several popular local descriptors evaluated on the CUFSF dataset using (a) static, and (b) cascaded static and dynamic local feature extraction.

TABLE 3. Parameter settings for all the pre-trained CNN models used in this work to extract deep learning features and to perform the matching.

Parameter Settings	Pre-trained CNN Models		
	VGG Face Descriptor [26]	**Light CNN** [27]	**ArcFace** [28]
Architecture	VGG-Very-Deep-16 CNN	LightCNN-29	LResNet50E
Input Size	$224 \times 224 \times 3$	144×144	112×112
Feature Dimension	4096	512	512
Feature Normalization	L_2-norm	L_2-norm	L_2-norm
Distance Measure	L_1	L_1	L_1

(using different local descriptor) was tested on the CUFS and CUFSF datasets, respectively. On CUFS, the results clearly indicate that MLBP, SIFT, and SURF do not improve the accuracy but worsens it when the cascaded static and dynamic local feature extraction is employed. Similarly, when tested on CUFSF, MLBP accuracy reduced significantly while SIFT and SURF exhibited no accuracy improvement. Overall, the results demonstrate that the HOG and DoGOGH exhibited comparable accuracy on the CUFS dataset and a significant accuracy improvement on the CUFSF dataset after applying the proposed extraction method. Of these two, the DoGOGH performed better.

Due to the fact that the feature can be extracted using either handcrafted or deep learning feature, therefore it is worth to compare the performance between the two. For the deep learning feature matching approach, we adopted the Siamese CNN architecture that commonly used for tasks that involve finding the similarity between two comparable images. The implementation was based on the pre-trained models (i.e., two identical CNNs) for feature extraction and the distance metric was used to compute the similarity of the features. Figure 8 illustrates the Siamese CNN architecture used in this work. To extract the deep learning features, we employed several deep learning pre-trained models (i.e., VGG Face Descriptor [26], Light CNN [27] and ArcFace [28]) that have been trained for face recognition. This was to ensure that the model extracts appropriate features. The computation of these models was based

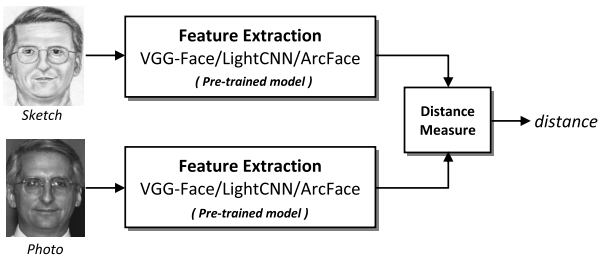


FIGURE 8. Deep Learning features matching approach based on Siamese CNN architecture used in this work for performance comparison.

on CNN implementation. In order to use these models for feature extraction, we removed the last classification layer and treated the output (i.e., at the fully connected layer) as the feature vector. Then, the resulting feature vector was normalized using L_2-norm. Refer to Table 3 for the details of each model. Once the feature vector is ready, for a fair comparison, the same similarity measure as in Algorithm 3 was employed to obtain the distances between sketches and photos. Then the rank-1 accuracies were computed based on these distances. Next, the accuracies were inserted in Table 2 for comparison. From the results obtained, Light CNN pre-trained model extracts better features as it gives the rank-1 accuracy of 89.71% and 38.27% for CUFS and CUFSF datasets, respectively. However, the proposed handcrafted feature is observed to give a better representation than the extracted deep learning feature in the context of matching sketch to photo using the simplest distance metric (i.e., L_1).

V. CONCLUSION

In this paper, we presented a new local feature extraction method based on a combination of static and dynamic local feature extraction in a cascaded manner. The results demonstrate that the proposed cascaded static and dynamic local feature extraction exhibits better accuracy in regard to matching face sketches with shape exaggeration to photos. This is because the shape exaggeration effect is addressed by employing dynamic local feature extraction for the n number of shortlisted candidates from static local feature extraction matching. Despite the fact that dynamic local feature extraction requires an exceptionally long time to extract the features and may reduce the discriminative power if applied on a large number of classes, cascaded static and dynamic local feature extraction is proposed and has been proven to solve these issues. To achieve further improvement, local feature extraction can be extracted only on some Patches of Interest (PoI), thus leading to our future work. Overall, the cascaded static and dynamic local feature extraction method exhibited better performance in comparison with a merely static approach.

REFERENCES

[1] X. Tang and X. Wang, "Face sketch synthesis and recognition," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, pp. 687–694, Oct. 2003.
[2] X. Tang and X. Wang, "Face sketch recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 50–57, Jan. 2004.
[3] Q. Liu, X. Tang, H. Jin, H. Lu, and S. Ma, "A nonlinear approach for face sketch synthesis and recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 1005–1010.
[4] X. Gao, J. Zhong, J. Li, and C. Tian, "Face sketch synthesis algorithm based on E-HMM and selective ensemble," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 4, pp. 487–496, Apr. 2008.
[5] X. Wang and X. Tang, "Face photo-sketch synthesis and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 11, pp. 1955–1967, Nov. 2009.
[6] W. Zhang, X. Wang, and X. Tang, "Lighting and pose robust face sketch synthesis," in *Computer Vision—ECCV* (Lecture Notes in Computer Science), vol. 6316, K. Daniilidis, P. Maragos, and N. Paragios, Eds. Berlin, Germany: Springer, 2010.
[7] N. Wang, X. Gao, D. Tao, and X. Li, "Face sketch-photo synthesis under multi-dictionary sparse representation framework," in *Proc. 6th Int. Conf. Image Graph.*, 2011, pp. 82–87.
[8] N. Wang, D. Tao, X. Gao, X. Li, and J. Li, "Transductive face sketch-photo synthesis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 9, pp. 1364–1376, Sep. 2013.
[9] C. Peng, X. Gao, N. Wang, D. Tao, X. Li, and J. Li, "Multiple representations-based face sketch–photo synthesis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 11, pp. 2201–2215, Nov. 2016.
[10] C. Peng, X. Gao, S. Member, N. Wang, and J. Li, "Superpixel-based face sketch–photo synthesis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 2, pp. 288–299, Feb. 2017.
[11] M. Zhu, N. Wang, X. Gao, and J. Li, "Deep graphical feature learning for face sketch synthesis," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, 2017, pp. 3574–3580.
[12] N. Wang, X. Gao, L. Sun, and J. Li, "Bayesian face sketch synthesis," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1264–1274, Mar. 2017.
[13] N. Wang, M. Zhu, J. Li, B. Song, and Z. Li, "Data-driven vs. model-driven: Fast face sketch synthesis," *Neurocomputing*, vol. 257, pp. 214–221, Sep. 2017.
[14] N. Wang, X. Gao, and J. Li, "Random sampling for fast face sketch synthesis," *Pattern Recognit.*, vol. 76, pp. 215–227, Apr. 2018.
[15] N. Wang, X. Gao, L. Sun, and J. Li, "Anchored neighborhood index for face sketch synthesis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 9, pp. 2154–2163, Sep. 2018.
[16] A. Radman and S. A. Suandi, "Robust face pseudo-sketch synthesis and recognition using morphological-arithmetic operations and HOG-PCA," *Multimedia Tools Appl.*, vol. 77, no. 19, pp. 25311–25332, 2018.
[17] B. Klare and A. K. Jain, "Sketch-to-photo matching: A feature-based approach," *Proc. SPIE*, vol. 7667, p. 766702, Apr. 2010.
[18] B. F. Klare, Z. Li, and A. K. Jain, "Matching forensic sketches to mug shot photos," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 639–646, Mar. 2011.
[19] H. K. Galoogahi and T. Sim, "Inter-modality face sketch recognition," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2012, pp. 224–229.
[20] H. K. Galoogahi and T. Sim, "Face sketch recognition by local radon binary pattern: LRBP," in *Proc. Int. Conf. Image Process. (ICIP)*, Sep./Oct. 2012, pp. 1837–1840.
[21] B. F. Klare and A. K. Jain, "Heterogeneous face recognition using kernel prototype similarities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1410–1422, Jun. 2013.
[22] M. A. A. Silva and G. C. Chávez, "Face sketch recognition from local features," in *Proc. Brazilian Symp. Comput. Graphic Image Process.*, 2014, pp. 57–64.
[23] D. Liu, J. Li, N. Wang, C. Peng, and X. Gao, "Composite components-based face sketch recognition," *Neurocomputing*, vol. 302, pp. 46–54, Aug. 2018.
[24] C. Galea and R. A. Farrugia, "Forensic face photo-sketch recognition using a deep learning-based architecture," *IEEE Signal Process. Lett.*, vol. 24, no. 11, pp. 1586–1590, Nov. 2017.
[25] D. Liu, N. Wang, C. Peng, J. Li, and X. Gao, "Deep attribute guided representation for heterogeneous face recognition," in *Proc. 27th Int. Joint Conf. Artif. Intell. (IJCAI)*, 2018, pp. 835–841.
[26] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Proc. Brit. Mach. Vis. Conf.*, 2015, vol. 1, no. 3, p. 6.
[27] X. Wu, R. He, Z. Sun, and T. Tan, "A light CNN for deep face representation with noisy labels," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 11, pp. 2884–2896, Nov. 2018.
[28] J. Deng, J. Guo, and S. Zafeiriou. (2018). "ArcFace: Additive angular margin loss for deep face recognition." [Online]. Available: https://arxiv.org/abs/1801.07698
[29] S. Setumin and S. A. Suandi, "Difference of Gaussian oriented gradient histogram for face sketch to photo matching," *IEEE Access*, vol. 6, pp. 39344–39352, 2018.
[30] J. Sommer. (2015). *10 Incredibly Realistic Sketches by the World's Most Successful Forensic Artist*. Accessed: Oct. 10, 2016. [Online]. Available: http://www.businessinsider.com/10-sketches-by-forensic-artist-lois-gibson-2015-7
[31] K. T. Taylor, *Forensic Art and Illustration*. Boca Raton, FL, USA: CRC Press, 2001.
[32] L. Gibson, *Forensic Art Essentials: A Manual for Law Enforcement Artists*. Amsterdam, The Netherlands: Elsevier, 2008.
[33] R. G. Uhl, Jr., and N. da Vitoria Lobo, "A framework for recognizing a facial image from a police sketch," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 1996, pp. 586–593.
[34] W. Zhang, X. Wang, and X. Tang, "Coupled information-theoretic encoding for face photo-sketch recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 513–520.
[35] B. Cao, N. Wang, X. Gao, and J. Li, "Asymmetric joint learning for heterogeneous face recognition," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 6682–6688.
[36] N. Wang, D. Tao, X. Gao, X. Li, and J. Li, "A comprehensive survey to face hallucination," *Int. J. Comput. Vis.*, vol. 106, no. 1, pp. 9–30, 2014.
[37] X. Gao, N. Wang, D. Tao, and X. Li, "Face sketch-photo synthesis and retrieval using sparse representation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 8, pp. 1213–1226, Aug. 2012.
[38] L. Jiao, S. Zhang, L. Li, F. Liu, and W. Ma, "A modified convolutional neural network for face sketch synthesis," *Pattern Recognit.*, vol. 76, pp. 125–136, Apr. 2018.
[39] L. Zhang, L. Lin, X. Wu, S. Ding, and L. Zhang, "End-to-end photo-sketch generation via fully convolutional representation learning," in *Proc. 5th ACM Int. Conf. Multimedia Retr.*, 2015, pp. 627–634.
[40] H. Roy and D. Bhattacharjee, "Face sketch-photo recognition using local gradient checksum: LGCS," *Int. J. Mach. Learn. Cybern.*, vol. 8, no. 5, pp. 1457–1469, 2017.
[41] C. Peng, X. Gao, N. Wang, and J. Li, "Graphical representation for heterogeneous face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 2, pp. 301–312, Feb. 2017.
[42] C. Peng, X. Gao, N. Wang, and J. Li, "Sparse graphical representation based discriminant analysis for heterogeneous face recognition," *Signal Process.*, vol. 156, pp. 46–61, Mar. 2019.

[43] X. Tang and X. Wang, "Face photo recognition using sketch," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2002, pp. 257–260.
[44] A. M. Martinez, "The ar face database," CVC, India, New Delhi, Tech. Rep. 24, 1998.
[45] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, "XM2VTSDB: The extended M2VTS database," in *Proc. 2nd Int. Conf. Audio Video-Based Biometric Person Authentication*, vol. 24, 1999, pp. 72–77.
[46] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, Oct. 2000.
[47] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, vol. 1, no. 1, pp. 886–893.
[48] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.
[49] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[50] A. Vedaldi and B. Fulkerson. (2008). *VLFeat: An Open and Portable Library of Computer Vision Algorithms*. Accessed: Nov. 2, 2017. [Online]. Available: http://www.vlfeat.org/
[51] H. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa, "On matching sketches with digital face images," in *Proc. 4th IEEE Int. Conf. Biometric, Theory, Appl. Syst. (BTAS)*, Sep. 2010, pp. 1–7.
[52] S. Klum, H. Han, A. K. Jain, and B. Klare, "Sketch based face recognition: Forensic vs. composite sketches," *Proc. Int. Conf. Biometrics*, 2013, pp. 1–8.
[53] S. J. Klum, H. Han, B. F. Klare, and A. K. Jain, "The FaceSketchID system: Matching facial composites to mugshots," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 12, pp. 2248–2263, Dec. 2014.
[54] P. Mittal, M. Vatsa, and R. Singh, "Composite sketch recognition via deep network—A transfer learning approach," in *Proc. Int. Conf. Biometrics*, 2015, pp. 251–256.
[55] S. Ouyang, T. Hospedales, Y. Z. Song, and X. Li, "Cross-modal face matching: beyond viewed sketches," in *Proc. Asian Conf. Comput. Vis.* Cham, Switzerland: Springer, Nov. 2014, pp. 210–225.
[56] Z. Chen, K. Wang, and C. Liu, "Fast face sketch-photo image synthesis and recognition," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 30, no. 10, pp. 1656008.1–1656008.13, 2016.

SAMSUL SETUMIN received the B.Eng. degree (Hons.) in electronic engineering from the University of Surrey, in 2006, and the M.Eng. degree in electrical-(electronic and telecommunication) from Universiti Teknologi Malaysia, in 2009. He is currently pursuing the Ph.D. degree with Universiti Sains Malaysia. Since 2010, he has been a Lecturer with Universiti Teknologi MARA Pulau Pinang, Malaysia. He was formerly a Test Engineer with Agilent Technologies (M) Sdn. Bhd. and later attached to Intel Microelectronics (M) Sdn. Bhd. for a year. His research interests include computer vision, image processing, and pattern recognition.

SHAHREL AZMIN SUANDI received the B.Eng. degree in electronic engineering, and the M.Eng. and D.Eng. degrees in information science from the Kyushu Institute of Technology, Fukuoka, Japan, in 1995, 2003, and 2006, respectively. He is currently an Associate Professor with the School of Electrical and Electronic Engineering, Universiti Sains Malaysia, Engineering Campus, Nibong Tebal, Malaysia. At USM, he serves as the Coordinator of the Intelligent Biometric Group, where he is the Founder of a biometric product, FaceBARS. Prior to joining the university, he was an Engineer with Sony Video (M) Sdn. Bhd. and Technology Park Malaysia Corporation Sdn. Bhd. for almost six years. His current research interests include face-based biometrics, real-time object detection and tracking, and pattern classification. He has served as a Reviewer for several international conferences and journals, including *IET Biometrics*, *IET Computer Vision*, *Multimedia Tools and Applications*, *Neural Computing and Applications*, the *Journal of Electronic Imaging*, the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, and the IEEE ACCESS.