IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received July 16, 2019, accepted July 26, 2019, date of publication August 7, 2019, date of current version August 20, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2933645

A Superpixel-Wise Approach for Face Sketch Synthesis

ABDULJALIL RADMAN[1,2] AND SHAHREL AZMIN SUANDI[1], (Senior Member, IEEE)
[1]Intelligent Biometric Group, School of Electrical and Electronic Engineering, Universiti Sains Malaysia Engineering Campus, Nibong Tebal 14300, Malaysia
[2]Faculty of Engineering and Information Technology, Taiz University, Taiz, Yemen

Corresponding author: Abduljalil Radman (abdurad@usm.my)

This work was supported by the Universiti Sains Malaysia Research University Individual Research Grant Scheme under Grant 1001/PELECT/8014056.

ABSTRACT Face sketch synthesis from an input photo has drawn great attention in law enforcement and digital entertainment applications. Currently, the input photo is simply divided into overlapped rectangular patches and transformed to a sketch through weighted average of corresponding sketch patches. However, the regular patches lead to defects in the structure of synthesized sketch. In addition, existing methods need to cross the whole training dataset in order to collect the corresponding sketch patches, which limits their usability with the big training datasets. In this paper, a superpixel-wise approach based on the Superpixel technique incorporated into the Locality-constraint Linear Coding (LLC), termed as SuperLLC, is proposed to enhance the facial structure of synthesized sketch and simultaneously maintain fixed computational complexity regardless of the training dataset size. First, the input photo is segmented into overlapped superpixels to find their corresponding sketch superpixels from the training dataset. The LLC is then imposed to regularize proper weights to reconstruct the target sketch from these sketch superpixels, with averaging the overlapped areas between adjacent superpixels. But before all these, for each input photo, a sub-training set is selected based on facial landmarks distance between the input photo and the training photos set. This insures a steady synthesis process. Both subjective and objective experiments on public face sketch databases are finally carried out to reveal the superior performance of the proposed SuperLLC method compared to state-of-the-art methods.

INDEX TERMS Face sketch synthesis, facial landmarks, locality-constraint linear coding (LLC), superpixel segmentation.

I. INTRODUCTION

Recently face sketch synthesis has drawn great efforts due to its wide applications ranging from social networks to law enforcement. It aims to generate a face sketch from a given photo by means of training photo-sketch pairs [1], [2]. Indeed, in numerous law enforcement cases, the culprit's photo is often unavailable because of his deliberately avoidance or limitations in the surveillance camera [3]. In such a situation, a forensic artist is asked to draw a sketch which depicts the culprit facial appearance according to an eyewitness verbal description. Thus, assisting law enforcement entities in expediting the automatic retrieval of a suspect photo from the police mugshot database using a forensic sketch is crucial. This can be done by transforming the photos of the mugshot database into sketches by a face sketch synthesis method, so as to significantly reduce the modality difference between the photo and sketch which is the main barrier in cross-modal face recognition [2]–[4].

To this end, several exemplar-based face sketch synthesis approaches have been introduced in the past two decades [5]–[9]. These approaches have proved its capability in learning the drawing style (e.g. shadow and stroke) without explicitly modeling between face photos and sketches [10]. In such approaches, a test photo is first divided into evenly sized patches with some overlapping in order to guarantee the spatial compatibility. A number of similar photo patches from the training set are then searched for each test photo patch. Next, their corresponding training sketch patches are utilized to synthesize the sketch patch of the test photo patch, with an assumption that the photo and sketch pair in the training set share a similar geometric structure [5], [10], [11].

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

However, the majority of existing exemplar-based face sketch synthesis methods in the literature, either select the best match sketch patch [5], [12], [13] or linearly merge group of selected candidate sketch patches [6], [7], [10], [14]–[16] to synthesize the test sketch patch.

Wang and Tang [5] exploited Markov Random Field (MRF) model with a multi-scale strategy to find dependencies between the test photo patch and the best sketch patch, as well as the compatibility between adjacent synthesized sketch patches so as to obtain a smooth synthesized sketch for the test photo. With the aim of overcoming illumination and pose variations, facial landmarks [17] and Retinex algorithm [18] were integrated into MRF [13]. Although these methods showed promising synthesis results, but these single patch-based techniques might lead to loss in the contents of synthesized sketch patch in view of the limited scale of the training set, especially if the best match patch does not really converge with the desired sketch patch.

To tackle single patch deficiencies, the authors of [6] proposed a Markov weight field model that is able to synthesize the target sketch patch through a linear combination of training sketch patches. In [7], the sketch synthesis problem was formulated as a spatial sketch denoising problem, where each desired sketch patch was constructed using a weighted sum of training sketch patches. Wang *et al.* [19] also proposed a weight-based face sketch synthesis method, taken into account both the test and training samples to find appropriate similar counterparts for each test photo patch. Then, the target sketch patch was generated according to the weights of these counterparts [20]. In fact, these face sketch synthesis techniques divide the face images into regular grid ignoring the inherent structure of the face image, which would introduce either noise or blurring in the synthesized sketch particularly around the edge components (e.g. nose, mouth and chin) [15], [16], [21]. Moreover, the combination weights for each patch is computed independently, which means that the neighboring constraint between adjacent patches is ignored [22]. To overcome these problems, superpixel segmentation and MRF model were applied for face sketch synthesis with the aim of better exploiting the inherent face structure [15]. Despite the observed improvement in the synthesis performance, the computational cost is notably high due to the convergence process of MRF which aims to achieve a visually acceptable state. Focusing on fast synthesis process, Wang et al. [10] proposed a fast-offline random sampled method called Fast-RSLCR, to effectively use random neighbor patches to reconstruct centralized target patch.

Hence, learning-based methods have attracted the researchers to solve such defects. Gao *et al.* [23] applied sparse coding to both the training photos and the training sketches to obtain a pre-synthesized sketch for the test photo. Subsequently, they utilized support vector regression to learn the high frequency relationship between the photo and sketch patch pairs with the aim of compensating the missing details in the pre-synthesized sketch, assuming that the sparse codes in both modalities are same. Likewise, Zhang *et al.* [24] presented a sparse representation method incorporated into Bayesian interference for face sketch synthesis. In the same context, Wang *et al.* [25] proposed a Model-Driven (MD) face sketch synthesis method based on linear regression to learn the mapping between the photo and sketch modalities, intentionally for fast synthesis performance. Sparse representation coefficients were further applied to intelligently minimize the linear reconstruction error between the test photo patch and its sparse neighbors [26], [27]. Most of these methods presumed that both the training sketch patch and its corresponding photo patch have the same sparse representation coefficients, ignoring the common facial structure [20], [22].

Moreover, deep learning networks also played an important role in face sketch synthesis. Zhang *et al.* [28] proposed end-to-end Fully Convolution Network (FCN) for pixel-wise face sketch synthesis. Due to its high-frequency filtering, FCN tends to produce blurry face synthesized sketches [9], [16]. Similarly, Generative Adversarial Network (GAN) was applied to synthesize images which are quite identical to natural images [29]. Since this method is also pixel-to-pixel synthesis method, noisy face sketches are often generated. As a post-processing process, back projection technique had been appended with GAN, termed as BP-GAN, so as to overcome this issue [9]. In spite of fine texture were synthesized by BP-GAN, deformation obviously occurred in the synthesized sketches [8]. In a different approach [8], Zhang et al. introduced a multi-domain adversarial learning method for the purpose of tackling blurs and deformations toward high-quality synthesis.

Nonetheless, except [10], all aforementioned methods consider the whole samples of the training set (i.e., photo and sketch) in order to synthesize every single patch of the target face sketch, which means that the storage and computational complexity are obviously increased with the training data amount. Furthermore, the majority of existing approaches rely on rectangular overlapped patches to find the best training neighbors of the test photo patches. Thereby, the facial components (e.g. chin, mouth, nose and hair borders) are not precisely represented by these rectangular patches, and therefore deformations appear in the synthesized sketches.

In this paper, instead of using the whole training data for synthesizing a face sketch of a test photo, we impose facial landmarks [17] to find some candidate training photos which are properly aligned with the test photo, to be used for synthesizing its corresponding sketch. This strategy not only speeds up the synthesis process but also ensures well face structure. In addition, we segment the test photo into superpixels rather than rectangular patches, and then apply the Locality-constrained Linear Coding (LLC) [30] on their corresponding training sketch superpixels to synthesize the target sketch. This proposed face sketch synthesis method on the basis of Superpixel and LLC is termed as SuperLLC. The remainder of this paper is organized as follows. Section II explains in detail the proposed SuperLLC face sketch synthesis method. Experimental results are presented

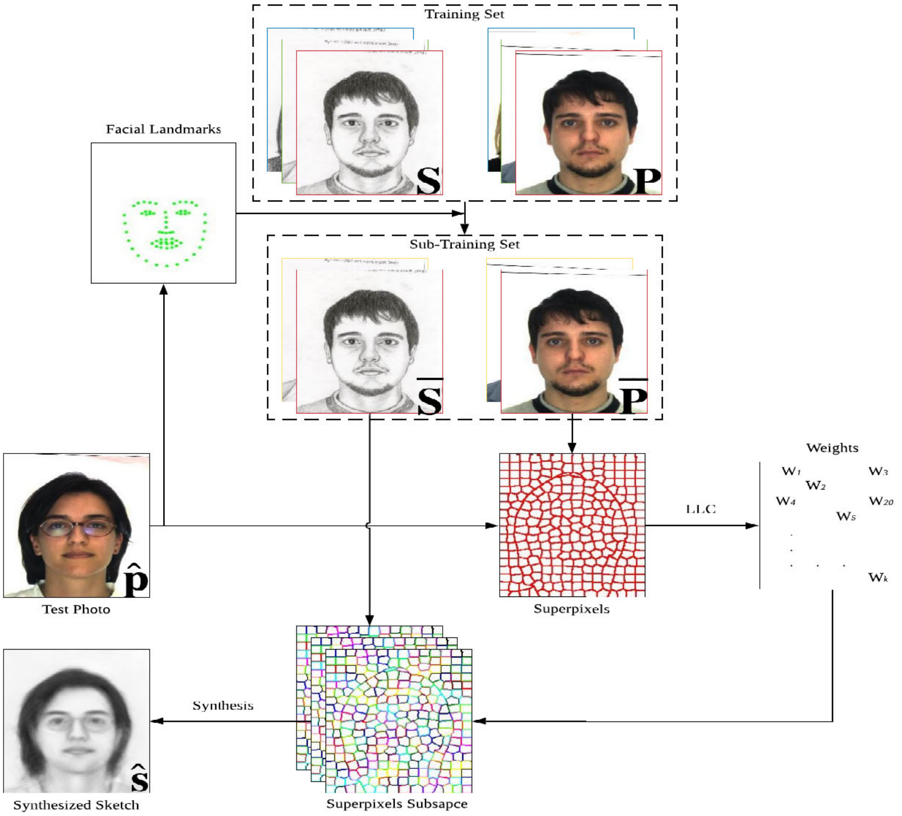


FIGURE 1. **The framework of the proposed SuperLLC face sketch synthesis method.**

and analyzed in Section III, while Section IV concludes the paper.

II. PROPOSED SUPERLLC FACE SKETCH SYNTHESIS METHOD

First, we will expose how facial landmarks are exploited to narrow down the training set for each test photo so as to synthesis its corresponding sketch. Second, we will show how superpixels are constructed to guarantee consistency between the neighbors of synthesized sketch. Finally, we will explain how these superpixels are represented by LLC to achieve the proposed SuperLLC face sketch synthesis method.

In summary, the proposed SuperLLC starts by selecting the most identical training pairs (i.e. subset of photos and their corresponding sketches) to the test photo based on the facial landmarks, to be utilized for the face sketch synthesis. Then, both the test photo and the photos of this sub-training set are segmented into superpixels with overlapping between adjacent superpixels with a view to attain local compatibility between neighbors. Since the superpixels of photos and sketches in the sub-training set form manifolds with similar local geometry, the proposed SuperLLC eventually imposes LLC to synthesize each target sketch superpixel by averaging a combination of training sketch superpixels using the same weights of their corresponding photo superpixels. Fig. 1 graphically illustrates the framework of the proposed SuperLLC face sketch synthesis method.

A. FACIAL LANDMARKS FOR SUB-TRAINING SET SELECTION

For better synthesis production, pairs of training photos and sketches which are geometrically aligned based on predefined points (i.e. eyes and mouth centers) are usually used. Consequently, any difference in pose between the test photo and training set will definitely cause degradation in the synthesis performance. Current face sketch synthesis methods handle this issue by using the whole training set to cover all possibilities of pose variations or by relaxing the searching area for candidates, and thus large computation and storage capabilities are required. To this end, we proposed to apply facial landmarks to adopt the most identical training pairs (i.e. subset of photos and sketches) to the test photo for synthesizing its corresponding sketch. Indeed, the facial landmarks of the training photos can be offline calculated, and then the Euclidean distance between the facial landmarks of the test photo and the training photos is employed to select the sub-training set for the synthesis process. Let $\mathbf{P} = \{\mathbf{p}_i\}_{i=1}^{M}$ and $\mathbf{S} = \{\mathbf{s}_i\}_{i=1}^{M}$ be the photos and sketches training set, respectively, where M is the training set size, $\mathbf{p}_i$ and $\mathbf{s}_i$ are photo and sketch pairs. For each test photo $\hat{\mathbf{p}}$, we used the facial landmarks method [17] to find its sub-training set of M, for

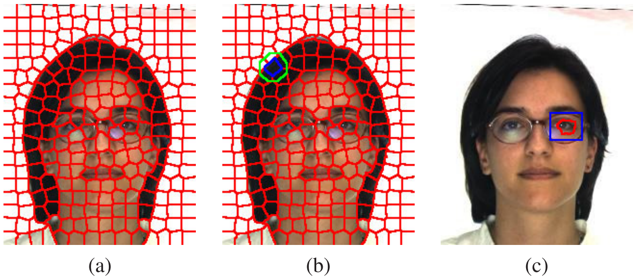


FIGURE 2. Face image with (a) non-overlapped superpixels. (b) Superpixels overlapping (the green contour represents the blue superpixel for overlapping). (c) The search area (blue rectangle) around superpixel (red contour).

synthesizing its corresponding sketch $\hat{\mathbf{s}}$. After obtaining the facial landmarks, we applied Euclidean distance (d_i) between the facial landmarks of the test photo $\hat{\mathbf{p}}$ and each training photo $\mathbf{p}_i$ as follows:

$$d_i = \sqrt{\sum_{j=1}^{c} (FL_{\hat{p}^j} - FL_{p_i^j})^2} \tag{1}$$

where $FL_{\hat{p}^j}$ and $FL_{p_i^j}$ are the facial landmark features for the test photo $\hat{\mathbf{p}}$ and the training photo $\mathbf{p}_i$ at the facial landmark j, respectively, whereas c is the total number of facial landmarks. Accordingly, N training photos with the smallest distances (d_i) and their corresponding training sketches were chosen to construct the sub-training set ($\overline{\mathbf{P}} = \{\overline{\mathbf{p}}_i\}_{i=1}^N$ and $\overline{\mathbf{S}} = \{\overline{\mathbf{s}}_i\}_{i=1}^N$). This procedure contributes to the face sketch synthesis through reducing the computational complexity by decreasing the training set for synthesis ($N \ll M$). Furthermore, it reduces mismatch between the sub-training set (N) and the test photo ($\hat{\mathbf{p}}$), and therefore mitigates the searching area for candidates in the sub-training set.

B. SUPERPIXEL SEGMENTATION AND OVERLAPPING

Given a test photo $\hat{\mathbf{p}}$, to synthesize its corresponding sketch $\hat{\mathbf{s}}$, the main target was to generate photo superpixel space $\mathbf{X}$ of the sub-training photos $\overline{\mathbf{P}}(\mathbf{X} \in \overline{\mathbf{P}})$, corresponding to K superpixels ($\hat{\mathbf{x}}^k$) of $\hat{\mathbf{p}}$ ($\hat{\mathbf{p}} = \{\hat{\mathbf{x}}^k\}_{k=1}^K$). In this paper, the SLIC method [31] was first employed to segment the test photo $\hat{\mathbf{p}}$ into K superpixels, as shown in Fig. 2a. Despite aligning $\hat{\mathbf{p}}$ with $\overline{\mathbf{p}}_i$ based on the facial landmarks (as described in II-A), there would be still misalignment between $\hat{\mathbf{p}}$ and $\overline{\mathbf{p}}_i$ due to the database settings (i.e. all photo and sketch pairs were aligned only on three points, the eyes and mouth centers), which might cause deformations in the synthesized sketch, $\hat{\mathbf{s}}$. Therefore, to ensure local compatibility between adjacent superpixels, all superpixels were then enlarged by a scale of e pixels to confer some ovelapping between adjacent superpixels. Fig. 2b illustrates how superpixles were overlapped (the green contour reveals the overlapping between the blue superpixel and its neighbors). Next, the superpixel space $\mathbf{X}$ were built up within search area $(2r + 1)^2$ around each $\hat{\mathbf{x}}^k$ (the blue rectangle in Fig. 2c), using the same superpixels map of the test photo $\hat{\mathbf{p}}$ (Fig. 2a). Hence, $\mathbf{X} = \{\mathrm{x}^k\}_{k=1}^K$, $\mathrm{x}^k = \{\overline{\mathbf{x}}_n^k\}_{n=1}^{N(2r+1)^2}$ is the subspace corresponding to $\hat{\mathbf{x}}^k$, and r denotes the search length around each $\hat{\mathbf{x}}^k$. To further reduce the computational complexity of the proposed SuperLLC method, only $Q\,\{1 \le Q \le N(2r+1)^2\}$ sketch superpixels y^k corresponding to x^k were utilized to synthesize each sketch superpixel $\hat{\mathbf{y}}^k$ corresponding to $\hat{\mathbf{x}}^k$. The nearest neighbor algorithm [32] was applied to find the Q. Accordingly, $\mathrm{x}^k = \{\overline{\mathbf{x}}_q^k\}_{q=1}^Q$ is the new subspace corresponding to $\hat{\mathbf{x}}^k$. Similarly, $\mathrm{y}^k = \{\overline{\mathbf{y}}_q^k\}_{q=1}^Q$ is the subspace to synthesize $\hat{\mathbf{y}}^k$.

C. SYNTHESIS VIA LOCALITY-CONSTRAINED LINEAR CODING (LLC)

Locality-constrained Linear Coding (LLC) [30] has been widely used in several computer vision applications, such as image classification [33], face hallucination [34]–[36], action recognition [37], and face recognition [38]. Herein, we took the advantage of LLC to represent each superpixel $\hat{\mathbf{y}}^k$ of the target sketch $\hat{\mathbf{s}}$ by a weighted linear combination of its subspace superpixels y^k.

Given prepared photo and sketch superpixel subspaces, x^k and y^k, respectively, which have Q accurate corresponding superpixels ($\overline{\mathbf{x}}_q^k$ and $\overline{\mathbf{y}}_q^k | q \in \{1, \ldots, Q\}$), any $\hat{\mathbf{y}}^k$ can be then reconstructed using the same weights calculated between $\hat{\mathbf{x}}^k$ and x^k. The optimal reconstruction weights are obtained by solving:

$$\min_{\mathbf{w}^k} \|\hat{\mathbf{x}}^k - \mathrm{x}^k \mathbf{w}^k\|_2^2 + \lambda \|\mathbf{d}^k \odot \mathbf{w}^k\|, \\ s.t.\ \mathbf{1}^T \mathbf{w}^k = 1, \quad \forall k \in \{1, \ldots, K\} \tag{2}$$

where $\odot$ denotes to the element-wise multiplication, λ is a free parameter applied to balance the contribution of the reconstruction error term ($\|\hat{\mathbf{x}}^k - \mathrm{x}^k \mathbf{w}^k\|_2^2$) and the regularization term ($\|\mathbf{d}^k \odot \mathbf{w}^k\|$), $\mathbf{d}^k$ is the Euclidean distance vector between $\hat{\mathbf{x}}^k$ and x^k, and $\mathbf{w}^k$ is the weight representation vector of the test photo superpixel $\hat{\mathbf{x}}^k$. Following LLC [30], the solution of Equation 2 can be derived as:

$$\tilde{\mathbf{w}}^k = \left(\mathbf{C}^k + \lambda diag(\mathbf{d}^k)\right) \setminus \mathbf{1}, \tag{3}$$

$$\mathbf{C}^k = (\mathrm{x}^k - 1\hat{\mathbf{x}}^{k^T})(\mathrm{x}^k - \mathbf{1}\hat{\mathbf{x}}^{k^T})^T, \tag{4}$$

$$\mathbf{w}^k = \tilde{\mathbf{w}}^k / 1^T \tilde{\mathbf{w}}^k, \tag{5}$$

where $\mathbf{C}^k$ denotes covariance matrix, $\mathbf{1}$ is a column vector of $1s$, and $diag(\mathbf{d})$ extends the vector $\mathbf{d}$ into a diagonal matrix. The target sketch superpixel $\hat{\mathbf{y}}^k$ is reconstructed from y^k by the same weights $\mathbf{w}^k$:

$$\hat{\mathbf{y}}^k = \mathrm{y}^k \mathbf{w}^k \tag{6}$$

Finally, the target sketch $\hat{\mathbf{s}}$ is then reconstructed from its superpixels $\hat{\mathbf{y}}^k$ with averaging overlapping areas. Algorithm 1 summarizes the proposed SuperLLC face sketch synthesis method.

III. EXPERIMENTAL RESULTS AND ANALYSIS

We subjectively and objectively evaluated the proposed SuperLLC method on the public available databases, the Chi-

(a) (b) (c) (d)

FIGURE 3. Examples of photo-sketch pairs from the CUFS and CUFSF databases. The upper row shows the photos, while their corresponding sketches drawn by the artist are presented in the lower row, from (a) CUHK Student database, (b) AR database, (c) XM2VTS database, and (d) CUFSF database.

Algorithm 1 SuperLLC

Input: The training photo set $\mathbf{P}$, the training sketch set $\mathbf{S}$, the test photo $\hat{\mathbf{p}}$

Step 1: According to facial landmarks [17] and Equation 1, find the sub-training set ($\overline{\mathbf{P}} \subset \mathbf{P}$ and $\overline{\mathbf{S}} \subset \mathbf{S}$) of $\hat{\mathbf{p}}$.

Step 2: Segment $\hat{\mathbf{p}}$ into K superpixels $\hat{\mathbf{p}} = \{\hat{\mathbf{x}}^k\}_{k=1}^{K}$, with overlapping e pixels between the adjacent superpixels.

Step 3: For each test superpixel $\hat{\mathbf{x}}^k$, create its corresponding training photo superpixels subspace $\mathrm{x}^k = \{\overline{\mathbf{x}}_n^k\}_{n=1}^{N(2r+1)^2}$ from $\overline{\mathbf{P}}$, where r denotes the search length around $\hat{\mathbf{x}}^k$ and N is the size of $\overline{\mathbf{P}}$.

Step 4: Reduce the training photo superpixels subspace by the nearest neighbor method [32], $\mathrm{x}^k = \{\overline{\mathbf{x}}_q^k\}_{q=1}^{Q}$ $\{1 \leq Q \leq N(2r+1)^2\}$. Likewise, reduce the corresponding training sketch superpixels subspace $\mathrm{y}^k = \{\overline{\mathbf{y}}_q^k\}_{q=1}^{Q}$, where $\mathrm{y}^k \in \overline{\mathbf{S}}$.

Step 5: Compute the weights $\mathbf{w}^k$ corresponding to $\hat{\mathbf{x}}^k$ as in Equation 5.

Step 6: Reconstruct the target sketch superpixel $\hat{\mathbf{y}}^k$ corresponding to $\hat{\mathbf{x}}^k$ using y^k and $\mathbf{w}^k$ based on Equation 6.

Step 6: Arrange all $\hat{\mathbf{y}}^k$ $\{k \in 1, \ldots, K\}$ into the whole target sketch $\hat{\mathbf{s}}$ taking into account averaging the overlapped areas.

Output: The target sketch $\hat{\mathbf{s}}$

nese University of Hong Kong (CUHK) face sketch (CUFS) database [5] and the CUHK face sketch FERET (CUFSF) database [39]. Fig. 3 shows some examples from these databases. All photos and sketches are cropped to the size of 250×200. The CUFS consists of 606 photo-sketch pairs from three different databases; the CUHK Student (e.g. Fig.3a) database [40], the AR (e.g. Fig.3b) database [41] and the XM2VTS (e.g. Fig.3c) database [42], which include 188 (88 pairs for training and 100 for testing), 123 (80 pairs for training and 43 for testing) and 295 (100 pairs for training and 195 for testing) pairs, respectively. The CUFSF database contains 1194 (250 pairs for training and 944 for testing) photo-sketch pairs from the FERET (e.g. Fig.3d) database [39].

In what follows, we would first introduce the parameters settings for the proposed SuperLLC method. Subsequently, we subjectively (visual quality assessment) compare the performance of SuperLLC with state-of-the-art methods on the CUFS and CUFSF databases. Over and beyond, we further verify the robustness of the proposed SuperLLC method in generating fine structures and textures. Afterward, objective (image quality assessment) evaluation is carried out to demonstrate the superiority of the proposed SuperLLC method compared with state-of-the-arts.

A. PARAMETERS SETTINGS

For the proposed SuperLLC method, there are three main parameters (i.e. the size of sub-training set N (Subsection II-A), the number of superpixels K (Subsection II-B), and the number of nearest neighbor superpixels Q (Subsection II-B)) and three auxiliary parameters (i.e. the overlapping between adjacent superpixels e (Subsection II-B), the length for search area r (Subsection II-B), and the regularization parameter λ (Equation 2)) need to be tuned for better synthesis performance.

Sub-Training Set (N): to overcome the scalability issue of existing face sketch synthesis methods, where the computational complexity is increased with the training set size. We applied facial landmarks distance (Equation 1) to limit the training set size (M) for each test photo $\hat{\mathbf{p}}$ into $N << M$. Since the smallest training set in our experiments is that belong to AR database which is 80 photo-sketch

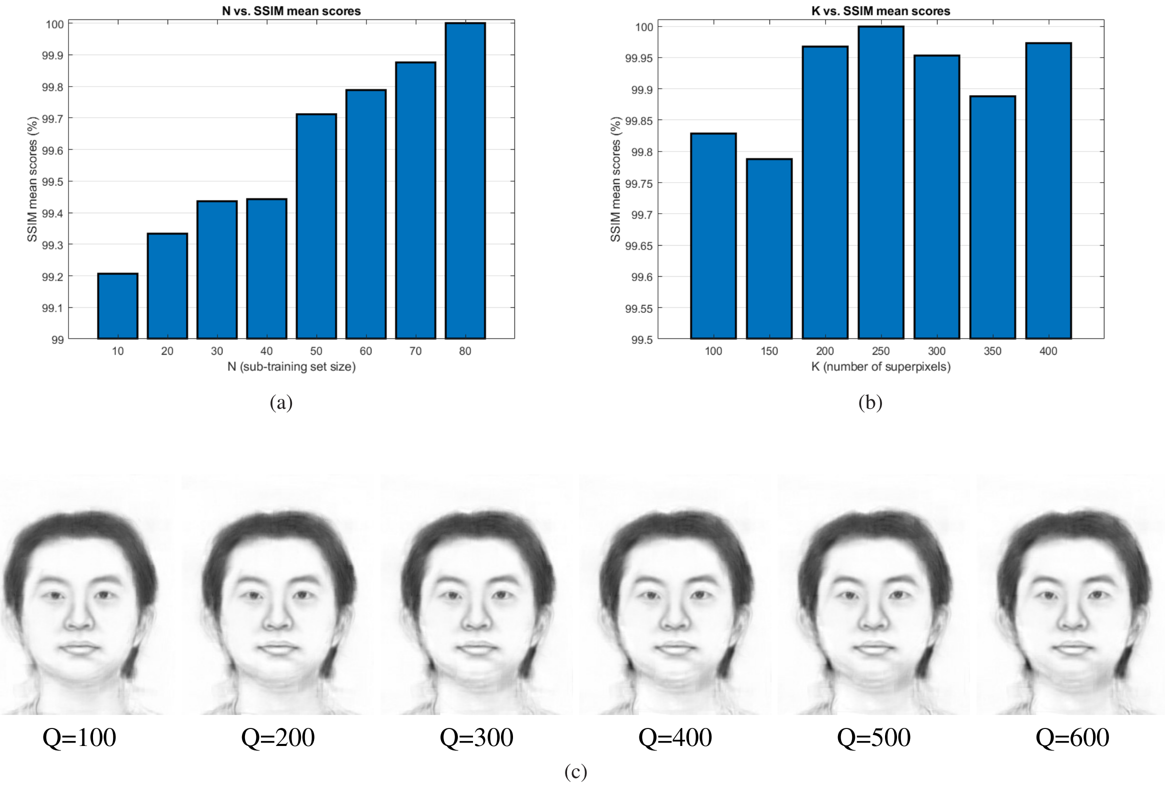


FIGURE 4. **The performance of the proposed SuperLLC method under parameter variations: (a) the sub-training set (N), (b) the number of superpixels (K), and (c) the number of nearest neighbors (Q).**

pairs, we chose N to be less than 80 for all databases. Herein, the CUHK Student database was employed to adjust the parameters, whereas the Structural SIMilarity index (SSIM) [43] was exploited to quantitatively determine the quality of M synthesized sketches (corresponding to 100 test photos) based on the training set (88 photo-sketch pairs). We ran the proposed SuperLLC method with N equals to 10, 20, 30, 40, 50, 60, 70 and 80, respectively. Fig. 4a plots the relationship between the normalized SSIM mean scores and N. It reveals that the quality of synthesized sketches is improved (i.e. higher SSIM) with larger sub-training set N, but definitely the larger N leads to low computation efficiency and storage consuming. However, it can be seen that SSIM mean scores grow slowly after $N = 50$, and thus N was established at 50 in our experiments. The low scale of training set (N) used in the synthesis process not only saves the storage space but also maintains same computational burden across the different sizes of databases. Furthermore, choosing N based on the alignment among the test photo $\hat{\mathbf{p}}$ and the sub-training set photos $\overline{\mathbf{P}} = \{\overline{\mathbf{p}}_i\}_{i=1}^{N}$ reduces the misalignment between $\hat{\mathbf{p}}$ and $\overline{\mathbf{p}}_i$, and thereby allows short length r for the search area $(2r+1)^2$. Hereby, r was set to 5 pixels.

Number of Superpixels (K): we tested the proposed SuperLLC method with 100, 150, 200, 250, 300, 350 and 400 superpixels, respectively. From Fig. 4b, it can be noticed that SSIM mean scores is obviously degraded after $K = 250$. This is due to the small superpixels (larger K) that cause mosaic effects. It can be also seen that $K = 400$ performs almost similar to $K = 250$, however, larger K means more superpixels ($\hat{\mathbf{y}}^k$) need to be synthesized and hence much computational complexity, so K was fixed at 250 in subsequent experiments. By means of that, $K(= 250)$ is segmented the test photo $\hat{\mathbf{p}}$ (250×200 pixels) into approximately 200 pixels for each superpixel, and therefore 11 pixels was assigned to e with a view to assure at least 50% overlapping between contiguous superpixels.

Number of Nearest Neighbors (Q): to further improve the computation efficiency and synthesized sketch quality, we restricted the photo superpixels subspace (x^k) corresponding to each test photo superpixel ($\hat{\mathbf{x}}^k$) to include only Q superpixels ($\mathrm{x}^k = \{\overline{\mathbf{x}}_q^k\}_{q=1}^{Q}$), as detailed in Subsection II-B. The effect of different numbers of Q on the proposed SuperLLC performance was presented in Fig.3c. It is clearly visible that the larger Q leads to deformation and blur around the face structure (e.g. chin edges as in Fig.3c ($Q = 200$ to $Q = 600$)). Moreover, Q is mainly affected the computational complexity of the proposed SuperLLC method as it is clear from Equation (6), where y^k depends on the number of Q. Hence, we retained $Q = 100$ for the proposed SuperLLC method. From our empirical study, the regularization parameter λ was given the value of 0.5, which can achieve well in calculating the weights (Equations 3 and 5) for reconstructing $\hat{\mathbf{y}}^k$ according to Equation (6).

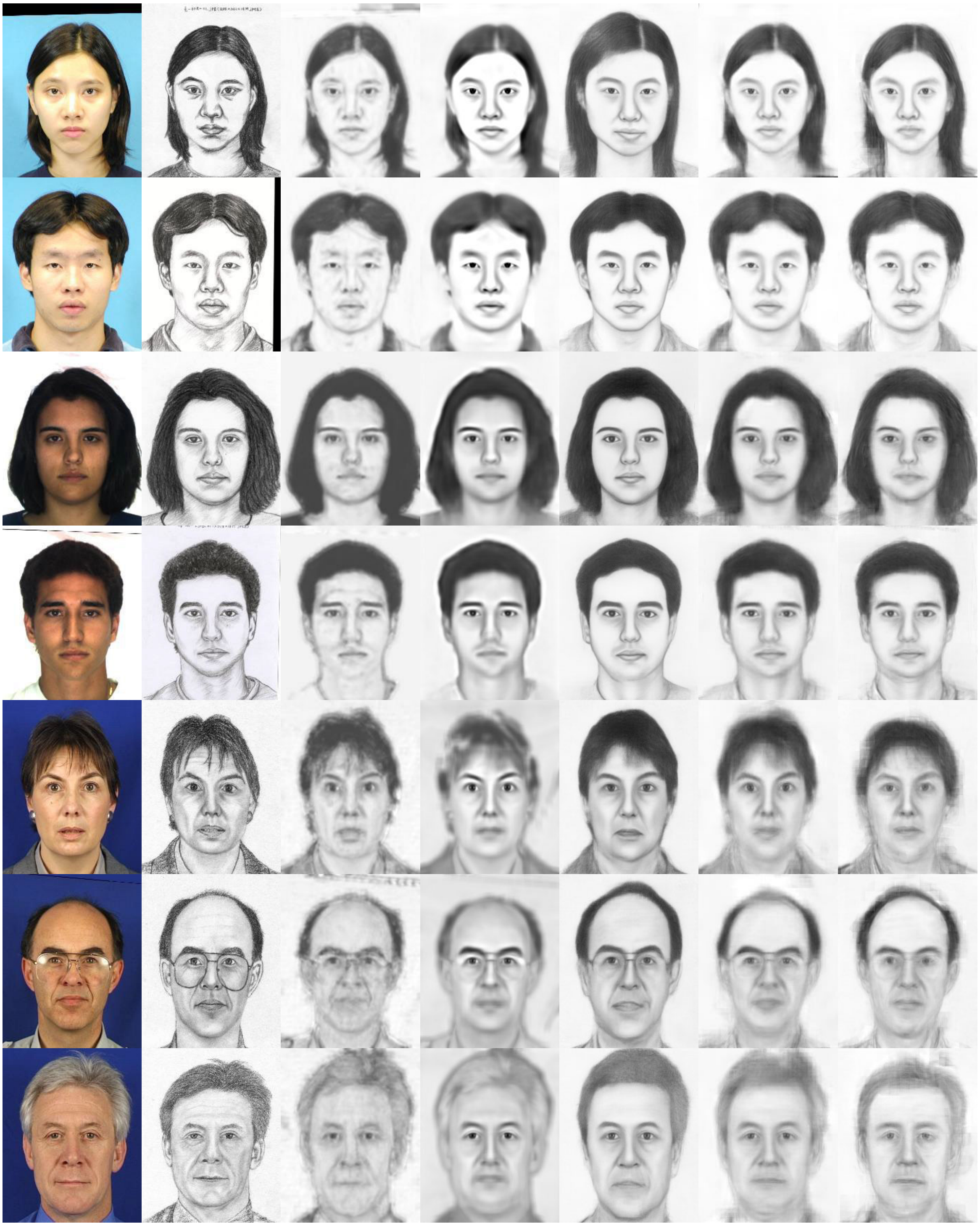


FIGURE 5. Some synthesized face sketches on the CUFS database by FCN [28], MD [25], BP-GAN [9], Fast-RSLCR [10], and the proposed SuperLLC method. The first four rows are based on photos from the CUHK Student (1^{st} and 2^{nd}) and AR (3^{rd} and 4^{th}) databases, while the last three rows are based on photos from the XM2VTS database.

B. SUBJECTIVE PERFORMANCE EVALUATION

In this section, the quality of synthesized face sketches obtained by the proposed SuperLLC method[1] is subjectively evaluated and compared to state-of-the-arts. Fig. 5 and Fig. 6 present some synthesized face sketches obtained by the proposed SuperLLC, FCN [28], MD [25], BP-GAN [9] and Fast-RSLCR [10] methods, on the CUFS database and the CUFSF database, respectively. From Fig. 5 and Fig. 6, it can

[1]The synthesized sketches of the proposed SuperLLC method are available online at http://ee.eng.usm.my/IBG/index.php/ms/downloads.

FIGURE 6. **Some synthesized face sketches on the CUFSF database by FCN [28], MD [25], BP-GAN [9], Fast-RSLCR [10], and the proposed SuperLLC method.**

be observed that FCN generates noisy sketches (especially on the CUFSF database), this is due to the fact that FCN method tries to synthesize the fine details of the face. Likewise, the synthesized sketches by MD method have fine textures but the noise effect still exists as depicted by Fig. 6 (i.e. on the CUFSF database). This is because of MD method is

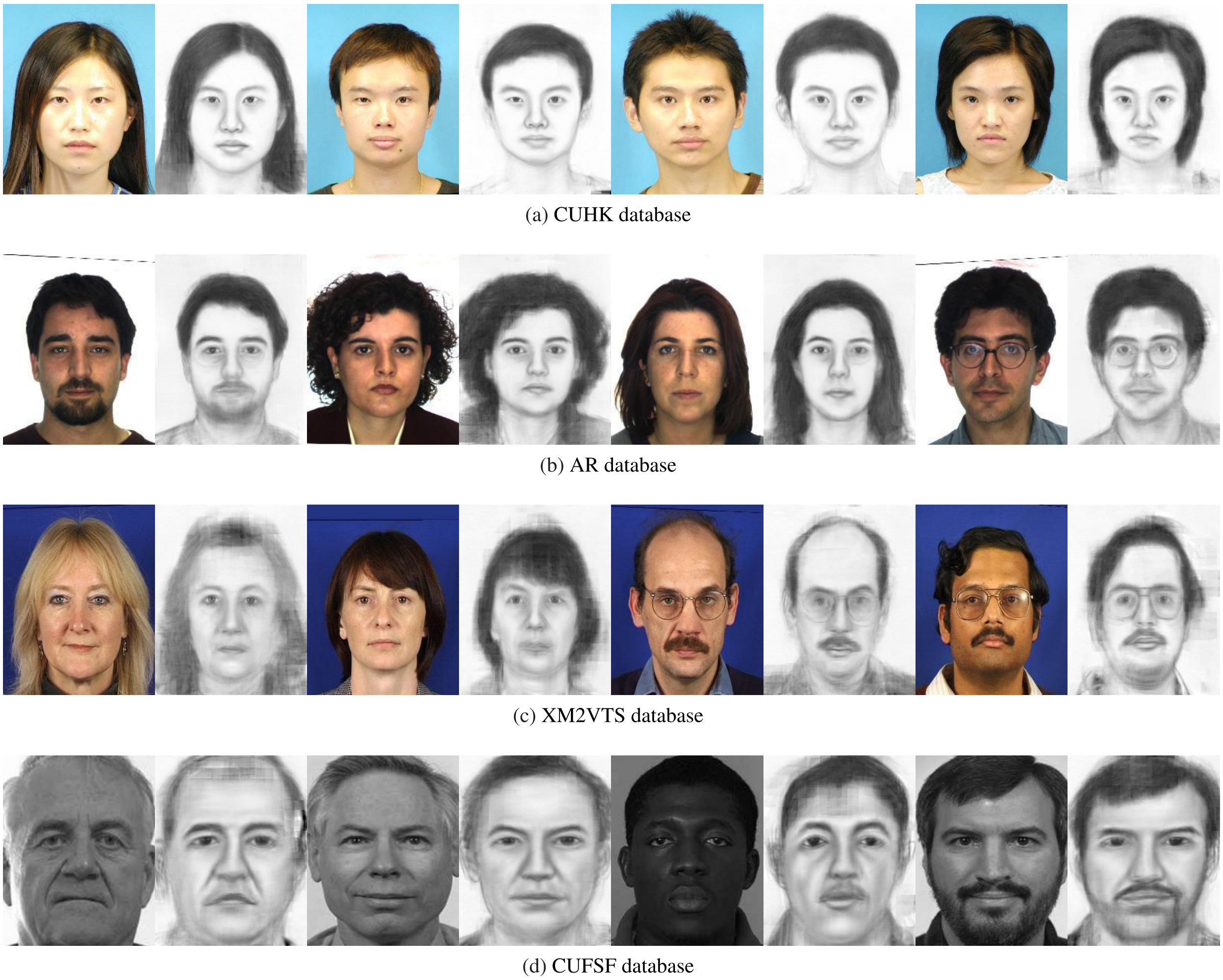
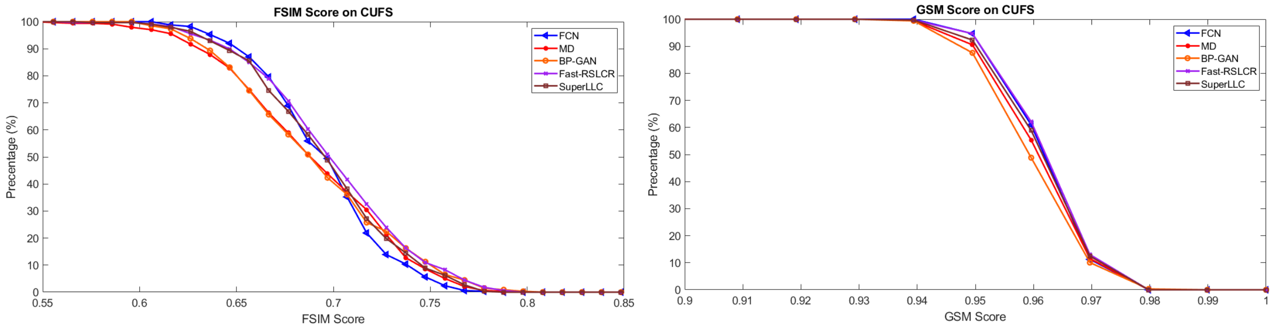


FIGURE 7. Examples of synthesized sketches by the proposed SuperLLC method on the CUFS database (a, b and c) and the CUFSF database (d).

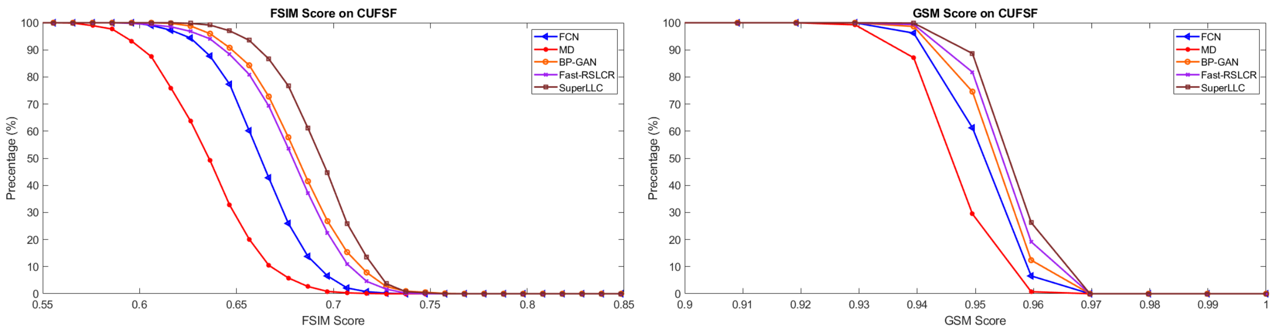
a linear method while the photos and sketches of the CUFSF database have many nonlinear aspects such as shape exaggeration, illumination variations, and skin colors. Even though MD method shows better synthesized sketches on the CUFS database than those ones by FCN method, but those synthesized by FCN method on the CUFSF are much better than of MD method. In overall, MD method performs the worst on the CUFSF database. On the contrary to FCN and MD methods, BP-GAN method presents smooth textures on the synthesized sketches. This because the main goal of BP-GAN was to remove the noise caused by GAN [29] which is a pixel-to-pixel synthesis method. However, the synthesized sketches by BP-GAN method are considerably blurred, this effect comes from exaggerating the back projection process around the high frequency regions. The synthesized sketches illustrated in Fig. 5 and Fig. 6 show that Fast-RSLCR method generates smooth appearance and less noise on both the CUFS and CUFSF databases, but on the other hand some blur around the face structure is noticed (e.g. on the mouth, eyes and chin). This is due to the huge number of samples used to reconstruct each patch of the face (i.e. 200 samples). In contrast to all previous methods, the sketches synthesized by the proposed SuperLLC method, on both the CUFS and CUFSF databases, are well structured, much clearer in appearance, less noisy and blurry. Indeed, the synthesized sketches by the proposed SuperLLC method present perfect structures for the face components such as the mouth, nose, eyes and chin. This is due to the superpixel technique, used in the proposed SuperLLC method, preserves the face structure better than the regular rectangular patches which have been used in the other methods (e.g. Fast-RSLCR). The nearest neighbor algorithm [32] and alignments among the test photo $\hat{\mathbf{p}}$ and sub-training set photos $\overline{\mathbf{P}} = \{\overline{\mathbf{p}}_i\}_{i=1}^{N}$ based on facial landmarks [17] also significantly contribute in finding the best appropriate superpixels $\mathrm{x}^k = \{\overline{\mathbf{x}}_q^k\}_{q=1}^{Q}$ for each test superpixel $\hat{\mathbf{x}}^k$. This illustrates the superiority of the proposed SuperLLC method over state-of-the-art methods.

More synthesized face sketches generated by the proposed SuperLLC method on the CUFS and CUFSF databases are presented in Fig. 7. These synthesized sketches clearly



(a) (b)

FIGURE 8. Objective quality assessment on synthesized sketches of the CUFS database.



(a) (b)

FIGURE 9. Objective quality assessment on synthesized sketches of the CUFSF database.

demonstrate that the proposed SuperLLC method could generate fine edges around the face wrinkles (e.g. the 2^{nd} and 4^{th} columns of CUFSF database (Fig.7d)), and well structures for eyeglasses (e.g. the 8^{th} column of AR database (Fig.7b), 6^{th} and 8^{th} columns of XM2VTS database (Fig.7c)). Furthermore, the proposed SuperLLC method performs well with blond hair (e.g. the 2^{nd} column of XM2VTS database (Fig.7c)) and black skin (e.g. the 6^{th} column of CUFSF database (Fig.7d)). All these prove the effectiveness of the proposed SuperLLC method.

C. OBJECTIVE PERFORMANCE EVALUATION

Since the essential target for the proposed SuperLLC method is to enhance the facial structure of synthesized sketches, we utilized the Feature SIMilarity (FSIM) index [44] and the Gradient Similarity Metric (GSM) [45], which are mainly developed to measure the changes in structure between the test and reference images, to objectively evaluate the synthesized sketches generated by the proposed SuperLLC method. FSIM [44] uses the phase congruency and the gradient magnitude for full reference image quality assessment, while GSM [45] considers both luminance and contrast structural changes to effectively assess the image quality. For this purpose, the sketches drawn by artists corresponding to the synthesized sketches were utilized as reference images. Fig. 8 and Fig. 9 illustrate the statistical curve of FSIM and GSM scores for the CUFS and CUFSF databases, respectively. The horizontal axis labels the FSIM/GSM scores and the vertical axis labels the percentage of synthesized sketches whose scores are larger than the score marked on the horizontal axis. Table 1 also presents the average FSIM/GSM scores on the synthesized sketches generated by different face sketch synthesis methods, on the CUFS and CUFSF databases. Even though BP-GAN method illustrated smooth and clear synthesized sketches in Fig. 5, FSIM and GSM curves in Fig. 8 as well as the average scores in Table 1 led to opposite conclusion, where BP-GAN method presents the lowest score on the CUFS database. Similar to what have been discussed in Subsection III-B, MD showed the worst performance on the CUFSF database as clearly reflected by FSIM and GSM curves in Fig.9, as well as the average scores of FSIM and GSM in Table 1. Fig.8 presented slight superiority from FCN and Fast-RSLCR methods over the proposed SuperLLC method, but the average of FSIM scores in Table 1 showed that the proposed SuperLLC method achieves better performance than FCN method. The superiority of Fast-RSLCR method on the CUFS database was because that Fast-RSLCR method exploits all the color channels to retrieve the best match from the training set, whereas the proposed SuperLLC method only relied on the grayscale channel to retrieve the corresponding superpixels from the sub-training set. This also proves the inconsistency of Fast-RSLCR method, where it

TABLE 1. Average FSIM/GSM scores on sketches synthesized by different methods.

Method	CUFS		CUFSF	
	FSIM	GSM	FSIM	GSM
FCN [28]	69.35	96.11	66.24	95.09
MD [25]	68.83	96.04	63.53	94.62
BP-GAN [9]	68.98	95.92	68.14	95.30
Fast-RSLCR [10]	69.79	96.17	67.75	95.47
SuperLLC	69.44	96.07	69.23	95.62

operated in full color space with the CUFS database and grayscale channel with the CUFSF database. In contrast to the CUFS database, the proposed SuperLLC method significantly outperforms all other methods on the CUFSF database, as depicted in Fig.9 and Table 1, which coincides with our conclusions in Subsection III-B.

IV. CONCLUSION

This paper proposes a superpixel-wise approach, called SuperLLC, for face sketch synthesis from an input photo. In contrast to existing approaches which divide the input photo into regular patches, the proposed SuperLLC method starts by segmenting the input photo into superpixels in order to assure well face structure. The proposed SuperLLC method not only generates fine structures and textures on the synthesized sketches, but also maintains fixed computational complexity through selecting limited numbers (i.e. the sub-training set) of photos from the training set, to be used in synthesizing the target sketch of the input photo. The proposed SuperLLC method ensures overlapping among adjacent superpixels to take into account the local compatibility issue. Then, the Locality-constraint Linear Coding (LLC) is exploited to find the reconstruction weights between the input photo superpixels and similar photo superpixels of the sub-training set, to be used for reconstructing the corresponding target sketch superpixels. Subjective and objective evaluations demonstrated the superiority of the proposed SuperLLC method on state-of-the-art face sketch synthesis methods. The proposed SuperLLC method also illustrated sketches with much textures and structures details. Furthermore, the synthesized sketches significantly showed the robustness of the proposed SuperLLC method to skin and hair variations. In the future, the proposed SuperLLC method would be extended to synthesis a sketch from a single training photo-sketch pair, exploiting facial landmarks or similar features to select the best training photo-sketch pair for each input photo.

REFERENCES

[1] 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.

[2] 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.

[3] A. Radman and S. A. Suandi, "Robust face pseudo-sketch synthesis and recognition using morphological-arithmetic operations and HOG-PCA," *Multimedia Tools Appl.*, vol. 79, no. 19, pp. 25311–25332, Oct. 2018.

[4] J. Jiang, Y. Yu, Z. Wang, and J. Ma, "Residual learning for face sketch synthesis," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 1952–1956.

[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] H. Zhou, Z. Kuang, and K.-Y. K. Wong, "Markov weight fields for face sketch synthesis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 1091–1097.

[7] Y. Song, L. Bao, Q. Yang, and M.-H. Yang, "Real-time exemplar-based face sketch synthesis," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 800–813.

[8] S. Zhang, R. Ji, J. Hu, X. Lu, and X. Li, "Face sketch synthesis by multidomain adversarial learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 5, pp. 1419–1428, May 2019.

[9] N. Wang, W. Zha, J. Li, and X. Gao, "Back projection: An effective postprocessing method for GAN-based face sketch synthesis," *Pattern Recognit. Lett.*, vol. 107, pp. 59–65, May 2018.

[10] N. Wang, X. Gao, and J. Li, "Random sampling for fast face sketch synthesis," *Pattern Recognit.*, vol. 76, pp. 215–227, Apr. 2018.

[11] C. Chen, X. Tan, and K.-Y. K. Wong, "Face sketch synthesis with style transfer using pyramid column feature," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, Mar. 2018, pp. 485–493.

[12] W. Zhang, X. Wang, and X. Tang, "Lighting and pose robust face sketch synthesis," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 420–433.

[13] A. Radman and S. A. Suandi, "Markov random fields and facial landmarks for handling uncontrolled images of face sketch synthesis," *Pattern Anal. Appl.*, vol. 22, no. 1, pp. 259–271, Feb. 2019.

[14] 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.

[15] C. Peng, X. Gao, 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.

[16] J. Jiang, Y. Yu, Z. Wang, X. Liu, and J. Ma, "Graph-regularized locality-constrained joint dictionary and residual learning for face sketch synthesis," *IEEE Trans. Image Process.*, vol. 28, no. 2, pp. 628–641, Feb. 2019.

[17] X. Yu, J. Huang, S. Zhang, W. Yan, and D. N. Metaxas, "Pose-free facial landmark fitting via optimized part mixtures and cascaded deformable shape model," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1944–1951.

[18] E. H. Land, "Recent advances in retinex theory and some implications for cortical computations: Color vision and the natural image," *Proc. Nat. Acad. Sci. USA*, vol. 80, no. 16, p. 5163, Aug. 1983.

[19] 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.

[20] M. Zhang, R. Wang, X. Gao, J. Li, and D. Tao, "Dual-transfer face sketch–photo synthesis," *IEEE Trans. Image Process.*, vol. 28, no. 2, pp. 642–657, Feb. 2019.

[21] M. Zhang, J. Li, W. Wang, and X. Gao, "Compositional model-based sketch generator in facial entertainment," *IEEE Trans. Cybern.*, vol. 48, no. 3, pp. 904–915, Mar. 2018.

[22] 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.

[23] 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.

[24] S. Zhang, X. Gao, N. Wang, and X. Li, "Face sketch synthesis from a single photo–sketch pair," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 2, pp. 275–287, Feb. 2017.

[25] 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.

[26] S. Zhang, X. Gao, N. Wang, and J. Li, "Robust face sketch style synthesis," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 220–232, Jan. 2016.

[27] S. Zhang, X. Gao, N. Wang, J. Li, and M. Zhang, "Face sketch synthesis via sparse representation-based greedy search," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2466–2477, Aug. 2015.

[28] 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.*, Jun. 2015, pp. 627–634.

[29] P. Isola, J.-Y. Zhu, T. Zhou, and A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. 30th IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 5967–5976.
[30] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 3360–3367.
[31] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.
[32] J. L. Bentley, J. H. Friedman, and R. A. Finkel, "An algorithm for finding best matches in logarithmic time," *ACM Trans. Math. Softw.*, vol. 3, pp. 209–226, Sep. 1977.
[33] J. Chen, Q. Li, Q. Peng, and K. H. Wong, "CSIFT based locality-constrained linear coding for image classification," *Pattern Anal. Appl.*, vol. 18, no. 2, pp. 441–450, May 2015.
[34] L. Liu, C. L. P. Chen, S. Li, Y. Y. Tang, and L. Chen, "Robust face hallucination via locality-constrained bi-layer representation," *IEEE Trans. Cybern.*, vol. 48, no. 4, pp. 1189–1201, Apr. 2018.
[35] J. Jiang, R. Hu, Z. Han, T. Lu, and K. Huang, "Position-patch based face Hallucination via locality-constrained representation," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2012, pp. 212–217.
[36] Y. Li, J. Liu, W. Yang, and Z. Guo, "Multi-pose face Hallucination via neighbor embedding for facial components," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2015, pp. 4037–4041.
[37] H. Rahmani, D. Q. Huynh, A. Mahmood, and A. Mian, "Discriminative human action classification using locality-constrained linear coding," *Pattern Recognit. Lett.*, vol. 72, pp. 62–71, Mar. 2016.
[38] X. Peng, L. Zhang, Z. Yi, and K. K. Tan, "Learning locality-constrained collaborative representation for robust face recognition," *Pattern Recognit.*, vol. 47, no. 9, pp. 2794–2806, Sep. 2014.
[39] W. Zhang, X. Wang, and X. Tang, "Coupled information-theoretic encoding for face photo-sketch recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 513–520.
[40] X. Tang and X. Wang, "Face photo recognition using sketch," in *Proc. Int. Conf. Image Process.*, Sep. 2002, vol. 1, p. 1.
[41] A. M. Martinez, "The ar face database," CVC, New Delhi, India, Tech. Rep. 24, 1998.
[42] 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*, Mar. 1999, vol. 964, pp. 965–966.
[43] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[44] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
[45] A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1500–1512, Apr. 2012.

ABDULJALIL RADMAN received the B.Eng. degree in electrical engineering from Sana'a University, Yemen, in 2002, and the M.Eng. and Ph.D. degrees in electrical, electronic, and systems engineering from the Universiti Kebangsaan Malaysia, Malaysia, in 2009 and 2014, respectively. From 2002 to 2007, he was an Assistant Lecturer with the Department of Electrical Engineering, Sana'a University, and the Department of Communication and Computer Engineering, Taiz University. He was a Postdoctoral Research Fellow of Universiti Sains Malaysia and Universiti Sains Islam Malaysia, from 2015 to 2017. From January 2018 to November 2018, he was an Assistant Professor with the Department of Communication and Computer Engineering, Taiz University, Yemen. He is currently a Postdoctoral Research Fellow of the Intelligent Biometric Group, School of Electrical and Electronic Engineering, Universiti Sains Malaysia, Malaysia. His research interests include image processing, computer vision, and pattern recognition.

SHAHREL AZMIN SUANDI received the B.Eng. degree in electronic engineering, 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 and the Deputy Dean of the Electrical and Electronic Engineering School, Universiti Sains Malaysia Engineering Campus (USM), Malaysia. 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 well-known international journals such *IET Computer Vision*, *IET Biometrics*, the SPIE *Journal of Electronic Imaging*, *Applied Soft Computing*, *Multimedia Tools and Applications*, and others. He has served as a Reviewer for well-known international conferences such as WACV and AVSS.