Face Sketch–Photo Synthesis and Retrieval Using Sparse Representation

Xinbo Gao, *Senior Member, IEEE,* Nannan Wang, Dacheng Tao, *Senior Member, IEEE,* and Xuelong Li, *Fellow, IEEE*

***Abstract*—Sketch–photo synthesis plays an important role in sketch-based face photo retrieval and photo-based face sketch retrieval systems. In this paper, we propose an automatic sketch–photo synthesis and retrieval algorithm based on sparse representation. The proposed sketch–photo synthesis method works at patch level and is composed of two steps: sparse neighbor selection (SNS) for an initial estimate of the pseudoimage (pseudosketch or pseudophoto) and sparse-representation-based enhancement (SRE) for further improving the quality of the synthesized image. SNS can find closely related neighbors adaptively and then generate an initial estimate for the pseudoimage. In SRE, a coupled sparse representation model is first constructed to learn the mapping between sketch patches and photo patches, and a patch-derivative-based sparse representation method is subsequently applied to enhance the quality of the synthesized photos and sketches. Finally, four retrieval modes, namely, sketch-based, photo-based, pseudosketch-based, and pseudophoto-based retrieval are proposed, and a retrieval algorithm is developed by using sparse representation. Extensive experimental results illustrate the effectiveness of the proposed face sketch–photo synthesis and retrieval algorithms.**

***Index Terms*—Face sketch–photo synthesis, sketch–photo-based retrieval, sparse neighbor selection, sparse representation.**

I. Introduction

ACROSS THE wide range of imaging modes, face retrieval is not limited to face photo retrieval. Example-based face database retrieval is mainly achieved by two means: photo-based face retrieval and sketch-based face retrieval. Photo-based face retrieval is widely applied in many fields, such as access control systems, Internet search engines, and video surveillance. However, in many cases such as case-solving and suspect-searching, a photo of the suspect cannot be obtained. An alternative method is to use a substitute sketch generated through the cooperation of an artist with a witness.

Manuscript received August 28, 2011; revised November 30, 2011; accepted January 10, 2012. Date of publication May 7, 2012; date of current version July 31, 2012. This work was supported by the National Basic Research Program of China (973 Program), under Grant 2012CB316400, by the National Natural Science Foundation of China, under Grants 61125204, 61125106, 91120302, 61172146, and 60832005, by the Fundamental Research Funds for the Central Universities, by the Natural Science Basis Research Plan, Shaanxi, China, under Grant 2009JM8004, by the open Project Program of the State Key Laboratory of CAD&CG at Zhejiang University, under Grant A1006, by Zhejiang University, and by the State Administration of STIND, under Grant B1320110042. This paper was recommended by Associate Editor A. Signoroni.

X. Gao and N. Wang are with the VIPS Laboratory, School of Electronic Engineering, Xidian University, Xi'an 710071, China (e-mail: xbgao@mail.xidian.edu.cn; nannanwang.xidian@gmail.com).

D. Tao is with the Center for Quantum Computation and Intelligent Systems, Faculty of Engineering and Information Technology, University of Technology, Sydney, NSW 2007, Australia (e-mail: dacheng.tao@uts.edu.au).

X. Li is with the Center for Optical Imagery Analysis and Learning, State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China (e-mail: xuelong_li@opt.ac.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2012.2198090

Sketch-based face retrieval can mainly be used in two situations: one is to identify a criminal, and the other is to determine whether there exist committed crimes, or further, how many times a criminal has broken the law. In the former situation, face retrieval can be performed on a mug shot image database with a face sketch as a query image. In the latter case, a face photo can be used as a query image for retrieval from a gallery of mug shot images. However, sketches and photos are heterogeneous and can be significantly different in their modality, both in geometry and in texture. Consequently, generic photo-photo face retrieval algorithms [1]–[6] are greatly challenged in performing heterogeneous image recognition. To reduce the differences between sketches and photos as much as possible, an advisable method is to convert them to the same representation space. Therefore, heterogeneous image transformation between sketches and photos for this purpose becomes a crucial task. Furthermore, face sketch synthesis can also be applied to some video-based industries, such as cartoon and animation. An illustration of face sketch synthesis from four frames extracted from video shots [44] is shown in Fig. 1.

There are two strategies for transforming heterogeneous face images into homogeneous ones, one being from a sketch to a photo and the other from a photo to a sketch. In this paper, face images include face photos and sketches, so "pseudoimage" may mean "pseudosketch" or "pseudophoto." For the sake of brevity, we use different notations to represent these two kinds of transformation. "Sketch$\rightarrow$photo" denotes the transformation from a sketch to a photo and a pseudophoto is generated. Correspondingly, "photo$\rightarrow$sketch" means the transformation from a photo to a sketch, from which a pseudosketch results. "Sketch–photo" presents sketch–photo pairs or the transformation between a sketch and a photo, which can be easily understood in context. Moreover, "example-based image retrieval" is referred to sketch-, photo-, pseudosketch, or pseudophoto-based retrieval.

A. Related Work

The research on face sketch–photo synthesis is still at an initial stage. In one of our previous works [7], sketches are classified into two major classes: 1) the simple line-drawing sketch and caricature; and 2) the complex sketch studied in this paper. Several algorithms have been proposed [8]–[12] to transform a photo to a line-drawing sketch or a caricature. Wang *et al.* proposed a sketch generation algorithm by imitating a pencil in [13]. Simple line-drawing sketches and caricatures may be easily recognized by human beings; however, this is a rather hard task for computers. Thus, it is difficult to perform face retrieval using the above two types of sketches. Recently, Klar *et al.* used forensic sketches (drawn by an artist according to the descriptions of a witness rather than from a static photo, as adopted in this paper) to do the matching in mug shot photos, with the help of multiscale local binary patterns and SIFT feature descriptors [14]. No reference is made in that paper of sketch–photo synthesis, however, only image retrieval.

Recent achievements in sketch–photo synthesis and retrieval can be classified into two categories: linear algorithms [15]–[17], [20], [22] and nonlinear algorithms [7], [18], [19], [24]–[27]. Tang *et al.* achieved noticeable progress [15]–[21] in recent years by applying the idea of principle component analysis to the photo→sketch transformation in [15]–[17], in which a query photo is first projected onto the photo training set. A pseudosketch is then synthesized from the sketch training set with the projection coefficients obtained above. Li *et al.* extended Tang's idea into a hybrid subspace and assumed that sketches and photos share the same projection vector [22]. Liu *et al.* proposed an algorithm [20] to hallucinate a photo from a sketch based on Bayesian tensor inference. Because the mapping between sketches and photos is not a simple linear mapping, these linear algorithms cannot obtain satisfying results, especially when hair regions are included.

Using the idea of locally linear embedding (LLE) [23], Liu *et al.* proposed an approach that transformed a photo into a pseudosketch by a nonlinear transformation [18] at the patch level. Given that sketches and photos are on the same manifold, they first found K nearest neighbors of a photo patch in the photo space, with which they can reconstruct the photo patch and thus obtain the reconstruction coefficient vector. Thereafter, a pseudosketch patch results with the obtained coefficient vector and the sketch patches corresponding to the K nearest photo patch neighbors. Finally, a pseudosketch is fused with these pseudosketch patches. Nevertheless, since the number of nearest neighbors is fixed, some of the nearest neighbors may be not closely correlated to the query patch, leading to a blurring effect. Wang and Tang proposed a new sketch–photo synthesis system using the multiscale Markov random field [19], investigating both sketch→photo and photo→sketch transformations. This method can generate a much better result compared with the foregoing algorithms. Unfortunately, in some cases, there are small deformations in the synthesized images. Zhang *et al.* proposed a lighting and pose robust face sketch synthesis algorithm based on the multiscale Markov random field [21]. In our previous works, we proposed several sketch–photo synthesis algorithms based on embedded hidden Markov model (EHMM) [7] and [24]–[27], which can be incorporated into the company of nonlinear methods. EHMM is first constructed to learn the nonlinear mapping between sketches (sketch patches) and photos (photo patches). By combining the selective ensemble strategy [28] with several learned models, a final pseudosketch is generated. Although these algorithms achieve good results on a small database, they need do many forward-backward and Viterbi iterations. Most of the existing algorithms to date are k-nearest neighbors (KNN)-based for neighbor selection, resulting in a low definition.

Since sketches and photos are represented in the same image space, generic face recognition algorithms [1], [2] can be applied to conduct face image retrieval. Recently, inspired by the idea of sparse representation, Wright *et al.* proposed a robust face recognition algorithm via sparse representation in [29], which consists of a simple model but can achieve a better result under well-controlled circumstances. Yang *et al.* also proposed several effective face recognition algorithms based on sparse representation [30]–[32]. Yuan and Yan presented a visual classification with a multitask joint covariate selection model [33], which can be efficiently optimized via an accelerated proximal gradient method. Considering the fact that sketches and photos are taken under well-controlled conditions (the faces to be studied in this paper are in a frontal pose with normal lighting, neutral expression, and no occlusions) and given the simplicity of the model, we introduce the sparse representation classification (SRC) [29] algorithm to verify the effectiveness of the proposed sparse neighbor selection-sparse representation-based enhancement (SNS-SRE) method. In addition, we conduct face sketch–photo retrieval on four different modes using the sparse-representation-based face recognition algorithm.

B. Proposed Method

Analyzing existing sketch–photo synthesis approaches, we find that there is a disadvantage that may handicap the face image retrieval process. The number of nearest neighbors in [18], [19], [21], and [24]–[27] is fixed, which results in the pseudoimages synthesized by existing methods being low in definition. In this paper, a new sketch–photo synthesis method, SNS-SRE, is proposed based on sparse representation to mitigate this adverse effect. Taking the pseudosketch synthesis as an example, the framework of the proposed algorithm is demonstrated in Fig. 2. The pseudophoto synthesis process can be obtained by switching the role of sketches and photos. It should be noticed that the proposed sketch–photo synthesis algorithm is patch-level-based. For consistency of the global face structure, each image in the database (both the sketch–photo training set and sketch–photo testing set) is divided into even areas in a raster scan order with overlap between adjacent patches.

We can see from Fig. 2 that SNS-SRE consists of two parts: sparse neighbor selection (SNS) obtains an initial estimate, and sparse-representation-based enhancement (SRE) enhances the definition of the initial estimate. Assuming that the sketch patches and photo patches lie in the same manifold (the difference between SNS and the idea in [23] is that the number of nearest neighbors is not fixed, but adaptively determined in this paper), a sketch patch shares the same sparse representation vector with the corresponding photo patch. Here, the sparse representation vector is obtained by projecting the

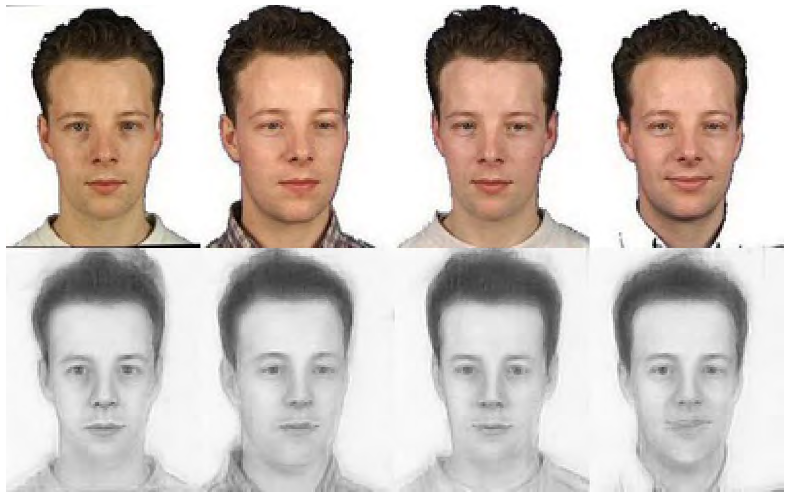


Fig. 1. Face sketch synthesis on four video frames using the proposed method. Top row is the frames and below is the corresponding synthesized sketches.

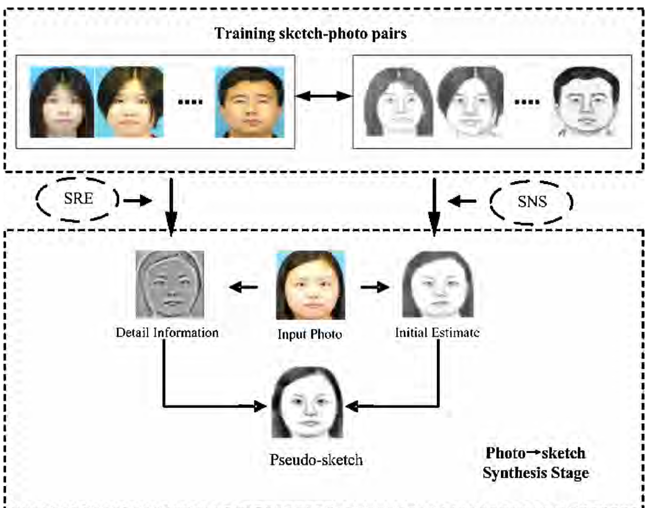


Fig. 2. Framework of the proposed SNS-SRE image synthesis algorithm.

sketch (photo) patch on a predefined sketch (photo) patch dictionary. The predefined sketch (photo) patch dictionary is constructed by a number of sketch (photo) patches randomly sampled from the sketch (photo) training set. Given an input photo patch, a pseudosketch patch can be obtained by a linear combination of the sketch patches whose corresponding coefficients in the sparse representation vector are nonzero, where the linear combination coefficient is the nonzero sparse representation coefficient normalized by the sum of elements in the sparse representation vector. To further enhance the definition of the pseudoimages, we propose a patch-derivative-based strategy via sparse representation inspired by the idea of [34] and [35]. Since the mean value of an image patch approximates the low frequency information of this patch and the derivatives correspond to the high frequency or detailed information, the proposed SRE strategy is adopted on the high frequency information working at patch level. Let us consider the situation of pseudosketch synthesis, which is no longer mentioned in the following context. A pseudophoto can be obtained easily by switching the roles of sketches and photos. In the training stage, the mean value of each sketch patch is subtracted to obtain the detailed information of this patch; the first-order derivative and the second-order derivative of each photo patch are extracted as its feature vector. In this model, we obtain two dictionaries through joint-training, meaning that these two dictionaries are learned but not predefined, as in SNS. A similar strategy to that being employed in SNS is then applied to the feature of the input patch. A synthesized derivative-level feature result is added to the corresponding patch in the initial estimate. This results in an enhanced pseudosketch patch. Fusing these enhanced patches, we shall obtain a pseudosketch with higher definition than existing methods.

The remainder of this paper is organized as follows. Sketch–photo synthesis and enhancement are introduced in Section II. Example-based face image retrieval via sparse representation is presented in Section III. Experimental results and performance evaluation are given in Section IV, and Section V concludes this paper.

II. Face Photo→Sketch Synthesis and Enhancement

As the first step for face database retrieval, sketch–photo synthesis bears a significant effect on the retrieval performance. In this section, we will introduce the face sketch synthesis method that can be easily extended to face photo synthesis by switching the roles of sketches and photos.

Sparse representation has been successfully applied in image super-resolution [34], [35], image denoising [39], and image restoration [40]. It admits a signal sparse approximation over an overcomplete dictionary. Let $D \in \mathbb{R}^{u \times N_{\text{atom}}}$ be an overcomplete dictionary consisting of N_{atom} atoms, where u is the dimensionality of each atom and $N_{\text{atom}} >> u$. A signal $y \in \mathbb{R}^u$ can be denoted as a sparse linear combination of these atoms, i.e., $y = Dw_0$, with $w_0 \in \mathbb{R}^{N_{\text{atom}}}$ being a vector of few nonzeros ($<< N_{\text{atom}}$) entries. This can be formulated by

$$w_0 = \arg\min_{w} \|w\|_0 \quad \text{s.t. } y = Dw \tag{1}$$

where $\|w\|_0$ counts the number of nonzero elements in w. According to some recent research [41], the optimization problem in (1) can be transformed into the following form to solve:

$$w_0 = \arg\min_{w} \|w\|_1 \quad \text{s.t. } y = Dw. \tag{2}$$

Here, the sparse-representation-based model was used to obtain an initial estimate and to enhance the details of the initial estimate.

A. SNS

In previous works [7], [18], [19], [21], and [24]–[27], the methods of K nearest neighbors are extensively used to synthesize a pseudosketch or a pseudophoto. However, it is not guaranteed that all these fixed K neighbors are the first K most related patches to the query patch. In contrast, sparse representation is capable of eliminating this problem to some extent by finding the fewest closely related patches to reconstruct the query patch.

Given a query photo P_i ($i = 1, \ldots N_q$, N_q is the number of query photos), it is first divided into n even patches with some overlap, i.e., $P_i = \{P_{i1}, \ldots, P_{in}\}$, where P_{ij} means the jth patch of photo i. In this paper, each patch is ordered lexicographically as a column vector. Let C_p denote a dictionary whose columns consist of patches sampled from the photo training set and C_s indicate the dictionary with the sketch patches corresponding to photo patches in C_p as its column vectors. Then, the sparse representation of P_{ij} can be represented as follows:

$$\min \left\|w_{ij}\right\|_1 \ \text{s.t. } \left\|C_p w_{ij} - P_{ij}\right\|_2^2 \leq \varepsilon \tag{3}$$

where $w_{ij} = (w_{ij1}, w_{ij2}, \cdots w_{ijc})^T$ indicates the coefficient vector and c is the number of columns in C_p or C_s. We can solve the above optimization problem for w_{ij}. Afterward, the neighborhood of photo patch P_{ij} is achieved according to the following criteria:

$$N(P_{ij}) = \left\{k | \delta(w_{ijk}) \neq 0, 1 \leq k \leq c\right\} \tag{4}$$

where $N(P_{ij})$ denotes the neighborhood for P_{ij} and $\delta(\cdot)$ is a neighbor selection function

$$\delta(w_{ijk}) = \begin{cases} w_{ijk}, & \left|w_{ijk}\right| \geq \sigma \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where σ is a small positive real number, which is set to 0.001 in our experiments. Once the neighborhood is determined, the weight for every neighbor in $N(P_{ij})$ is calculated as follows:

$$W_{ijk} = \delta(w_{ijk}) / sum(\delta(w_{ij})) \tag{6}$$

where $sum(\delta(w_{ij}))$ denotes the summation of $\delta(w_{ijk}), 1 \leq k \leq c$. Finally, the corresponding initial estimate sketch patch can be synthesized as

$$S_{ij} = C_s W_{ij}. \tag{7}$$

For each photo patch in the query photo, we iterate the above steps and then fuse all these synthesized patches into a whole sketch by averaging the overlapping areas.

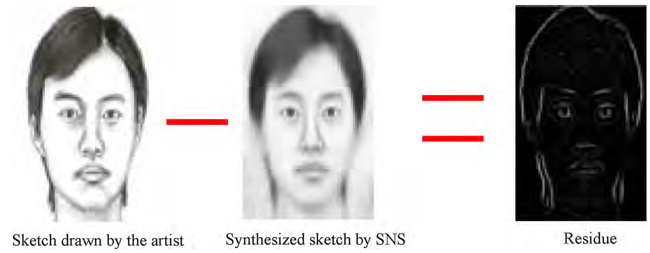


Fig. 3. Motivation of sparse-representation-based enhancement.

B. SRE

In most existing methods [7], [15]–[18], [20], [24]–[27], the target synthesized image (patch) was acquired by a linear combination of candidate images (patches). Indeed, this would result in some over-smooth or blurring effect due to the fact that linear combination can be seemed to be a low-pass filter to some extent (actually linear combination is a weighted average and average is thought to be a low-pass filter). Then, some high-frequency information may be filtered by this process, which can be shown in Fig. 3.

Using the joint-training strategy to be described in Section II-C, we can obtain two normalized dictionaries: a sketch patch feature dictionary D_p and a photo patch feature dictionary D_s. After the extraction of the feature vector of photo patch P_{ij}, a sparse representation of this patch can be achieved. The sparse representation problem can be expressed, as presented in Section II-A, by

$$\min \ \|w\|_1 \ \text{s.t. } \left\|D_p w - \mathbb{F} P_{ij}\right\|_2^2 \leq \varepsilon \tag{8}$$

where $\mathbb{F}$ denotes a feature extracting operator. If $\mathbb{F}$ is an identity operator, $\mathbb{F} P_{ij}$ means a vector of patch intensities as described in Section II-A. Formula (8) can be rewritten as follows with a Lagrangian multiplier:

$$\min \ \lambda \|w\|_1 + \left\|D_p w - \mathbb{F} P_{ij}\right\|_2^2 \tag{9}$$

where λ is the penalty factor for sparsity and is set to 0.1 in all our experiments. Formula (9) is a linear regression problem about the sparse representation coefficient vector w. Indeed, the above optimization problem in (3) is a special case of (9). Thus, they can be solved by the same approach such as [38] and [42].

The first-order and second-order derivatives can be used as the features for the photo patches. For the sketch, we subtract the mean intensity value for each patch so as to be in the same order of magnitude with derivatives. The four filters used to extract the derivatives are as follows:

$$\begin{array}{ll} d_1 = [-1, 0, 1] & d_2 = f_1^T \\ d_3 = [1, 0, -2, 0, 1] & d_4 = f_3^T. \end{array} \tag{10}$$

For each photo patch, four feature vectors are concatenated to one vector as the final feature representation for this photo patch. Assuming a photo or a sketch is divided into patches of size $s \times s$, the feature vector for a sketch patch has a size of s^2 and thus the feature vector for a photo patch has a size of $4s^2$. One can also take the mean intensity value subtracted by its mean as the input feature for the photo patches, which

Algorithm 1 Algorithm for photo→sketch enhancement

Input: a query photo P_i, sketch patch feature dictionary D_S, photo patch feature dictionary D_p, and the penalty factor λ for sparsity.

1) Divide P_i into n patches, for each patch P_{ij}, extract the feature vector and solve the following formula for the sparse representation coefficient vector w:

$$\min\ \lambda \|w\|_1 + \left\|D_p w - f_{P_{ij}}\right\|_2^2.$$

2) Reconstruct the feature vector of the patch $f_{S_{ij}}$: $f_{S_{ij}} = D_s w$.
3) Fuse all feature vectors of pseudosketch patches $f_{S_{ij}}(j = 1, \ldots n)$ into one feature vector. The overlapping regions are averaged.

Output: the pseudosketch $\tilde{S}_i$.

was adopted in our experiments since this strategy has small inputting feature dimension and stable performance. After the feature vector for sketches and photos is ready, a coupled dictionary can be trained using the algorithm shown in Tabel III. We still use $\tilde{S}_i$ to denote the pseudosketch obtained from the first step. Given a query photo P_i, for each patch P_{ij} of P_i, a feature vector $f_{P_{ij}}$ of size $4s^2$ is obtained from image intensity values by subtracting the mean intensity value of P_{ij}. Using the photo patch dictionary D_p, the sparse representation of this feature vector can be found according to the following formula:

$$\min\ \lambda \|w\|_1 + \left\|D_p w - f_{P_{ij}}\right\|_2^2. \tag{11}$$

Once w is obtained, a feature vector $f_{S_{ij}}$ of the pseudosketch patch can be reconstructed as

$$f_{\tilde{S}_{ij}} = D_s w. \tag{12}$$

The algorithm for the photo→sketch enhancement method is shown in Algorithm 1. Just by exchanging several variables (the input dictionaries, the feature extracting operator), we can obtain the solution to the optimization problem in (3).

Finally, $f_{\tilde{S}_{ij}}$ is added to the initial estimate patch S_{ij} to obtain the final enhanced pseudosketch. The whole algorithm for the enhancement of face pseudosketch synthesis is shown in Algorithm 2.

C. Joint-Training for Dictionary

As [34] and [35] did, we adopt the joint-training strategy to learn two dictionaries: the sketch patch feature dictionary and the photo patch feature dictionary. We first randomly sample 80 000 patches in the training set. Assuming the sampled image patches to be $\{T_P, S_P\}$, $T_P = \{T_{P11}, \ldots T_{P1M}, \ldots T_{PL1}, \ldots T_{PLM}\}$ is the sampled photo patches and $T_S = \{T_{S11}, \ldots T_{S1M}, \ldots T_{SL1}, \ldots T_{SLM}\}$ is the corresponding sketch patches, where M denotes the number of sampled patches in each image and N the number of training sketch–photo pairs. Our goal is to train a photo patch feature dictionary and the corresponding sketch patch feature dictionary. We assume that a sketch patch and the corresponding photo patch have the same sparse representation. Under this

Algorithm 2 Whole algorithm for photo→sketch synthesis

Input: A query photo P_i, photo training set T_p, sketch training set T_S, the penalty factor λ for sparsity.

Part 1: Original pseudosketch synthesis stage
Using the SNS method in Section II-A with the sketch–photo training pairs to obtain an initial estimate.
Output: an initial estimate with low definition.

Part 2: Enhancement stage

1) Learn the dictionary pair:
   - Step 1. Sample the feature vector for T_p and T_S;
   - Step 2. Use the algorithm in Algorithm 3 with the feature vector to learn the sketch patch feature dictionary and the photo patch feature dictionary.
2) Synthesize the enhanced part:
   For $I = 1$:***n*,** n is the number of patches.
   - Step 1. For the ith query photo patch, extract the feature vector;
   - Step 2. Use the algorithm in Algorithm 1 to find the sparse representation of the ith feature vector according to the following formula:

     $$\min\ \lambda \|w\|_1 + \left\|D_p w - f_{P_{ij}}\right\|_2^2$$

   - Step 3. Synthesize the i-th feature vector according to the formula $f_{S_{ij}} = Dw_{ij}$ and add it to the corresponding initial estimate patch.
     End
   - Step 4. Fuse these n patches into the final pseudosketch.

Output: an enhanced pseudosketch.

consideration, we have the following constrained objective function to optimize:

$$D_p = \arg\min_{\{D_p, X\}} \left\|T_P - D_p X\right\|_2^2 + \alpha \|X\|_1 \tag{13}$$

$$D_s = \arg\min_{\{D_s, X\}} \|T_S - D_s X\|_2^2 + \alpha \|X\|_1 \tag{14}$$

where X is the sparse representation matrix, ℓ^1-norm $\|X\|_1$ is for sparsity, and ℓ^2-norm is used to guarantee the fidelity of the approximation. We can combine (13) and (14) into a formula

$$\begin{aligned}\min_{\{D_p, D_s, X\}} \frac{1}{N_d} \left\|T_P - D_p X\right\|_2^2 + \frac{1}{M_d} \|T_S - D_s X\|_2^2 \\ +\alpha \left(\frac{1}{N_d} + \frac{1}{M_d}\right) \|X\|_1\end{aligned} \tag{15}$$

where N_d and M_d are the dimensionality of the photo patch and sketch patch, respectively, expressed in vectors. Here, $1/N_d$ and $1/M_d$ are used to balance (13) and (14). If $N_d = M_d$, $1/N_d$ and $1/M_d$ can be omitted. Further, we can rewrite (15) as follows:

$$\min_{\{D_p, D_s, X\}} \|I - DX\|_2^2 + \beta \|X\|_1 \tag{16}$$

where $I = \left[T_P^T/\sqrt{N_d}, T_S^T/\sqrt{M_d}\right]^T$, $D = [D_p^T/\sqrt{N_d}, D_s^T/\sqrt{M_d}]^T$, and $\beta = \alpha(1/N_d + 1/M_d)$, which is set to 0.05 in all

Algorithm 3 Algorithm for joint-training for dictionary

Input: Initialize the matrix D as a Gaussian matrix and normalize each atom to a unit vector; and the penalty factor β for sparsity.

1) Fix D, update C with the following formula:

$$C = \arg\min_{C} \|I - DC\|_2^2 + \beta \|C\|_1 .$$

2) Fix C, update D with as follows:

$$D = \arg\min_{D} \|I - DC\|_2^2 , \quad \text{s.t.} \|D_i\|_2^2 \leq 1, i = 1, 2, \ldots K.$$

The formula is a quadratic programming with square constraint and
there are a lot of algorithms for solving it efficiently.

3) Iterate 1) and 2) until convergent is achieved.

Output: the dictionary matrix D.

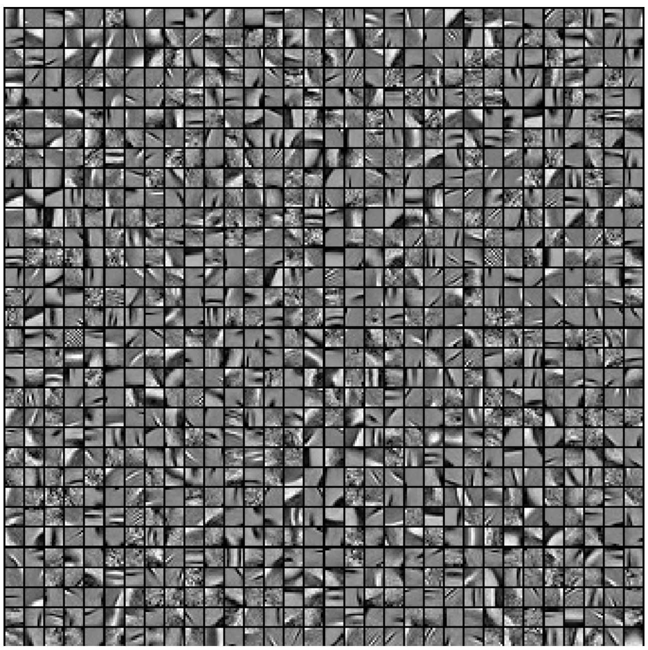


Fig. 4. Photo patch feature dictionary trained by (16) using 80 000 sketch–photo patch pairs sampled from our newly constructed VIPSL database. There are 1024 dictionary atoms in total with each atom of size 9×9.

our experiments. To make the solution unique, the dictionary matrix D should be normalized to unit 2-norm. In our joint-training process, we can also extract sketch or photo patch features as the sampling data. We can see that (16) is not convex for D and X. However, when either D or X is fixed, (16) is convex for X or D. So we can adopt the alternative iteration method to optimize (16), and the corresponding algorithm is shown in Algorithm 3. A learned dictionary with (16) is shown in Fig. 4.

III. Example-Based Face Database Retrieval

Combined with the applications in case-solving and suspect-searching, example-based face database retrieval includes two situations: sketch-based face photo database retrieval and photo-based face sketch database retrieval. For the former,

Algorithm 4 Sparse representation-based retrieval algorithm

Input: The matrix of training samples $Z = [Z_1, Z_2, \ldots Z_c] \in \mathbb{R}^{m\times t}$ for c classes, a test sample $x \in \mathbb{R}^m$, and the penalty factor for sparsity λ

Step 1. Solve the ℓ^2-minimization problem

$$\hat{\gamma}_1 = \arg\min_{\gamma} \|\gamma\|_1 \quad \text{subject to } \|Z\gamma - x\|_2 \leq \varepsilon.$$

Step 2. Compute the residuals $r_i(x) = \left\| x - Z\delta_i(\hat{\gamma}_1) \right\|_2$, $i = 1, \ldots c_l$ where $\delta_i : \mathbb{R}^t \rightarrow \mathbb{R}^t$ is the characteristic function that selects the coefficients associated with the ith class. For $x \in \mathbb{R}^t$, $\delta_i(x) \in \mathbb{R}^t$ is a new vector whose only nonzero entries are the entries in x that are associated with class i.

Output: identity$(x) = \arg \min_i r_i(x)$.

TABLE I
Different Face Retrieval Modes

Query Image	Face Database (Training Set)
Pseudophoto	Photo database
Sketch	Pseudosketch database
Pseudosketch	Sketch database
Photo	Pseudophoto database

TABLE II
Subjective Quality Assessment Results-MOS Value

Pseudoimages	[18]	[24]	[25]	SNS-SRE
CUHK-pseudosketch	3.4	3.6	–	4.35
CUHK-pseudophoto	2.7	–	3.75	4.65
VIPSL-pseudosketch	1.9	2.8	–	4.2
VIPSL-pseudophoto	3.05	–	3.65	3.9

we can first transform the query sketch into a pseudophoto and later retrieve the associated candidates in the face photo database. An alternative method is to transform the face photo database into a face pseudosketch database and then retrieve it in the face pseudosketch database with the query sketch. For the latter, the means of retrieval are similar to the first situation. A query photo can either be transformed into a pseudosketch or the face sketch database can be transformed into a face pseudophoto database. Thus, there are four retrieval modes in all, as shown in Table I.

Inspired by the idea of sparse representation, Wright *et al.* proposed a robust face recognition algorithm via sparse representation in [29]. Instead of learning a feature dictionary by using generic sparse representation algorithms, they took the training samples directly as atoms of the dictionary. A test sample can then be projected on the dictionary to obtain the sparse representation coefficient vector via ℓ^1-minimization. The identity of the test sample is determined by these coefficients combining with the training samples with minimal estimator. In this paper, we combine the sparse-representation-based face retrieval method with example-based images (sketches, photos, pseudosketches, and pseudophotos) into an example-based face database retrieval algorithm.

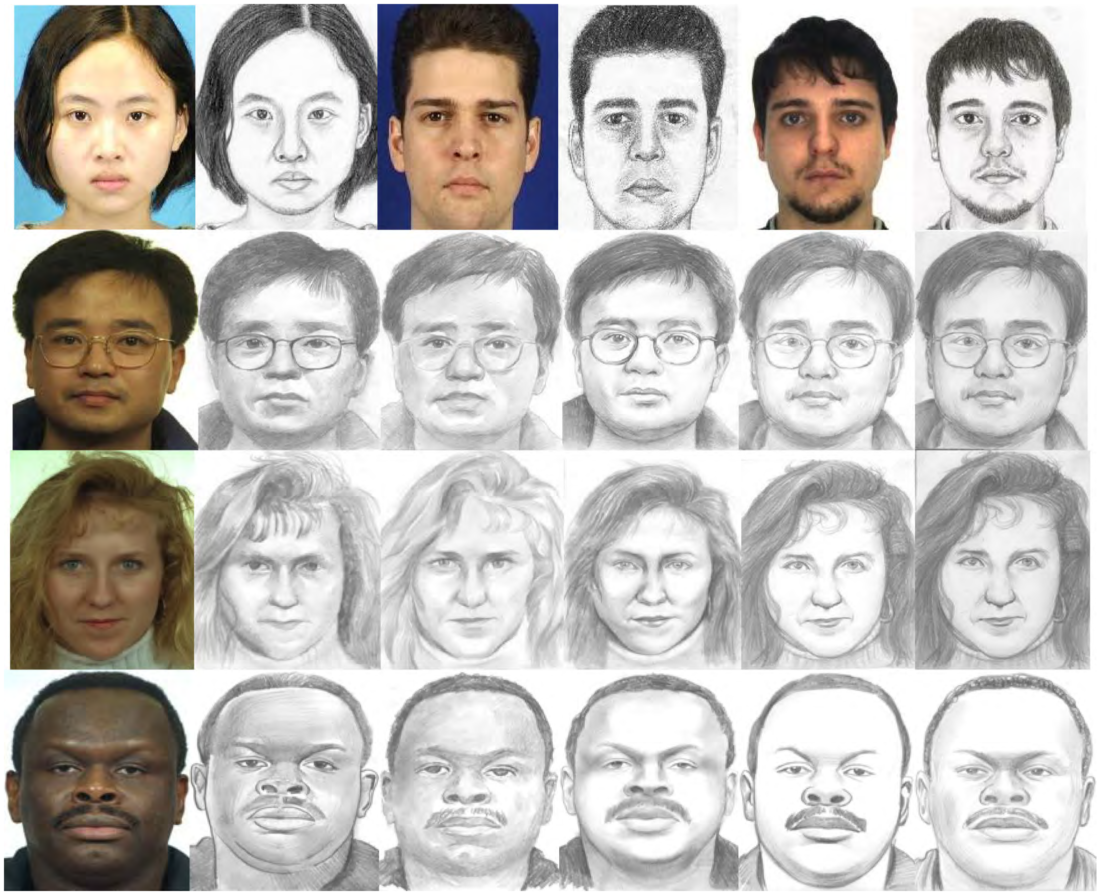


Fig. 5. Examples used in experiments. Images of the top row come from CUFS database, images of the last three rows come from VIPSL database.

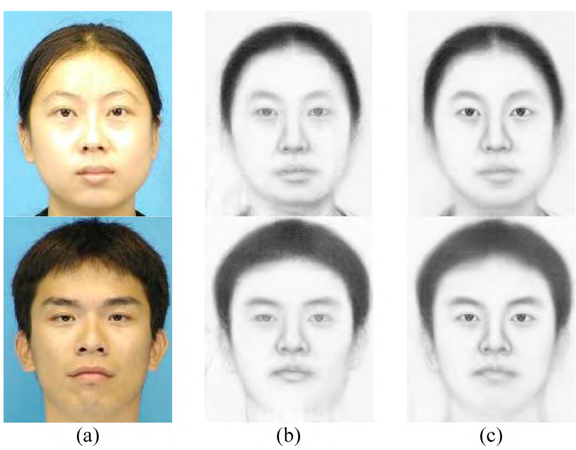


Fig. 6. Comparisons between SNS and KNN-based methods. (a) Original photos. (b) Sketches synthesized by [18]. (c) Sketches synthesized by the proposed SNS method.

Before conducting the retrieval, sketches and photos are transformed into homogeneous images using the algorithm in Algorithm 2. In this section, we do not distinguish sketches, photos, pseudosketches, and pseudophotos, and thus give them a common name-image. In order to distinguish the difference from the dictionary above, here we use the capital letter Z as a substitute. For a given image of size $w \times h$, it is arranged as a column vector $v \in \mathbb{R}^m (m = wh)$ in a lexicographical order. Supposing $Z = [Z_1, Z_2, \ldots Z_{c_l}] \in \mathbb{R}^{m\times t}$, where t is the total number of training samples and c_l is the number of classes, the subset $Z_i = [v_{i1}, v_{i2}, \ldots v_{it_i}] \in \mathbb{R}^{m\times t_i}$ $(i = 1, \ldots c_l)$ consists of t_i training samples from the ith class. Each column of Z is normalized to have a unit ℓ^2-norm. In [29], Wright *et al.* suggested that a test sample $x \in \mathbb{R}^m$ can be approximately written as a linear combination of the training samples from the same class, that is

$$x = \gamma_{i1} v_{i1} + \gamma_{i2} v_{i2} + \ldots \gamma_{it_i} v_{it_i}. \tag{17}$$

Taking the whole training matrix Z into account, (17) can be rewritten as follows:

$$x = Z\gamma \tag{18}$$

where $\gamma = [0, \ldots 0, \gamma_{i1}, \gamma_{i2}, \ldots \gamma_{it_i}, 0, \ldots 0]^T \in \mathbb{R}^t$ is a coefficient vector whose entries are zero except those associated with the ith class. Then, the identity information is shown in the coefficients. The sparse-representation-based retrieval algorithm is shown in Algorithm 4.

IV. Experiments and Analysis

This section conducts experiments on two face sketch–photo databases to illustrate the effectiveness of the proposed algorithm. One database is a public database:[1] CUFS, released by the Multimedia Laboratory, Chinese University of Hong

[1]Available at http://mmlab.ie.cuhk.edu.hk/facesketch.html.

Fig. 7. Pseudosketch synthesis results on the CUFS database. (a) Original photos. (b) Pseudosketches generated by [18]. (c) Pseudosketches generated by SRE-LLE method. (d) Pseudosketches generated by [24]. (e) Pseudosketches generated by SRE-EHMM method. (f) Pseudosketches generated by the proposed SNS-SRE method.

TABLE III
SRC VERSUS GENERIC RECOGNITION METHODS WITH SKETCHES AS THE TRAINING SET (%)

	[18]	[24]	SNS-SRE
Eigenface	77.6	91.9	83.8
Fisherface	63.7	91.7	73.6
LPP	33.2	61.7	43.7
SRC	91.4	92.2	**93.1**

TABLE IV
SRC VERSUS GENERIC RECOGNITION METHODS WITH PSEUDOSKETCHES AS THE TRAINING SET (%)

	[18]	[24]	SNS-SRE
Eigenface	85.4	87.6	87.7
Fisherface	87.1	89.3	91.6
LPP	78.9	82.2	83.9
SRC	91.3	92.1	**93.7**

Kong, Shatin, Hong Kong. The other is newly constructed by our Video and Image Processing System Laboratory (and is called the VIPSL database). The former includes 606 face sketch–photo pairs from three subdatabases, which consists of 188, 295, 123 sketch–photo pairs, respectively [19], [43], [44]. Our newly constructed VIPSL database contains 200 face photos in total, selected from the FERET database [45], [46], the FRAV2D database [47], and the Indian Face Database.[2] For each face, there are five sketches drawn by five different artists based on the same photo taken in frontal pose, under normal lighting conditions, and with neutral expression. Fig. 5 shows some examples, in which all face sketches and photos are cropped and normalized both in geometric and grayscale. The size of each image in our experiments is unified to 163×200.

[2]Available at http://vis-www.cs.umass.edu/∼vidit/IndianFaceDatabase/, 2002.

A. Sketch–Photo Synthesis

In order to validate the effectiveness of the proposed SNS-SRE face sketch–photo synthesis method, the algorithms in [18], [24], and [25] are used as comparisons in this section. For the algorithm in [18], some key parameters are set as follows. The number of nearest neighbors is set to 5, and the patch size is 36×36 with 75% regions overlapping. The larger the patch size, the more serious the blocking

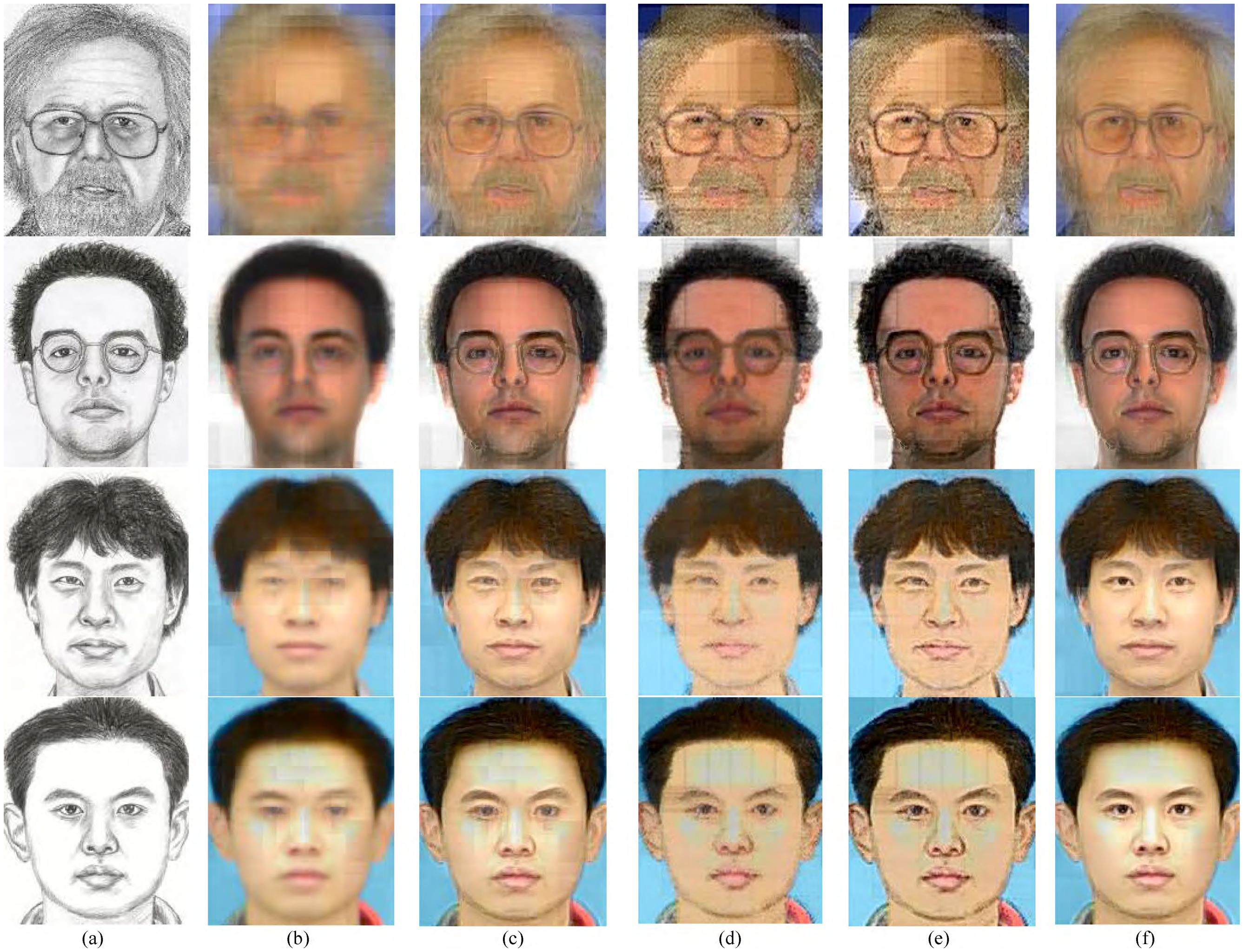


Fig. 8. Pseudophoto synthesis results on the CUFS database. (a) Original sketches. (b) Pseudophotos generated by [18]. (c) Pseudophotos generated by SRE-LLE method. (d) Pseudophotos generated by [25]. (e) Pseudophotos generated by SRE-EHMM method. (f) Pseudosketches generated by the proposed SNS-SRE method.

effect. If the patch size is smaller, the noise is obvious and more nearest neighbors will result in oversmoothing. For algorithms [24] and [25], we adopt the same settings as given in the literature and employ the leave-one-out strategy [48] to synthesize a pseudosketch or a pseudophoto. This means leaving out one photo or sketch as the test sample, while the other sketch–photo pairs are used as training samples. For the proposed SNS method, the patch size is 36×36 with 28×28 area overlapping and patch size 9×9 with 7×7 overlapping for the SRE enhancement stage. For CUHK student and XM2VTS database, 88 and 100 sketch–photo pairs are applied for training, and leave-one-out strategy for Purdue AR database. Since there are five sketches for each face photo in the VIPSL database, there are five groups of sketch–photo pairs, which are denoted by A, B, C, D, and E. Each group of sketch–photo pairs is composed of sketches drawn by the same artist and the corresponding face photos. For each group, 100 sketch–photo pairs are selected for training and the remaining 100 sketch–photo pairs are used for testing. In this way, we can obtain five pseudoimages of each person.

Some examples generated by the proposed SNS and KNN-based (e.g., [18]) approaches are shown in Fig. 6. The blurring

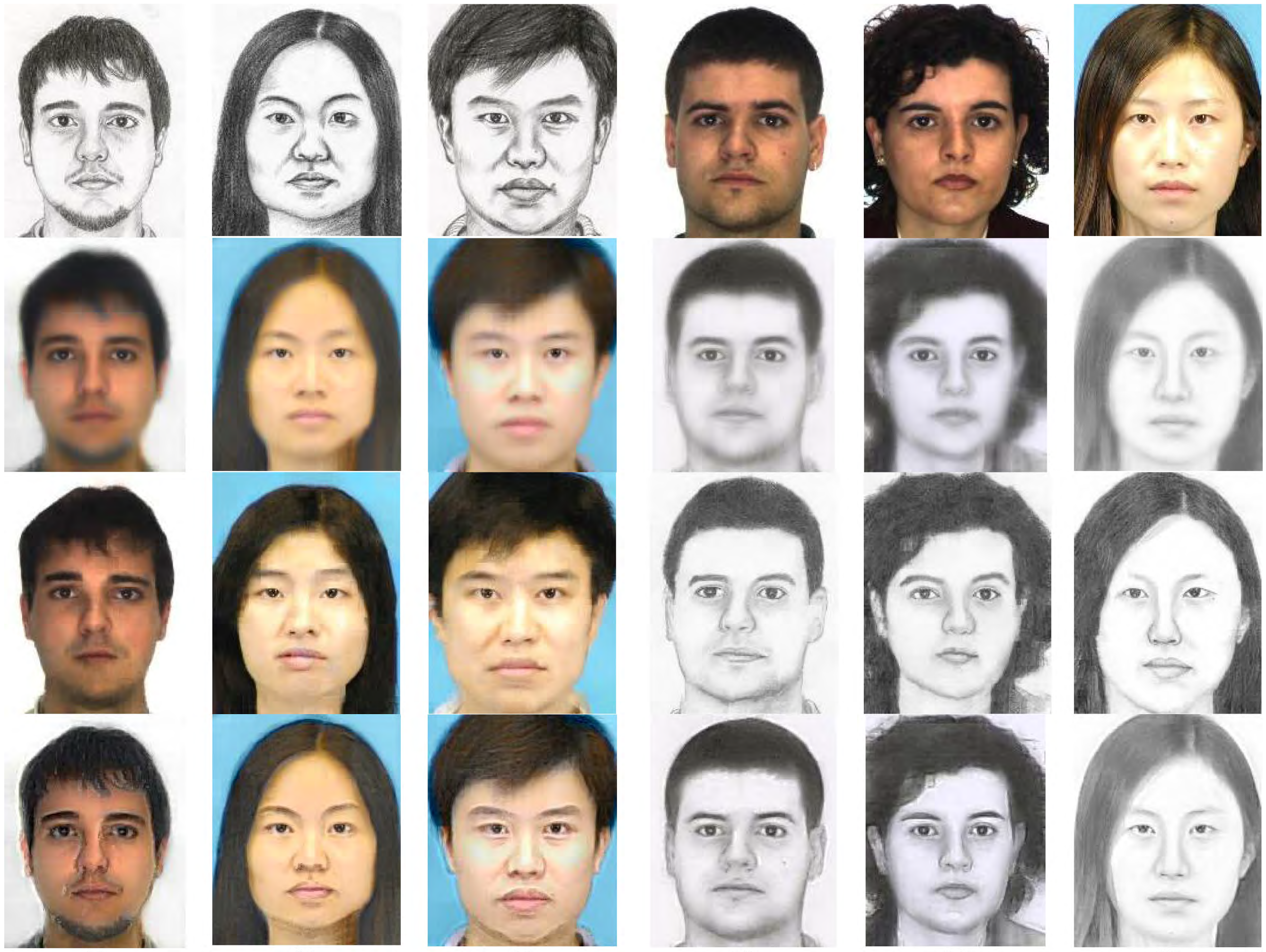
TABLE V

SRC Versus Generic Recognition Methods With Photos as the Training Set (%)

	[18]	[25]	SNS-SRE
Eigenface	71.5	74.7	80.4
Fisherface	52.8	69.3	72.2
LPP	30.8	51.2	45.7
SRC	90.1	94.6	**96.5**

effect in the pseudosketches generated by [18] is evident. Fortunately, there are fewer noises in the sketches synthesized by the proposed SNS method. The similar local search strategy can be adopted as in [19] or [49] for the propose method.

To further validate the effectiveness of the proposed SRE method, we combine it with the algorithm in [18], [24], and [25]. For the sake of simplicity, let SRE-LLE denote the combination of SRE and [18] and SRE-EHMM denote the combination of SRE and [24] or [25]. Some experimental pseudosketches and pseudophotos on the CUFS database are shown in Figs. 7 and 8, respectively. Since the method in [19] is one of state-of-the-arts, we have also provided some



(a) (b) (c) (d) (e) (f)

Fig. 9. Comparisons between MRF method [19] and the proposed SNS and SNS-SRE methods. (a)-(f) Six different people. The top row lists the original photos or sketches drawn by the artist. The second row lists the results of the proposed SNS method. The third row lists the results of MRF. The last row lists the results of the proposed SNS-SRE method.

TABLE VI
SRC VERSUS GENERIC RECOGNITION METHODS WITH PSEUDOPHOTOS AS THE TRAINING SET (%)

	[18]	[25]	SNS-SRE
Eigenface	84.9	88.6	89.1
Fisherface	89.9	88.8	90.2
LPP	86.2	87.3	81.3
SRC	88.3	94.1	**95.3**

TABLE VII
FACE RECOGNITION USING PHOTO AS THE TRAINING SET (%)

Algorithm	[18]	[25]	SNS-SRE
Recognition rate	81	86.5	**89**

comparisons between this method[3] (denoted as "MRF") and the proposed method on the CUFS full dataset in Fig. 9. In addition, the results on the VIPSL database are presented in

[3]All the results of the method in [18] come from the authors' public website http://www.ee.cuhk.edu.hk/~xgwang/sketch_multiscale.html.

TABLE VIII
FACE RECOGNITION USING SKETCH AS THE TRAINING SET (%)

Algorithm/Number of Training Samples per Class	1	3	5
[18]	83	94.5	98
[24]	84.5	95	98
SNS-SRE	**90.5**	**97**	**99**

Figs. 10 and 11. It can be seen that the pseudosketches and pseudophotos synthesized by [18], [24], and [25] have either a blurring effect, heavy noise, or low definition. For synthesized images on the VIPSL database especially, due to the diversity of face photos and sketches, algorithms [18], [24], and [25] do not work as well. However, the proposed SNS-SRE method performs well. The combination of the SRE method with [18], [24], and [25] can also achieve a more competitive result than these three algorithms alone. From Fig. 9, we can see that although the MRF method achieves a noiseless appearance while our SRE step may introduce little noise, it is subjected to some deformation especially for the mouth, beard, moustache,

Fig. 10. Synthesized pseudosketches on VIPSL database. (a) Original photos. (b)–(e) Pseudosketches generated by [18], SRE-LLE, [24], SRE-EHMM, respectively. (f) Pseudosketches generated by the proposed SNS-SRE method.

hair, and chin area. From our analysis, this is due to the fact that in this method each patch of the synthesized image originally comes from the training image and if there is no patch with the same structure in the training dataset, the MRF method still selects one patch as its counterpart that results in the deformation. While for the proposed method, each patch of the synthesized image is generated from a combination of some candidate patches that may interpolate some patches that do not appear in the training set. To further substantiate the proposed method's superiority, the following two group experiments are implemented: subjective image quality assessment and image retrieval, which can be viewed as an objective image quality assessment.

B. Subjective Image Quality Assessment

In order to validate that the pseudosketches and pseudophotos synthesized by the proposed algorithm have good image quality, we invited 20 volunteers to score one quarter of the pseudoimages produced by all four algorithms ([18], [24], [25], and the proposed SNS-SRE). The mean opinion score (MOS) that has been used in ITU-T p.910, a standard in multimedia services, was then calculated. The criterion for choosing pseudoimages is detailed as follows. We first randomly choose one quarter of the face photos in the CUHK student database and VIPSL database. The sketches and associated pseudoimages are shown at the same time. For every volunteer, one original image (sketch or photo) and its corresponding pseudoimages are shown at the same time. Thereafter, they score each pseudoimage. Taking the pseudosketch as an example, a sketch drawn by the artist is used as a reference image. Pseudosketches generated by [18], [24], and SNS-SRE are assessed in parallel. For each pseudosketch, we compute the mean score of 20 scores graded by volunteers. Finally, the algorithm is measured by calculating the mean score of these pseudosketches generated by the same algorithm. In this experiment, the score ranges from 1 to 5; thus five rates, i.e., *bad*, *poor*, *fair*, *good*, and *excellent*, are graded. The score from every volunteer must be a multiple of 0.5 considering the limitation of human visual resolution, but the final mean score is not limited to this constraint. The MOS value of an image is calculated as follows:

$$MOS(l) = \frac{1}{20}\sum_{i=1}^{20} A(i,l) \tag{19}$$

where $A(i,l)$ denotes the lth image's score from the ith volunteer. The final results are shown in Table II.

Fig. 11. Synthesized pseudophotos on VIPSL database. (a) Original sketches. (b)–(e) Pseudophotos generated by [18], SRE-LLE, [25], SRE-EHMM, respectively. (f) Pseudophotos generated by the proposed SNS-SRE method.

In Table II, "–" denotes that the algorithm is not used to synthesize the corresponding pseudoimage. The notations [18], [24], and [25] denote the corresponding algorithms in the literature. From Table II, it can be found that the proposed SNS-SRE algorithm does enhance the subjective quality of the pseudoimages. For the CUHK student database, especially, the scores of the pseudosketches and pseudophotos generated by the SNS-SRE algorithm reach 4.35 and 4.65, respectively. These algorithms achieve a higher score on the CUHK student database than on the VIPSL database. This is because the photos in the CUHK student database are taken with the same background, but the photos in the VIPSL database are taken against various backgrounds and include people of multiple races, as shown in Fig. 5. It is therefore more difficult for the VIPSL database to learn the various backgrounds correctly. Subsequently, we conduct an example-based retrieval to further illustrate the effectiveness of the proposed algorithm.

C. Example-Based Retrieval

As listed in Table I, there are four retrieval modes. Here, "example-based" denotes "sketch-based," "photo-based," "pseudosketch-based," or "pseudophoto-based." To illustrate the effectiveness of the SRC algorithm applied to our special topic, we compare the retrieval results with three popular generic recognition algorithms[4] (eigenfaces [50], Fisherfaces [51], and LPP [52]) by using images and pseudoimages from, or generated from, the CUFS database. The comparison results are reported in Tables III–VI. It should be noted that in these four tables, the items in first row ([18], [24], and SNS-SRE) were applied to synthesize sketches or photos and in the column (eigenface, Fisherface, LPP, and SRC) were exploited to perform face recognition. From these tables, we can see that the SRC approach outperforms the three generic recognition methods to some extent on the special topic of synthesized image retrieval. The method in [19] is one of the state-of-the-art methods and they reported 93.3% and 96.3% recognition rates corresponding to 93.7% and 96.5% of our method for pseudosketch-based and pseudophoto-based retrieval, respectively.

From an analysis of the sparse-representation-based face retrieval algorithm, the number of training samples within the same class affects the retrieval results. Thus, experiments are designed on the VIPSL database to demonstrate this conclusion. Since there is only one face photo in the VIPSL

[4]The MATLAB codes for these three traditional algorithms are downloaded from the website http://www.zjucadcg.cn/dengcai/Data/data.html.

TABLE IX

FACE RECOGNITION USING PSEUDOSKETCH AS THE TRAINING SET (%)

Algorithm/Number of Training Samples per Class	1	3	5
[18]	87.5	88.5	89
[24]	89.5	89.5	**90.5**
SNS-SRE	**90**	**91**	**90.5**

TABLE X

FACE RECOGNITION USING PSEUDOPHOTO AS THE TRAINING SET (%)

Algorithm/Number of Training Samples per Class	1	3	5
[25]	84	92	**96**
[18]	81	88	92
SNS-SRE	**86**	**93.5**	**96**

database, 100 photos are arranged to form the dictionary for the retrieval on the photo database, with one training sample per class. The results are shown in Table VII. For the remaining three retrieval modes, the number of training samples per class is set to 1, 3, and 5, respectively. The result is shown in Tables VIII–X.

Since the photos in the VIPSL database have various backgrounds and persons in the database come from different areas or even different races, there are more challenges in retrieval from the database. As shown in Tables VIII and X, the number of training samples per class affects the recognition rate. With the increase in the number of training samples, the recognition rate can improve remarkably. However, for face recognition using pseudosketches as the training set, the number of training samples has little impact on the final recognition rate, with only 0.5%–1.5% improvement for all three algorithms. All in all, the proposed SNS-SRE algorithm performs better than the algorithms proposed in [18], [24], and [25].

V. CONCLUSION

In this paper, we proposed a novel face sketch–photo synthesis and retrieval algorithm based on sparse representation. The algorithm consists of two stages. In the synthesis stage, we first approached a sparse neighbor selection method to obtain an initial estimate of the pseudoimage. An enhanced strategy was then introduced to improve the definition of the initial estimate. In the subsequent retrieval stage, we applied the sparse-representation-based face recognition algorithm to perform retrieval. Extensive experimental results on two databases showed that the proposed algorithm improved the quality of synthesized images. In this paper, we just used five sketches of a face photo for retrieval. In a future investigation, we will study how to use these five sketches to synthesize a higher quality sketch or a photo. Furthermore, we will use the idea of image denoising to further improve the definition of pseudoimages.

ACKNOWLEDGMENT

The authors appreciate the helpful comments and suggestions from the anonymous reviewers.

REFERENCES

[1] W. Zhao, R. Chellappa, A. Rosenfeld, and P. Phillips, "Face recognition: A literature survey," *ACM Comput. Survey*, vol. 34, no. 4, pp. 399–458, 2003.

[2] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 947–954.

[3] N. Guan, D. Tao, Z. Luo, and B. Yuan, "Manifold regularized discriminative nonnegative matrix factorization with fast gradient descent," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 2030–2048, Jul. 2011.

[4] N. Guan, D. Tao, Z. Luo, and B. Yuan, "Non-negative patch alignment framework," *IEEE Trans. Neural Netw.*, vol. 22, no. 8, pp. 1218–1230, Aug. 2011.

[5] T. Zhou, D. Tao, and X. Wu, "Manifold elastic net: A unified framework for sparse dimension reduction," *Data Mining Knowl. Discovery*, vol. 22, no. 3, pp. 340–371, 2011.

[6] D. Tao, X. Li, X. Wu, and S. Maybank, "Geometric mean for subspace selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 260–274, Feb. 2009.

[7] 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.

[8] S. E. Brennan, "Caricature generator," M.S. thesis, MIT Media Lab., Massachusetts Instit. Technol., Cambridge, MA, 1982.

[9] M. Tominaga, S. Fukuoka, K. Murakami, and H. Koshimizu, "Facial caricaturing with motion caricaturing in PICASSO system," in *Proc. IEEE/ASME Int. Conf. Advanced Intell. Mechatronics*, Jun. 1997, pp. 30–37.

[10] H. Koshimizu, M. Tominaga, T. Fujiwara, and T. Murakami, "On KANSEI facial processing for computerized facial caricaturing system PICASSO," in *Proc. IEEE Int. Conf. Syst. Man Cybern.*, Oct. 1999, pp. 294–299.

[11] Y. Li and H. Kobatake, "Extraction of facial sketch based on morphological processing," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 1997, pp. 316–319.

[12] H. Chen, N. Zheng, Y. Xu, and H. Shum, "An example-based facial sketch generation system," *J. Softw.*, vol. 14, no. 2, pp. 202–208, 2003.

[13] J. Wang, H. Bao, and W. Zhou, "Automatic image-based pencil sketch rendering," *J. Comput. Sci. Technol.*, vol. 17, no. 5, pp. 347–356, 2002.

[14] B. Klar, Z. Li, and A. Jain, "Matching forensic sketches to mug shot photos," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 639–646, Mar. 2011.

[15] X. Tang and X. G. Wang, "Face photo recognition using sketch," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2002, pp. 257–260.

[16] X. Tang and X. G. Wang, "Face sketch recognition," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 14, no. 1, pp. 50–57, Jan. 2004.

[17] X. Tang and X. Wang, "Face sketch synthesis and recognition," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2003, pp. 687–694.

[18] Q. Liu and X. Tang, "A nonlinear approach for face sketch synthesis and recognition," in *Proc. IEEE. Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 1005–1010.

[19] 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.

[20] W. Liu, X. Tang, and J. Liu, "Bayesian tensor inference for sketch-based facial photo hallucination," in *Proc. Int. Joint Conf. Artif. Intell.*, Jan. 2007, pp. 2141–2146.

[21] W. Zhang, X. Wang, and X. Tang, "Lighting and pose robust face sketch synthesis," in *Proc. Eur. Conf. Comput. Vis.*, vol. 6. Sep. 2010. pp. 420–433.

[22] Y. Li, M. Savvides, and V. Bhagavatula, "Illumination tolerant face recognition using a novel face from sketch synthesis approach and advanced correlation filters," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, May 2006, pp. 357–360.

[23] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, Dec. 2000.

[24] X. Gao, J. Zhong, D. Tao, and X. Li, "Local face sketch synthesis learning," *Neurocomputing*, vol. 71, nos. 10–12, pp. 1921–1930, 2008.

[25] B. Xiao, X. Gao, X. Li, and D. Tao, "A new approach for face recognition by sketches in photos," *Signal Process.*, vol. 89, no. 8, pp. 1531–1539, 2009.

[26] J. Zhong, X. Gao, and C. Tian, "Face sketch synthesis using E-HMM and selective ensemble," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Apr. 2007, pp. 485–488.

[27] B. Xiao, X. Gao, D. Tao, Y. Yuan, and J. Li, "Photo-sketch synthesis and recognition based on subspace learning," *Neurocomputing*, vol. 73, nos. 4–6, pp. 840–852, 2010.

[28] Z. H. Zhou, J. Wu, and W. Tang, "Ensemble neural networks: Many could be better than all," *Artif. Intel.*, vol. 137, nos. 1–2, pp. 239–263, 2002.

[29] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[30] M. Yang and L. Zhang, "Gabor feature based sparse representation for face recognition with Gabor occlusion dictionary," in *Proc. Eur. Conf. Comput. Vis.*, vol. 6. Sep. 2010, pp. 448–461.

[31] M. Yang, L. Zhang, D. Zhang, and J. Yang, "Metaface learning for sparse representation based face recognition," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 1601–1604.

[32] L. Zhang, M. Yang, Z. Feng, and D. Zhang, "On the dimensionality reduction for sparse representation based face recognition," in *Proc. Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 1237–1240.

[33] X. Yuan and S. Yan, "Visual classification with multi-task joint sparse representation," in *Proc. IEEE. Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 3493–3500.

[34] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[35] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *Proc. IEEE. Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[36] W. Freeman, E. Pasztor, and O. Carmichael, "Learning low-level vision," *Int. J. Comput. Vis.*, vol. 40, no. 1, pp. 25–47, 2000.

[37] J. Sun, N. Zheng, H. Tao, and H. Shum, "Image hallucination with primal sketch priors," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2003, pp. 729–736.

[38] H. Lee, A. Battle, R. Raina, and A. Ng, "Efficient sparse coding algorithms," in *Proc. Adv. Neural Inform. Process. Syst.*, 2007, pp. 801–808.

[39] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[40] J. Mairal, G. Sapiro, and M. Elad, "Learning multiscale sparse representations for image and video restoration," *SIAM Multiscale Model. Simul.*, vol. 7, no. 1, pp. 214–241, 2008.

[41] D. L. Donoho, "For most large underdetermined systems of linear equations, the minimal ℓ^1-norm near-solution approximates the sparsest near-solution," *Comm. Pure Appl. Math.*, vol. 59, no. 7, pp. 907–934, 2006.

[42] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *J. Roy. Statist. Soc. B*, vol. 58, no. 1, pp. 267–288, 1996.

[43] A. Martinez and R. Benavente, "The AR face database," Ohio State Univ., Columbus, OH, CVC Tech. Rep. 24, 1998.

[44] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, "XM2VTSDB: The extended M2VTS database," in *Proc. Int. Conf. Audio Video-Based Person Authentication*, 1999, pp. 72–77.

[45] P. J. Phillips, H. Wechsler, J. Huang, and P. Rauss, "The FERET database and evaluation procedure for face recognition algorithms," *Image Vis. Comput. J.*, vol. 16, no. 5, pp. 295–306, 1998.

[46] P. Phillips, H. Moon, S. Rizvi, and P. 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] A. Serrano, I. de Diego, C. Conde, E. Cabello, L. Shen, and L. Bai, "Influence of wavelet frequency and orientation in an SVM-based parallel Gabor PCA face verification system," in *Proc. Conf. Intell. Data Eng. Automated Learning*, vol. 4881. Dec. 2007, pp. 219–228.

[48] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *J. Roy. Statist. Soc.*, vol. 36, no. 2, pp. 111–147, 1974.

[49] N. Wang, X. Gao, D. Tao, and X. Li, "Face sketch-photo synthesis via multi-dictionary sparse representation framework," in *Proc. Int. Conf. Image Graph.*, Aug. 2011, pp. 82–87.

[50] M. Turk and A. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE. Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 1991, pp. 586–591.

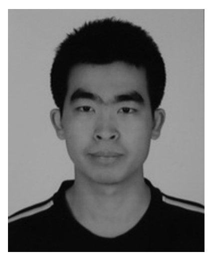
[51] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces versus fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[52] X. He, "Locality preserving projections," Ph.D. dissertation, Comput. Sci. Dept., Univ. Chicago, Chicago, IL, 2005.

Xinbo Gao (M'02–SM'07) received the B.Eng., M.Sc., and Ph.D. degrees in signal and information processing from Xidian University, Xi'an, China, in 1994, 1997, and 1999, respectively.

From 1997 to 1998, he was a Research Fellow with the Department of Computer Science, Shizuoka University, Shizuoka, Japan. From 2000 to 2001, he was a Post-Doctoral Research Fellow with the Department of Information Engineering, Chinese University of Hong Kong, Shatin, Hong Kong. Since 2001, he has been with the School of Electronic Engineering, Xidian University. Currently, he is a Professor of pattern recognition and intelligent systems and the Director of the VIPS Laboratory, Xidian University. His current research interests include computational intelligence, machine learning, computer vision, pattern recognition, and wireless communications. In these areas, he has published five books and around 150 technical articles in refereed journals and proceedings, including the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON NEURAL NETWORKS, the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, *Pattern Recognition*, and so on.

Dr. Gao is on the editorial boards of journals, including *EURASIP Signal Processing* (Elsevier) and *Neurocomputing* (Elsevier). He has served as the General Chair/Co-Chair or the Program Committee Chair/Co-Chair or a PC Member for around 30 major international conferences. He is currently a fellow of IET.



Nannan Wang received the B.Sc. degree in information and computation science from the Xi'an University of Posts and Telecommunications, Xi'an, China, in 2009. He is currently pursuing the Ph.D. degree in intelligent information processing with the VIPS Laboratory, School of Electronic Engineering, Xidian University, Xi'an.

Since September 2011, he has been a Visiting Ph.D. Student with the University of Technology, Sydney, NSW, Australia. His current research interests include machine learning and computer vision.

Dacheng Tao (M'07–SM'12) received the B.Eng. degree from the University of Science and Technology of China, Hefei, China, the M.Phil. degree from the Chinese University of Hong Kong, Shatin, Hong Kong, and the Ph.D. degree from the University of London, London, U.K.

Currently, he is a Professor with the Center for Quantum Computation and Information Systems, Faculty of Engineering and Information Technology, University of Technology, Sydney, NSW, Australia. He mainly applies statistics and mathematics for data analysis problems in data mining, computer vision, machine learning, multimedia, and video surveillance. He has authored or co-authored more than 100 scientific articles in top journals, including IEEE T-PAMI, T-KDE, T-IP, NIPS, AISTATS, AAAI, CVPR, ECCV, ICDM, ACM T-KDD, and KDD.

Dr. Tao has received many Best Paper Awards.

Xuelong Li (M'02–SM'07–F'12) is currently a Researcher (Full Professor) with the State Key Laboratory of Transient Optics and Photonics and the Director of the Center for Optical Imagery Analysis and Learning, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China.