

Transductive Face Sketch-Photo Synthesis

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

Abstract—Face sketch-photo synthesis plays a critical role in many applications, such as law enforcement and digital entertainment. Recently, many face sketch-photo synthesis methods have been proposed under the framework of inductive learning, and these have obtained promising performance. However, these inductive learning-based face sketch-photo synthesis methods may result in high losses for test samples, because inductive learning minimizes the empirical loss for training samples. This paper presents a novel transductive face sketch-photo synthesis method that incorporates the given test samples into the learning process and optimizes the performance on these test samples. In particular, it defines a probabilistic model to optimize both the reconstruction fidelity of the input photo (sketch) and the synthesis fidelity of the target output sketch (photo), and efficiently optimizes this probabilistic model by alternating optimization. The proposed transductive method significantly reduces the expected high loss and improves the synthesis performance for test samples. Experimental results on the CUHK face sketch dataset demonstrate the effectiveness of the proposed method by comparing it with representative inductive learning-based face sketch-photo synthesis methods.

Index Terms—Probabilistic graph model, quadratic programming, sketch-photo synthesis, transductive learning

I. INTRODUCTION

FACE sketch-photo synthesis facilitates many applications such as law enforcement and digital entertainment and has attracted increasing attention [1], [2], [3], [4]. In applications like law enforcement, a photo of the suspect is not always available due to limitations in environmental conditions or the suspect’s intentional attempt at concealment. A sketch drawn by an artist cooperating with an eyewitness might be used as a substitute for a photo, but the significant differences between sketches and mug-shots make straightforward matching to identify the suspect difficult [5], [6].

To reduce the discrepancies between a sketch and its corresponding photo, face sketch-photo synthesis is applied as a pre-process before matching by synthesizing a sketch

This work is supported by the National Basic Research Program of China(973 Program) (Grant No. 2012CB316400), the National Natural Science Foundation of China (Grant Nos: 61125204, 61125106, 61172146, 91120302, 61072093, 61201294, and 61001203), the Fundamental Research Funds for the Central Universities (Grant No. K5051202048 and K50513100009), and Shaanxi Innovative Research Team for Key Science and Technology (No.2012KCT-02)

N. Wang, X. Gao, J. Li are with VIPS Laboratory, School of Electronic Engineering, Xidian University, Xi’an 710071, Shaanxi, P. R. China (e-mail: nannanwang.xidian@gmail.com; xbgao@mail.xidian.edu.cn).

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

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

from an input photo or a photo from a source sketch. The key issue of face sketch-photo synthesis is how to learn the mapping relation between sketches and their counterparts. Thus, face sketch-photo synthesis is related to other cross-style image transfers such as face super-resolution (or face hallucination) [7], [8], artistic rendering [9], [10], [11], inter-modality biometrics [12], [13], [14], portrait painting [15], and heterogeneous image transformation [16]. In addition, face sketch-photo synthesis helps artists to simplify the animation procedure.

A. Previous Work: Face Sketch-Photo Synthesis

Existing methods can be classified into three main categories: subspace learning-based, sparse representation-based, and Bayesian inference-based.

Subspace learning mainly refers to linear subspace-based methods (e.g. principal component analysis, or PCA [17]), and nonlinear subspace methods such as manifold learning-based methods (e.g. local linear embedding, or LLE [18]). Tang and Wang [6], [19] proposed a linear face sketch synthesis method based on PCA called eigensketch transformation. They assumed that a source input photo and the target output sketch shared the same projection coefficients obtained from the PCA procedures. The coefficients are first obtained by projecting the input photo onto the training photos. The target sketch is then synthesized from a linear combination of training sketches weighted by the obtained projection coefficients. Tang and Wang [20] then proposed an improved method by separating the shape from the texture. The eigensketch transformation method [6], [19] is applied to shape and texture to compute the target shape and texture, respectively. The synthesized shape and texture are fused to obtain the final target sketch. Based on a similar idea to [6] and [19], Li et al. [21] proposed a hybrid subspace method for face photo synthesis by concatenating the training sketches and photos. Since the mapping between the whole sketches and corresponding photos is not linear, the above methods hardly synthesize a realistic sketch, especially when the hair region was included. Liu et al. [22] improved these linear models by exploiting LLE. Images are divided into overlapping patches. For each test photo patch, K photo patches are selected from the training set according to the Euclidean distance metric. Simultaneously, the corresponding K sketch patches are taken as the candidates. The weights for combining the K candidates are calculated by minimizing the least square of the reconstruction residual between the test photo patch and its K nearest neighbors. The target sketch patch is generated from a linear combination of those K candidates. Finally the target sketch is fused from these patches with an averaged overlapping area, which incurs a

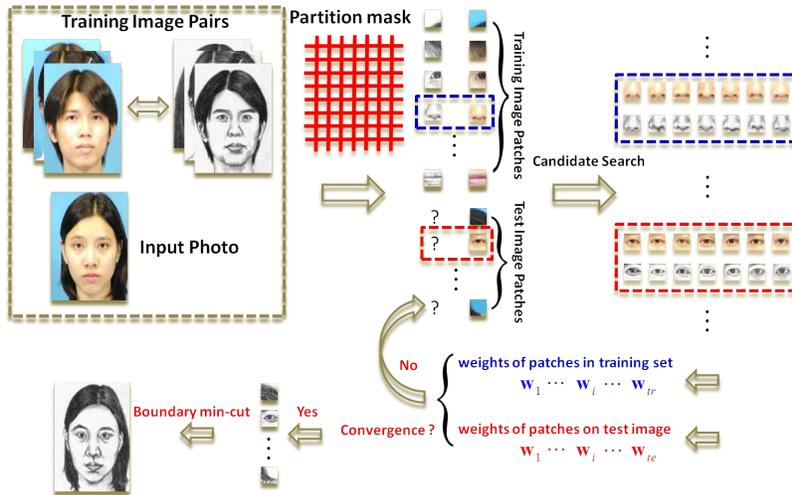


Fig. 1. Framework of face sketch synthesis

degree of blurring. The same LLE procedure was applied to generate an initial estimate in Liu et al.’s work [23]. They explored the idea of face hallucination to further compensate the residual which might be lost in the initial estimation procedure.

Sparse representation has wide applications in compressed sensing [24], image processing [25], computer vision and pattern recognition [26]. It can decompose a signal into a linear combination of atoms (or bases) weighted by a sparse vector. Chang et al. [27] introduced sparse coding [28] to face sketch synthesis. They first learned two coupled dictionaries (sketch patch dictionary and photo patch dictionary) via sparse coding. The coupled dictionaries denotes that they are learned from joint training by concatenating these two sub-dictionaries into one dictionary, so that the input training photo patch and the corresponding sketch patch have the same sparse representation coefficients. A test photo patch is then decomposed on the photo patch dictionary, weighted by the sparse representation coefficients. Hence from the linear combination of the atoms of the sketch patch dictionary, the target sketch patch is produced. Considering that the weighted combination of the candidates may result in the loss of high frequency information, Wang et al. proposed a two-step framework [29], [16] to further enhance the definition of the target output. In the first stage, a sparse feature selection algorithm is explored to synthesize an initial image which found closely related neighbors adaptively through sparse representation. In the second stage, the sparse coding strategy is explored to learn the mapping between the high frequency information of sketch patches and photo patches [29]. This two-step framework was extended to heterogeneous image transformations [16]. Ji et al. [30] also used sparse representation to perform face sketch synthesis from a regression perspective. Wang et al. [31] proposed a semi-coupled dictionary learning scheme which allowed the sketch patch and the corresponding photo patch to have different sparse representation coefficients, improving the flexibility of the mapping relation between sketches and photos.

Bayesian inference-based methods include embedded hidden Markov model (E-HMM)-based and Markov random fields (MRF)-based methods. Gao et al. [5] modeled the mapping between sketches and their photo counterparts by E-HMM. A face is decomposed into five super-states (corresponding to forehead, eye, nose, mouth, and chin) and sub-states (states in each super-state). Assuming the sketch and the photo have the same transition probability matrix, a joint training strategy is adopted. They then improved the proposed model in a local patch-based form which synthesized more detail [32], [33], [34]. All the above methods (subspace learning-based, sparse representation-based and E-HMM based) independently synthesize a target image patch and ignore the neighboring relation between overlapping patches. Taking the neighboring relation between overlapping patches as a regularization term, a number of MRF-based methods were proposed [3], [4], [35]. Wang and Tang [3] explored MRF to model the interaction between sketch-photo pairs and neighborhoods through two compatibility functions: data compatibility (constraints on the fidelity between the test photo patch and the target sketch patch) and spatial compatibility (constraints on the neighboring patches of the target sketch). They then extended this method by proposing a lighting and pose robust face sketch synthesis algorithm [4] which integrated shape prior, local evidence (data compatibility) and neighboring compatibility into a unified framework. Considering the fact that the MAP criterion used in [3] and [4] only selected one best matching patch as the target patch, which brought in some deformations due to insufficient image patches in the training set, Zhou et al. [35] proposed a weighted MRF model which selected K patches to interpolate new target patches not existing in the training set.

B. The Proposed Approach

All the aforementioned methods conduct face sketch synthesis or photo synthesis from the inductive learning perspective. Although they have obtained promising synthesis performance, these methods may result in high losses for a particular set

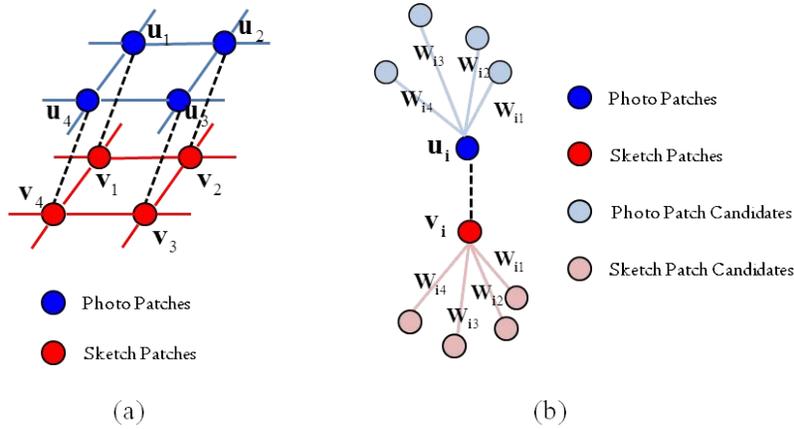


Fig. 2. Illustration of the constructed graph. (a) graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$. Photo patches (or sketch patches) can represent either training photo patches (training sketch patches) or test photo patches (target sketch patches) because we will construct the model from the perspective of transductive learning; (b) illustration of the candidate selection criterion. The number of nearest neighbors is $K = 4$. Weights on edges illustrate the similarity between a patch and its candidates.

of test samples. This is because inductive learning minimizes the empirical loss for training samples [36], [37]. In contrast, transductive learning algorithms minimize the expected loss for test samples by incorporating the given test samples into the learning process. Therefore, a transductive learning-oriented method may significantly reduce the high expected loss and improve the synthesis performance for the given test samples.

This paper presents a novel transductive face sketch-photo synthesis method. We mainly focus on face images in a frontal pose with normal lighting and neutral expression. All sketches and photos in this paper are divided into even patches with some overlap between neighboring patches. We design a probabilistic graphic model to model the relationship between sketch-photo patch pairs. This model takes both the reconstruction fidelity of the input photo (sketch) and the synthesis fidelity of the target output sketch (photo) into account. Furthermore, the relation between neighboring sketch patches is considered as a prior regularization on the hidden parameters. An alternative optimizing scheme is adopted to solve the proposed probabilistic model which converges in a small number of iterations. Finally, a min cut algorithm [9] is adopted to find the minimum error boundary to stitch the overlapping areas. The proposed method has the capability to handle both sketch synthesis and photo synthesis, because these two procedures are symmetric. In this paper, we take face sketch synthesis as an example to introduce our method, shown in Fig. 1.

The contributions of this paper are twofold: (1) it considers the face sketch-photo synthesis problem from a transductive view; and (2) it proposes an efficient and effective probabilistic framework for face sketch-photo synthesis.

In this paper, excepted when noted, we utilize a bold lowercase letter to denote a column vector, a bold uppercase letter to denote a matrix, and regular lowercase and uppercase letters to denote a scalar. The organization of the rest of the paper is as follows: Section II introduces the proposed transductive probabilistic method and implementation details. Experimental results and analyses are presented in Section III.

Section IV concludes the paper.

II. TRANSDUCTIVE FACE SKETCH SYNTHESIS

Suppose there are N sketch-photo pairs in the training set, represented by $(\mathbf{p}^1, \mathbf{s}^1), \dots, (\mathbf{p}^N, \mathbf{s}^N)$ where $\mathbf{p}^i, i = 1, \dots, N$ denotes the i -th photo and $\mathbf{s}^i, i = 1, \dots, N$ is the corresponding sketch in the training set. An input test photo can be denoted as \mathbf{x} and its target output is represented by \mathbf{y} . Each image should be divided into M patches; the training image patches, test photo patches and target sketch patches can then be expressed as follows respectively:

$$(\{\mathbf{p}_1^1, \dots, \mathbf{p}_M^1\}, \{\mathbf{s}_1^1, \dots, \mathbf{s}_M^1\}), \dots, (\{\mathbf{p}_1^N, \dots, \mathbf{p}_M^N\}, \{\mathbf{s}_1^N, \dots, \mathbf{s}_M^N\}), \{\mathbf{x}_1, \dots, \mathbf{x}_M\}, \{\mathbf{y}_1, \dots, \mathbf{y}_M\}$$

We can construct a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ to denote the relationship between sketch and photo patches, as shown in Fig. 2(a). In the MAP-MRF model [3], [4], only one matched sketch patch was selected for each test photo patch in the sketch synthesis process. As shown in [35], this method cannot generate a new sketch patch that does not exist in the training sketch patch set and thus some mismatches or deformations around the mouth and eyes are incurred. In our method, as shown in Fig. 2(b), we select K nearest neighbors as the candidates to synthesize the target sketch patch which can interpolate a new sketch patch that does not appear in the training sketch patches. This is because variables in MAP-MRF have discrete states, while variables in the proposed model have a continuous state space spanned by the selected K nearest neighbors, which allows the proposed model to synthesize sketch patches missing in the training set. It is reasonable to assume that sketch patches are similar if their corresponding counterparts are similar. The weights between an input photo patch and its neighbors are supposed to be the same as those between the target sketch patch and its candidates.

A. Probabilistic Model

Let \mathbf{X} denote the data matrix containing both training and test photo patches, \mathbf{Y} represent the corresponding training

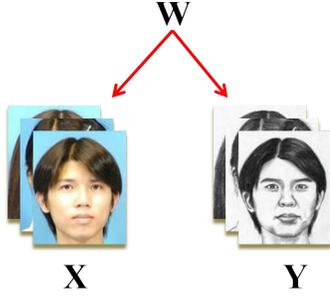


Fig. 3. Illustration of the generative process of photo patches and sketch patches from common hidden parameters.

sketch patches and target output sketch patches, and \mathbf{W} consist of the weight vectors. They are all arranged in a row form which means each row vector in \mathbf{X} denotes a photo patch and \mathbf{Y} and \mathbf{W} have a similar meaning:

$$\mathbf{X} = \begin{bmatrix} (\mathbf{p}_1^1)^T \\ \vdots \\ (\mathbf{p}_M^N)^T \\ \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_M^T \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_{(NM)} \\ \mathbf{X}_{(NM+1)} \\ \vdots \\ \mathbf{X}_{((N+1)M)} \end{bmatrix}$$

$$\mathbf{Y} = \begin{bmatrix} (\mathbf{s}_1^1)^T \\ \vdots \\ (\mathbf{s}_M^N)^T \\ \mathbf{y}_1^T \\ \vdots \\ \mathbf{y}_M^T \end{bmatrix} = \begin{bmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_{(NM)} \\ \mathbf{Y}_{(NM+1)} \\ \vdots \\ \mathbf{Y}_{((N+1)M)} \end{bmatrix}, \mathbf{W} = \begin{bmatrix} \mathbf{W}_1 \\ \vdots \\ \mathbf{W}_{(NM)} \\ \mathbf{W}_{(NM+1)} \\ \vdots \\ \mathbf{W}_{((N+1)M)} \end{bmatrix}$$

where \mathbf{X}_i , \mathbf{Y}_i , and $\mathbf{W}_i, i = 1, \dots, (N+1)M$ denote the i -th row vector of \mathbf{X} , \mathbf{Y} , and \mathbf{W} respectively (the same meaning as the following similar representations).

We can model the generative process of photo patches and sketch patches by their common hidden parameters \mathbf{W} as shown in Fig. 3. Sketch-photo patch pairs are sampled from the joint distribution $P(\mathbf{X}, \mathbf{Y}, \mathbf{W})$ governed by the weight matrix \mathbf{W} . We can then decompose the generative process as

$$\begin{aligned} P(\mathbf{Y}, \mathbf{X}, \mathbf{W}) &= P(\mathbf{Y}, \mathbf{X} | \mathbf{W})P(\mathbf{W}) \\ &= P(\mathbf{Y} | \mathbf{X}, \mathbf{W})P(\mathbf{X} | \mathbf{W})P(\mathbf{W}) \end{aligned} \quad (1)$$

where \mathbf{W} controls the generation of sketches and photos, $P(\mathbf{W})$ denotes the prior probability of the hidden parameters \mathbf{W} , $P(\mathbf{X} | \mathbf{W})$ is the conditional probability of generating photos \mathbf{X} given \mathbf{W} , and $P(\mathbf{Y} | \mathbf{X}, \mathbf{W})$ indicates the conditional probability of generating sketches \mathbf{Y} given corresponding photos \mathbf{X} and parameters \mathbf{W} . The prior probability is usually treated as a regularization term on \mathbf{W} . In this paper, we explore the compatibility constraint on overlapping regions between

neighboring sketch patches to model $P(\mathbf{W})$

$$\begin{aligned} P(\mathbf{W}) &\propto \prod_{(i,j) \in \mathcal{E}} \exp\left\{ -\frac{\|\sum_{k \in \mathcal{N}(i)} \mathbf{W}_{ik} \mathbf{Y}_k^{(i,j)} - \sum_{l \in \mathcal{N}(j)} \mathbf{W}_{il} \mathbf{Y}_l^{(j,i)}\|^2}{2\sigma_r^2} \right\} \\ &s.t. \sum_{k \in \mathcal{V}} \mathbf{W}_{ik} = 1, \mathbf{W}_{ik} \geq 0, \forall i \in \mathcal{V} \end{aligned} \quad (2)$$

where $\mathcal{N}(i)$ denotes the indices of neighbors of the node i , \mathbf{W}_{ik} represents the element located on the i -th row and the k -th column of the matrix \mathbf{W} , $\mathbf{Y}_k^{(i,j)}$ means the pixel intensities of the overlapping region (determined by the neighboring relation of nodes i and j) of the k -th neighbor of the node i , and $\mathbf{Y}_l^{(j,i)}$ means the pixel intensities of the overlapping region (determined by the neighboring relation of nodes i and j) of the l -th neighbor of the node j . Note that elements in the i -th $i = 1, 2, \dots, (N+1)M$ row of the weight matrix \mathbf{W} are 0 except for elements located in the neighborhood of the i -th node. This prior indicates the compatibility of the overlapping regions of neighboring sketch patches.

In face sketch synthesis, \mathbf{X} is observed, so $P(\mathbf{X}|\mathbf{W})$ is the likelihood function. It can be modeled as a Gaussian distribution:

$$P(\mathbf{X}|\mathbf{W}) \propto \prod_{i \in \mathcal{V}} \exp\left\{ -\frac{\|\mathbf{X}_i - \sum_{j \in \mathcal{N}(i)} \mathbf{W}_{ij} \mathbf{X}_j\|^2}{2\sigma_{dp}^2} \right\} \quad (3)$$

The probability $P(\mathbf{X}|\mathbf{W})$ mainly considers the reconstruction fidelity between a photo patch \mathbf{X}_i and its nearest neighbors $\mathbf{X}_j, j \in \mathcal{N}(i)$.

From Fig. 3, we find that given \mathbf{W} , \mathbf{Y} is conditionally independent of \mathbf{X} , i.e. $P(\mathbf{X}, \mathbf{Y}|\mathbf{W}) = P(\mathbf{X}|\mathbf{W})P(\mathbf{Y}|\mathbf{W})$ and then $P(\mathbf{Y}|\mathbf{X}, \mathbf{W}) = P(\mathbf{Y}|\mathbf{W})$. It is straightforward that the probability $P(\mathbf{Y}|\mathbf{W})$ can be modeled as a product of a series of independent and identically distributed normal distributions:

$$\begin{aligned} P(\mathbf{Y}|\mathbf{X}, \mathbf{W}) &= P(\mathbf{Y}|\mathbf{W}) \\ &\propto \prod_{i \in \mathcal{V}} P(\mathbf{Y}_i - \sum_{j \in \mathcal{N}(i)} \mathbf{W}_{ij} \mathbf{Y}_{(j)}) \\ &\propto \prod_{i \in \mathcal{V}} \exp\left\{ -\frac{\|\mathbf{Y}_i - \sum_{j \in \mathcal{N}(i)} \mathbf{W}_{ij} \mathbf{Y}_j\|^2}{2\sigma_{ds}^2} \right\} \end{aligned} \quad (4)$$

We can then reformulate the joint probability as in equation (5) on the next page. In face sketch synthesis, we are given an input photo as the test image and a number of sketch-photo pairs as the training set, and the objective is to infer the sketch corresponding to the input. This can be formulated as a maximum a posterior probability estimation problem (6). According to (5), the above maximization problem (6) is equivalent to the minimization of the objective function (7), where $\mathbf{M} = (\mathbf{I} - \mathbf{W})^T(\mathbf{I} - \mathbf{W})$, $\alpha = \frac{\sigma_{ds}^2}{\sigma_{dp}^2}$, $\beta = \frac{\sigma_{ds}^2}{\sigma_r^2}$, \mathbf{I} is the identity matrix with the same size as \mathbf{W} , the superscript T is the transpose operator and $tr(\cdot)$ is the trace operator.

We apply the alternating minimization method to the above problem to obtain a local solution according to the following two steps:

$$\begin{aligned}
 P(\mathbf{X}, \mathbf{Y}, \mathbf{W}) &\propto \prod_{i \in \mathcal{V}} \left\{ \exp\left\{ -\frac{\|\mathbf{Y}_{i \cdot} - \sum_{j \in \mathcal{N}(i)} \mathbf{W}_{ij} \mathbf{Y}_j\|^2}{2\sigma_{ds}^2} \right\} \exp\left\{ -\frac{\|\mathbf{X}_{i \cdot} - \sum_{j \in \mathcal{N}(i)} \mathbf{W}_{ij} \mathbf{X}_j\|^2}{2\sigma_{dp}^2} \right\} \right\} \\
 &\quad \prod_{(i,j) \in \mathcal{E}} \exp\left\{ -\frac{\|\sum_{k \in \mathcal{N}(i)} \mathbf{W}_{ik} \mathbf{Y}_k^{(i,j)} - \sum_{l \in \mathcal{N}(j)} \mathbf{W}_{il} \mathbf{Y}_l^{(j,i)}\|^2}{2\sigma_r^2} \right\} \\
 &= \exp\left\{ -\sum_{i \in \mathcal{V}} \left\{ \frac{\|\mathbf{Y}_{i \cdot} - \sum_{j \in \mathcal{N}(i)} \mathbf{W}_{ij} \mathbf{Y}_j\|^2}{2\sigma_{ds}^2} + \frac{\|\mathbf{X}_{i \cdot} - \sum_{j \in \mathcal{N}(i)} \mathbf{W}_{ij} \mathbf{X}_j\|^2}{2\sigma_{dp}^2} \right\} \right. \\
 &\quad \left. - \sum_{(i,j) \in \mathcal{V}} \frac{\|\sum_{k \in \mathcal{N}(i)} \mathbf{W}_{ik} \mathbf{Y}_k^{(i,j)} - \sum_{l \in \mathcal{N}(j)} \mathbf{W}_{il} \mathbf{Y}_l^{(j,i)}\|^2}{2\sigma_r^2} \right\}
 \end{aligned} \tag{5}$$

$$\begin{aligned}
 &\max_{\mathbf{W}, \mathbf{y}_1, \dots, \mathbf{y}_M} P(\mathbf{W}, \mathbf{y}_1, \dots, \mathbf{y}_M | \mathbf{x}_1, \dots, \mathbf{x}_M, \mathbf{X}_1, \dots, \mathbf{X}_{(NM)}, \mathbf{Y}_1, \dots, \mathbf{Y}_{(NM)}) \\
 &\Leftrightarrow \max_{\mathbf{W}, \mathbf{y}_1, \dots, \mathbf{y}_M} \frac{P(\mathbf{Y}, \mathbf{X}, \mathbf{W})}{P(\mathbf{X}, \mathbf{Y}_1, \dots, \mathbf{Y}_{(NM)})} \\
 &\Leftrightarrow \max_{\mathbf{W}, \mathbf{y}_1, \dots, \mathbf{y}_M} P(\mathbf{Y}, \mathbf{X}, \mathbf{W})
 \end{aligned} \tag{6}$$

$$\begin{aligned}
 &\min_{\mathbf{W}, \mathbf{y}_1, \dots, \mathbf{y}_M} \sum_{i \in \mathcal{V}} \left\{ \|\mathbf{Y}_{i \cdot} - \sum_{j \in \mathcal{N}(i)} \mathbf{W}_{ij} \mathbf{Y}_j\|^2 + \alpha \|\mathbf{X}_{i \cdot} - \sum_{j \in \mathcal{N}(i)} \mathbf{W}_{ij} \mathbf{X}_j\|^2 \right\} \\
 &\quad + \beta \sum_{(i,j) \in \mathcal{V}} \left\| \sum_{k \in \mathcal{N}(i)} \mathbf{W}_{ik} \mathbf{Y}_k^{(i,j)} - \sum_{l \in \mathcal{N}(j)} \mathbf{W}_{il} \mathbf{Y}_l^{(j,i)} \right\|^2 \\
 &\quad s.t. \quad \sum_{k \in \mathcal{V}} \mathbf{W}_{ik} = 1, \mathbf{W}_{ik} \geq 0, \forall i \in \mathcal{V} \\
 &\Leftrightarrow \min_{\mathbf{W}, \mathbf{y}_1, \dots, \mathbf{y}_M} \text{tr}(\mathbf{Y}^T \mathbf{M} \mathbf{Y}) + \alpha \text{tr}(\mathbf{X}^T \mathbf{M} \mathbf{X}) + \beta \sum_{(i,j) \in \mathcal{V}} \left\| \sum_{k \in \mathcal{N}(i)} \mathbf{W}_{ik} \mathbf{Y}_k^{(i,j)} - \sum_{l \in \mathcal{N}(j)} \mathbf{W}_{il} \mathbf{Y}_l^{(j,i)} \right\|^2 \\
 &\quad s.t. \quad \sum_{k \in \mathcal{V}} \mathbf{W}_{ik} = 1, \mathbf{W}_{ik} \geq 0, \forall i \in \mathcal{V}
 \end{aligned} \tag{7}$$

(1) fixing \mathbf{W} , update $\mathbf{y}_1, \dots, \mathbf{y}_M$ by solving (8)

$$\min_{\mathbf{y}_1, \dots, \mathbf{y}_M} \text{tr}(\mathbf{Y}^T \mathbf{M} \mathbf{Y}) \tag{8}$$

(2) then fixing $\mathbf{y}_1, \dots, \mathbf{y}_M$ to be the above obtained value, and update \mathbf{W} by solving (9)

$$\begin{aligned}
 &\min_{\mathbf{W}} \|\mathbf{U} - \mathbf{W}\mathbf{U}\|^2 \\
 &\quad + \beta \sum_{(i,j) \in \mathcal{V}} \left\| \sum_{k \in \mathcal{N}(i)} \mathbf{W}_{ik} \mathbf{Y}_k^{(i,j)} - \sum_{l \in \mathcal{N}(j)} \mathbf{W}_{il} \mathbf{Y}_l^{(j,i)} \right\|^2 \\
 &\quad s.t. \quad \sum_{k \in \mathcal{V}} \mathbf{W}_{ik} = 1, \mathbf{W}_{ik} \geq 0, \forall i \in \mathcal{V}
 \end{aligned} \tag{9}$$

where $\mathbf{U} = [\mathbf{X} \ \sqrt{\alpha} \mathbf{Y}]$ and we name each row of \mathbf{U} a *pairwise patch* since it consists of two patches extracted from a sketch-photo pair. We alternately conduct (8) and (9) until convergence. In the following section, we will introduce details of these two steps.

B. Implementation Details

In (9), each row of \mathbf{W} has at most K nonzero elements which correspond to K selected nearest neighbors. We de-

compose the update procedure of \mathbf{W} into $N + 1$ independent sub-problems, each of which takes sketch-photo patch pairs extracted from the same person as the test set. This means, to update the elements of the $(j \times M + 1)$ -th row to the $(j + 1) \times M$ -th row $j = 1, \dots, N$, sketch-photo patches extracted from $1, \dots, j-1, j+1, \dots, N$ -th sketch-photo pairs are taken as the training set, and the photo patches exacted from the j -th photo are the test set. The update of elements of the $(N \times M + 1)$ -th row to the $(N + 1) \times M$ -th row considers sketch-photo patches extracted from the $1, \dots, N$ -th sketch-photo pairs as the training set.

Each of the above sub-problems is equal to the quadratic programming problem defined in (10). We use $\mathbf{u}^i, i = 1, \dots, M$ to denote the M referred pairwise patches (concatenation of the corresponding sketch patches and photo patches), \mathbf{A}_i to denote a $2d \times (MK)$ matrix with $[(i-1) \times K + k]$ -th column denoting the k -th nearest neighbor of the patch \mathbf{u}^i (other columns are zero vectors), and $\mathbf{D}^{(i,j)}$ to denote a $d_{ij} \times (MK)$ matrix with the $[(i-1) \times K + k]$ -th column representing the overlapping pixel intensities of the k -th nearest neighbor of the patch \mathbf{u}^i (other columns are zero vectors, d_{ij} represents

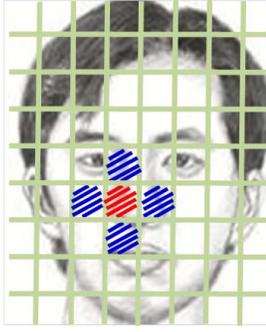


Fig. 4. A sub-problem in (10). Each sub-problem involves at most five patches as shown in the figure (also including each patch’s K nearest neighbors selected from the training set).

the number of overlapping pixels).

$$\begin{aligned} & \min_{\mathbf{w}} \mathbf{w}^T \mathbf{A} \mathbf{w} - 2 \mathbf{w}^T \mathbf{b} \\ & \text{s.t. } \mathbf{C} \mathbf{w} = \mathbf{e} \\ & \mathbf{w}_{i,k} \geq 0, \forall i \in \{1, \dots, M\}, k \in \{1, \dots, K\} \end{aligned} \quad (10)$$

where \mathbf{w} is a column vector size of $M \times K$, composed of the concatenation of all M row vectors of \mathbf{W} corresponding to the same person (only elements corresponding to the K nearest neighbors of each row are collected), $\mathbf{w}_{i,k}$ denotes the $[(i - 1) \times K + k]$ -th element of \mathbf{w} , $\mathbf{A} = \sum_{i=1}^M (\mathbf{A}^i)^T \mathbf{A}^i + \beta \sum_{(i,j) \in \mathcal{E}} (\mathbf{D}^{(i,j)} - \mathbf{D}^{(j,i)})^T (\mathbf{D}^{(i,j)} - \mathbf{D}^{(j,i)})$, $\mathbf{b} = \sum_{i=1}^M (\mathbf{A}^i)^T \mathbf{u}^i$, \mathbf{e} is an M -vector whose elements all equal 1, and \mathbf{C} is an $M \times (MK)$ matrix to normalize the sum of every K elements of \mathbf{w} from the same row of \mathbf{W} to be one.

In (10), \mathbf{A} is of size $(MK) \times (MK)$ and it is usually large-scale, so the above optimization problem is nearly computationally intractable. To make the above large-scale problem solvable, we can decompose it into sub-problems by exploring the Markov property according to [35]. Each sub-problem only refers to a group of patches comprising a patch and its neighboring patches, as illustrated in Fig. 4 (also including each patch’s K nearest neighbors selected from the training set). Here, only one group of patches is shown as an example (the red patch is surrounded by its neighboring blue patches). In fact, each patch marked in blue is also surrounded by its neighboring patches marked in red. All the pairwise patches of a single character can be grouped into two sets: patches marked in red and patches marked in blue. These sub-problems share a similar quadratic form to (10) but the meaning of each symbol is changed accordingly. We use the interior point method [38] to solve the quadratic programming problem for the weights in each sub-problem.

Note that the weights $\mathbf{W}_1, \dots, \mathbf{W}_{(NM)}$ of the training sketch-photo patch pairs can be computed once and then stored for the subsequent update process. Thus, in each update procedure of \mathbf{W} , only M weight vectors ($\mathbf{W}_{(NM+1)}, \dots, \mathbf{W}_{((N+1)M)}$) need to be calculated. In our experiments, elements of each row of \mathbf{W} which correspond to the K nearest neighbors are initialized to $\frac{1}{K}$.

For the optimization problem (8), we can find its closed-

form solution by

$$\frac{d \text{tr}(\mathbf{Y}^T \mathbf{M} \mathbf{Y})}{d \mathbf{Y}} = 0 \quad (11)$$

Denoting $\mathbf{M} = [\mathbf{M}_{tr} \quad \mathbf{M}_{te}]$, $\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_{tr} \\ \mathbf{Y}_{te} \end{bmatrix}$ and substituting them into (11), we have

$$\mathbf{M}_{tr} \mathbf{Y}_{tr} + \mathbf{M}_{te} \mathbf{Y}_{te} = 0 \quad (12)$$

Thus,

$$\mathbf{Y}_{te} = \begin{bmatrix} \mathbf{y}_1^T \\ \vdots \\ \mathbf{y}_M^T \end{bmatrix} = -((\mathbf{M}_{te})^T \mathbf{M}_{te})^{-1} (\mathbf{M}_{te})^T \mathbf{M}_{tr} \mathbf{Y}_{tr} \quad (13)$$

The above procedure can be decomposed into M similar sub-problems (see (14)), each of which solves a smaller-scale problem for one patch $\mathbf{y}_i, i = 1, \dots, M$. Thus, the parallel computation is applicable to improve efficiency.

$$\min_{\mathbf{y}_i} \text{tr}((\mathbf{Y}_i)^T \mathbf{M}^i \mathbf{Y}_i) \quad (14)$$

$$\mathbf{Y}^i = \begin{bmatrix} \tilde{\mathbf{S}}_i^{(1)} \\ \vdots \\ \tilde{\mathbf{S}}_i^{(N)} \\ \tilde{\mathbf{Y}}_i \end{bmatrix}, \tilde{\mathbf{S}}_i^{(j)} = \begin{bmatrix} \tilde{\mathbf{s}}_i^{(j1)} \\ \vdots \\ \tilde{\mathbf{s}}_i^{(jk)} \\ \tilde{\mathbf{s}}_i^{(j)} \end{bmatrix}, \tilde{\mathbf{Y}}_i = \begin{bmatrix} \tilde{\mathbf{y}}_i^1 \\ \vdots \\ \tilde{\mathbf{y}}_i^K \\ \tilde{\mathbf{y}}_i \end{bmatrix}$$

where $\mathbf{M}_i = (\mathbf{I} - \mathbf{W}^i)^T (\mathbf{I} - \mathbf{W}^i)$, $\tilde{\mathbf{s}}_i^{jk}$ denotes the k -th candidate of the training sketch patch $\mathbf{s}_i^{(j)}$, $\tilde{\mathbf{y}}_i^k$ represents the k -th candidate of the target sketch patch \mathbf{y}_i , $k = 1, \dots, K, i = 1, \dots, M, j = 1, \dots, N$, \mathbf{W}^i is a sparse matrix consisting of elements of the matrix \mathbf{W} , and let $W_{[l][n]}^i$ denote the element located on the m -th row and n -th column of the matrix \mathbf{W}^i , $m, n = 1, \dots, (N + 1) \times (K + 1)$. Most elements of \mathbf{W}^i are 0 except some diagonal elements being 1 (note that $W_{[l][l]} = 0, l = K + 1, 2(K + 1), \dots, (N + 1) \times (K + 1)$) and elements of rows indexed by multiple of $K + 1$ (most of the elements in these rows are 0, except for those corresponding candidates). The proposed algorithm is summarized in Algorithm 1:

Algorithm 1: Transductive Face Sketch Synthesis

Input: Training sketch-photo pairs $(\mathbf{p}^1, \mathbf{s}^1), \dots, (\mathbf{p}^N, \mathbf{s}^N)$, test photo \mathbf{x} , K, α, β , size of a patch, size of the overlapping region, size of the search region for nearest neighbors;

Step 1: Divide all images into even patches with some overlap. For every photo patch of each training photo $\mathbf{p}^i, i = 1, \dots, N$, find its K nearest neighbors from the remaining training photos within the search region around the location of the query photo patch. K candidate sketch-photo patch pairs are obtained. A similar procedure is applied to the patches of the test photo except that the K nearest neighbors are searched in all training photos.

Step 2 : For each training sketch-photo patch pair, initialize the corresponding K weights to be $\frac{1}{K}$, and then solve an atomic quadratic programming problem using an interior point method [38] to achieve the weights. Store all the weights



Fig. 5. Examples used in experiments: the two leftmost images come from the CUHK Student database, the second two come from the AR database, the third two come from XM2VTS database, the two rightmost come from the CUHK FERET database.

which can construct the first NM rows of the weight matrix \mathbf{W} . In the following update procedure, we keep them unchanged.

Step 3 : Initialize the weights for patches of test photo \mathbf{x} to be $\frac{1}{K}$ and then obtain the initial target sketch patch as the linear combination of their corresponding K candidates weighted by $\frac{1}{K}$.

Step 4: Update the weights as in step 2 for each sketch-photo patch pairs extracted from the target sketch and test photo respectively.

Step 5: Update each target sketch patch according to (14).

Step 6: Iterate step 4 and step 5 until convergence or until the maximum iterations are reached.

Step 7: Stitch all target sketch patches into a whole sketch through the algorithm [9].

Output: The target output sketch \mathbf{y} .

III. EXPERIMENTAL RESULTS AND ANALYSIS

We conduct experiments on the CUHK face sketch database [3] and the CUHK face sketch FERET database [14]. The CUHK face sketch database consists of 606 sketches corresponding to 606 faces collected from the CUHK student database (including 188 photos), the AR database (including 123 photos) [39], and the XM2VTS database (including 295 photos) [40]. The CUHK face sketch FERET database includes 1,194 subjects from the FERET database [41]. Each subject has one sketch-photo pair. Photos in this database have lighting variations and sketches have larger shape exaggeration in comparison to sketches in the CUHK face sketch database. Examples of these databases are shown in Fig. 5.

A. Face Sketch-Photo Synthesis

In the CUHK student database, we choose 88 sketch-photo pairs for model training and the rest for testing. In the AR database, a leave-one-out likelihood strategy is adopted in which we leave 23 out. In the XM2VTS database, 100 sketch-photo pairs are collected as the training set and the rest as the test set. Some synthesized sketches and photos are shown in Fig. 6 and Fig. 7. In our experiments, parameter settings are as follows: the size of each patch is 10×10 , the size of overlap between neighboring patches is 5×5 , the search region for K nearest neighbors is 20×20 and a similar search strategy is adopted as in [3], $\alpha = 0.1$ and $\beta = 0.25$. For the method [22], the size of each patch is 20×20 with $2/3$ area overlapping (14 pixels), and the number of nearest neighbors is 5. The effect of the neighborhood size K is shown in Fig. 8. Blocking artifacts become serious with the decrease of the neighborhood size due to the incompatibility between neighboring sketches. With the

increase of neighborhood size, the synthesized sketches seem to be blurring. In all our experiments, we carefully set $K = 10$.

The most time-consuming part of our proposed method lies in the neighbor searching phase. Around the search region for some given input photo patch, we first find the best match patch from each training photos and then select top K most similar photo patches. The corresponding K sketch patches in the training set are taken as the candidates. The complexity of this process is $O(cp^2MN)$. Here c is the number of candidates in the search region around one patch, p is the patch size, M is the number of patches on each image, and N is the number of sketch-photo pairs in the training set. In our implementations, we utilized the patch matching strategy in [3] which utilized integral computation [42] and the 2D fast Fourier transform. It takes about five minutes to synthesize a sketch by our method running on a 3.1 GHz CPU computer.

Fig. 9¹ shows the comparison of the proposed method with the approach in [3]. It can be seen that the proposed method achieves clearer detail around the mouth and the eyes. As we analyzed above in Section II, any target output patch achieves its value among K states due to the MAP strategy adopted by [3]. This may result in deformation of some detail, especially around the eyes, mouth, and chin, since it is hard to find any two pairs of eyes which look exactly alike. However, this may be alleviated by a linear combination of patches of different people which expands the discrete state space to a continuous one. Thus, a new patch can be synthesized which does not exist in the training set.

Fig. 10 and Fig. 11 show the comparison between the LLE-based method [22] and Markov weighted fields method [35] for synthesized sketches and photos respectively. The LLE-based method neglected the neighboring relation and synthesized each patch independently, which might result in incompatibility phenomena. Though the method [35] imposed a constraint on the neighborhood, some noise or blurring exists around the nose, mouth, and chin. The figures below indicate that the proposed method could improve this disadvantage. From the first column of Fig. 11, we can see that the Markov weighted fields method still results in some deformation in the mouth and hair regions. There is more variability among the faces in the XM2VTS database, and thus none of these three methods achieved as good a performance as they did on the CUHK student database and AR database. However, the proposed method still obtains better results than the other two methods.

In the CUHK face sketch FERET database, we randomly

¹ All synthesized sketches using method [3] in this paper are from Dr. Wei Zhang's implementation in his paper [34].

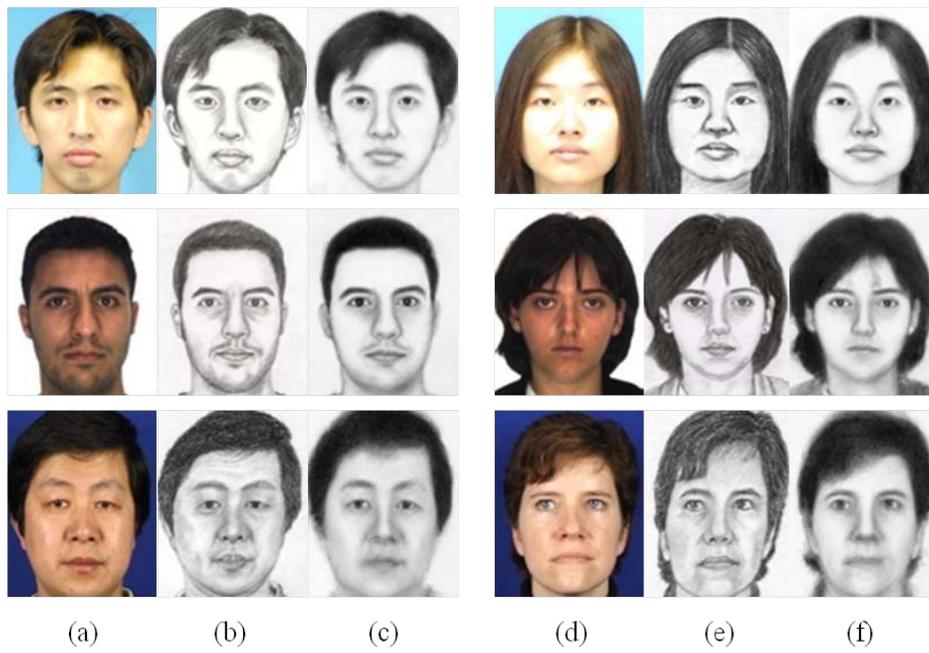


Fig. 6. Synthesized sketches from input test photos. (a), (d) input photos; (b), (e), sketches drawn by the artist; (c), (f), synthesized sketches from input photos

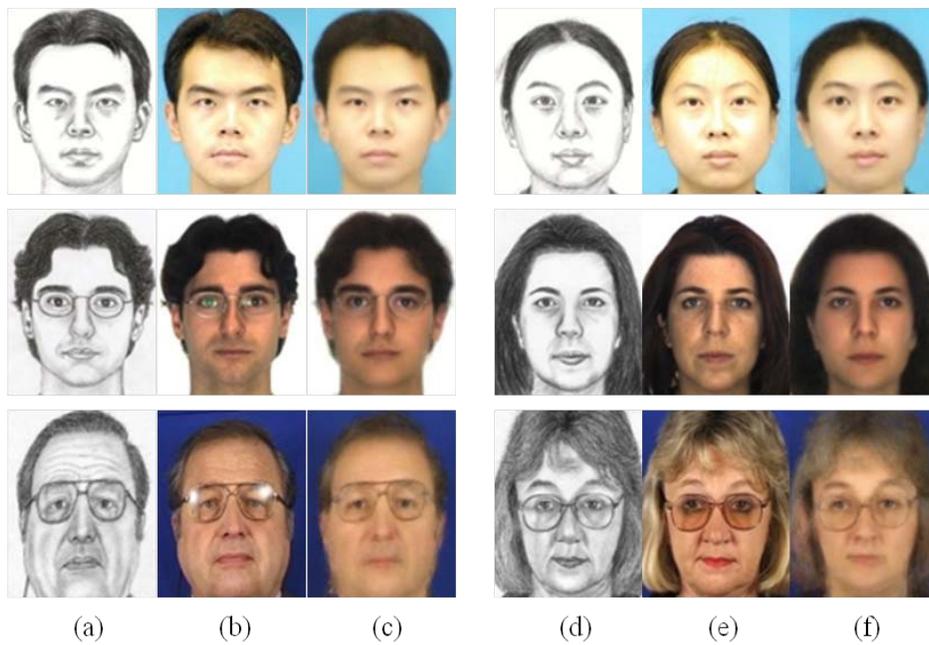


Fig. 7. Synthesized photos from input test sketches. (a), (d) input sketches; (b), (e), ground truth photo; (c), (f), synthesized photos from input sketches.



Fig. 8. The effect of the neighborhood size K .



Fig. 9. Comparison between the proposed method and MRF-based method [3] for synthesized sketches. Top row is the ground truth, second row is the result of [3], and bottom row is the result of the proposed method.

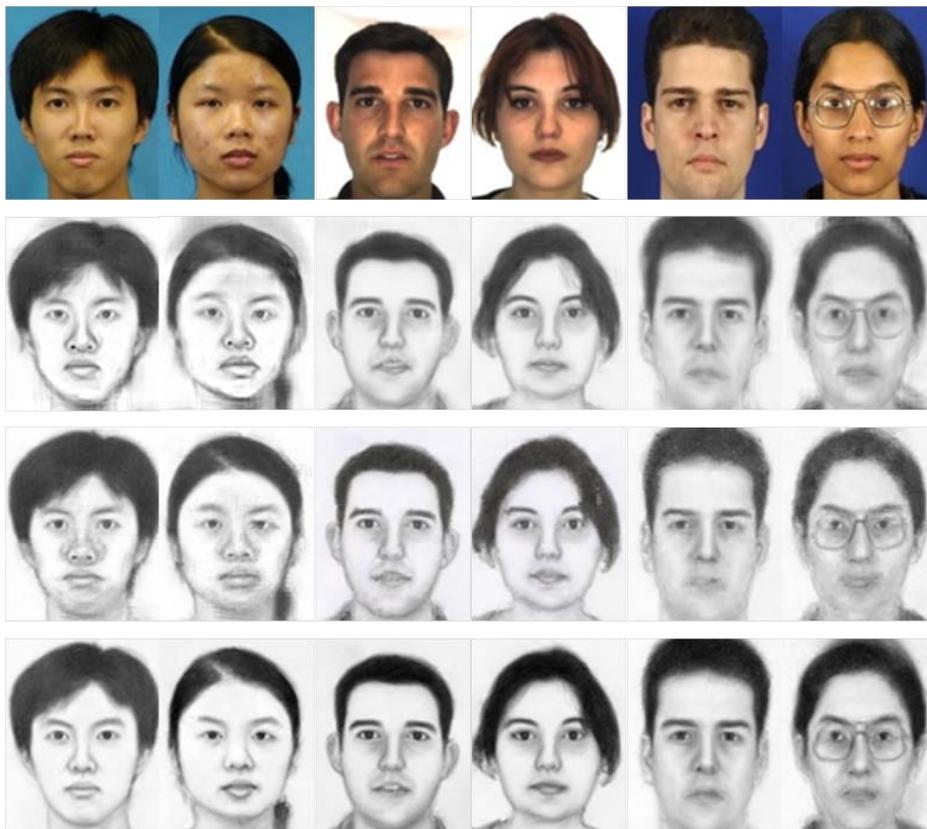


Fig. 10. Comparison of proposed method, LLE-based method [22] and the approach [35] for synthesized sketches. Top row shows input photos, the second row shows the results of method [22], the third row shows the results of method [35], and the last row is generated by the proposed method.

choose 250 sketch-photo pairs as the training set and rest 944 pairs for testing. The parameter settings are the same as the previous descriptions. Some example results are shown in Fig. 12. The proposed method obtains reasonable results on this challenge database.

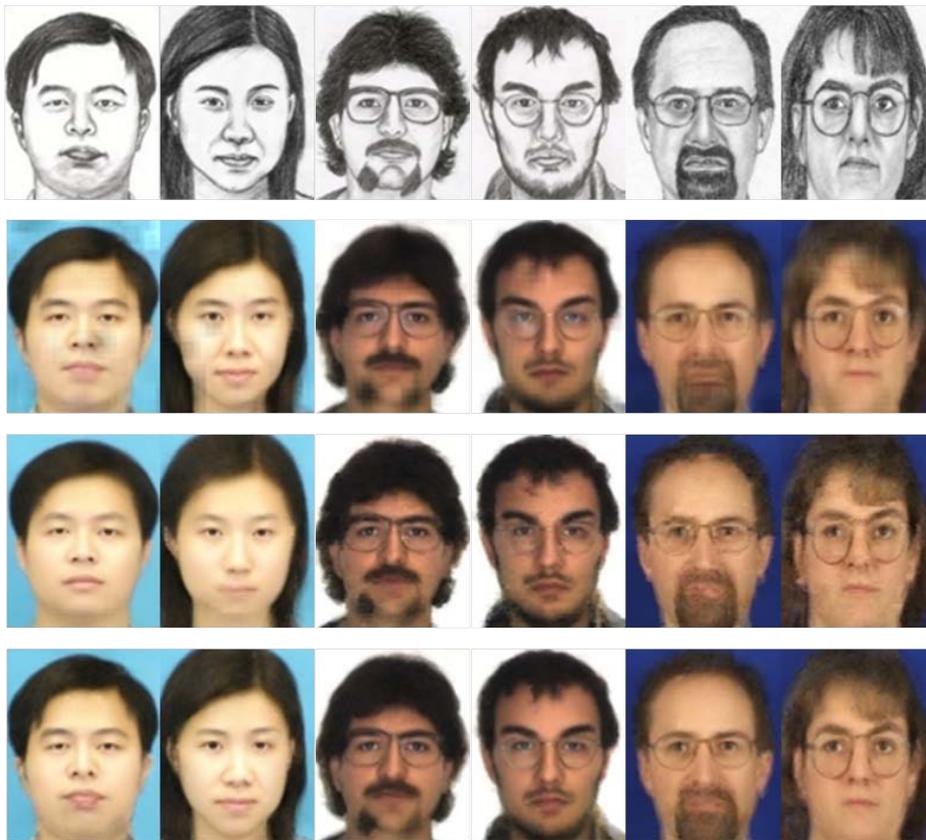


Fig. 11. Comparison of the proposed method, LLE-based method [22] and the approach [35] for synthesized photos. Top row shows input sketches, the second row shows the results of method [22], the third row shows the results of method [35], and the last row is generated by the proposed method.

B. Face Sketch Recognition

We performed face sketch recognition on the three databases together in the following two ways: (a) first transform all face photos in the gallery to sketches using our proposed face sketch synthesis method, and then a query sketch is identified from the synthesized gallery sketches; (b) first transform the query sketch to a photo exploiting the above synthesis method by switching the role of sketches and photos, and subsequently, match the synthesized photo to the gallery photo to identify the person. We apply several face recognition methods to perform face recognition: Fisherface [43], Null-space LDA (NLDA) [44], and random sampling LDA (RS-LDA) [45].

The 606 sketch-photo pairs in all three databases (the CUHK Student database, the AR database, and the XM2VTS database) are separated into three subsets. Subset 1 containing 153 pairs is used for synthesis training. Subset 2 is utilized for training the classifiers, including 153 pairs. When training the classifiers for strategy (a), photos in subset 2 are first transformed to sketches by taking subset 1 as the training set. It is similar for the strategy (b). Subset 3 consists of the remaining 300 pairs and is taken as the test set.

Table I compares the rank one recognition accuracies. Sketches/photos are synthesized by methods developed in [20], [3] and the proposed method. The aforementioned three different face recognition approaches are used for classification. In the table, “_SS” denotes synthesizing sketches, “_SP” denotes synthesizing photos, “MMRF” denotes multi-scale

MRF, and “TFSP” denotes the proposed transductive method. The table shows that the proposed transductive face sketch synthesis method combined with RS-LDA achieves the highest recognition rate 97.7%.

Table II compares the cumulative match scores of the proposed method with three other methods [3], [20] and [22]. Wang and Tang [3] reported the first match accuracy of 96.3% and the tenth match rate 99.7% using the RS-LDA face recognition approach. Our method obtains promising performance with the first match accuracy 97.7% and tenth match rate 99.7%. Though the proposed method obtained promising performance by comparing with existing face sketch-photo synthesis methods, its performance is not as good as feature-based recognition methods ([13] and [14] achieved accuracy of 99.47% and 99.87% respectively on this database). However, the main advantage of the proposed method lies in its promising synthesis performance.

We conduct some tentative experiments on the CUHK face sketch FERET database. We randomly choose 250 synthesized sketch-photo pairs to train random sampling LDA [45]. The rest 694 sketch-photos form the test set. Lighting variations and shape exaggerations bring great challenge to intensity-based face recognition methods, such as random sampling LDA method used in this paper. Also, existing face sketch-photo synthesis methods cannot properly learn the lighting variations of photos in the synthesis process. This is due to the fact that sketches drawn by artists do not carry the lighting



Fig. 12. Example results on the CUHK FERET database

TABLE I
RANK ONE RECOGNITION ACCURACY USING DIFFERENT SYNTHESIS METHODS AND FACE RECOGNITION APPROACHES (%)

	Eigentransformation [20]	MMRF_SS [3]	MMRF_SP [3]	TFSP_SP	TFSP_SS
Fisherface [43]	79.7	89.3	93.3	91.3	96.3
NLDA [44]	84.0	90.7	94.7	93.7	96.3
RS-LDA [45]	90.0	93.3	96.3	95.7	97.7

TABLE II
CUMULATIVE MATCH SCORES USING DIFFERENT FACE SKETCH RECOGNITION APPROACH (%)

	1	2	3	4	5	6	7	8	9	10
Eigentransformation [20]	90.0	94.0	96.7	97.3	97.7	97.7	98.3	98.3	99.0	99.0
Nonlinear approach [22]	87.7	92.0	95.0	97.3	97.7	98.3	98.7	99.0	99.0	99.0
MMRF_SS+RS-LDA [3]	93.3	94.6	97.3	98.3	98.3	98.3	98.3	99.0	99.0	99.0
MMRF_SP+RS-LDA [3]	96.3	97.7	98.0	98.3	98.7	98.7	99.3	99.3	99.7	99.7
TSPS_SP+RS-LDA	95.7	97.3	98.0	98.3	99.0	99.0	99.0	99.0	99.3	99.3
TSPS_SS+RS-LDA	97.7	98.0	98.3	98.7	98.7	99.0	99.0	99.7	99.7	99.7

variations. Thus, the recognition performance by synthesized photos is poor (43.7%) by only using intensity values. By contrast, we achieved an accuracy of 72.62% using sketches synthesized by the proposed method. This accuracy may be further improved by extracting comprehensive discriminative features, which exceeds the scope of this paper. Note that the proposed method is mainly for synthesis utilizations and can be applied to the entertainment driven applications, as well as face super-resolution. Zhang et al. [14] reported that MRF-based synthesis method [3] achieved an accuracy of 43.66%

at most, which is lower than the proposed method. Zhang et al.'s feature-based method [14] achieves an accuracy of 98.7% owing to the discriminative information and corresponding customized encoding strategy.

IV. CONCLUSION

In this paper, we presented a transductive face sketch-photo synthesis method. We first constructed a probabilistic framework to model the sketch-photo generation process. We then gave an alternative optimization method to solve the

proposed method, which can be further decomposed into a number of sub-problems. Experiment results show that the proposed method can achieve a better visual image quality than several other methods. Furthermore, a higher face recognition rate was reached utilizing photos synthesized by the proposed method compared to other methods. The proposed method can also be applied to other fields such as face image super-resolution, heterogeneous image transformation, facial animation in combination with some 3D face reconstruction [46] and modeling [47] techniques, and so on. In the future, we will explore the relationship between semisupervised-based methods [48] and transductive-based methods [49] to further introduce the graphical structure of faces into the model to assist the synthesis process. In addition, we will focus on the relation between the face recognition rate and the perceptual quality of synthesized images because they may not just have a simple linear relation. Moreover, we would like to explore ensemble embedding strategy [50] to fuse different features in the candidate search step to improve the robustness of the proposed method.

ACKNOWLEDGEMENT

The authors would like to thank the Editor-In-Chief Prof. D. Liu, the handling associate editor and all three anonymous reviewers for their constructive and helpful comments and suggestions. We would also like to thank Mr. Hao Zhou (The University of Hong Kong) for providing us with his results of reference [35], Prof. Xiaogang Wang (The Chinese University of Hong Kong) for sharing the CUHK sketch database with us (website: <http://mmlab.ie.cuhk.edu.hk/facesketch.html>) and Mr. Wei Zhang (The Chinese University of Hong Kong) for supplying us with his results of reference [3].

REFERENCES

- [1] S. Iwashita, Y. Takeda, and T. Onisawa, "Expressive face caricature drawing," in *Proceedings of IEEE International Conference on Fuzzy Systems*, 1999, pp. 1597–1602.
- [2] H. Koshimizu and M. Tominaga, "On kanse face processing for computerized face caricaturing system picasso," in *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, 1999, pp. 294–299.
- [3] X. Wang and X. Tang, "Face photo-sketch synthesis and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 1955–1967, 2009.
- [4] W. Zhang, X. Wang, and X. Tang, "Lighting and pose robust face sketch synthesis," in *Proceedings of European Conference on Computer Vision*, 2010, pp. 420–423.
- [5] X. Gao, J. Zhong, J. Li, and C. Tian, "Face sketch synthesis using e-hmm and selective ensemble," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 4, pp. 487–496, 2008.
- [6] X. Tang and X. Wang, "Face sketch recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 1–7, 2004.
- [7] S. Baker and T. Kanade, "Hallucinating faces," in *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, 2000, pp. 83–88.
- [8] C. Liu, H. Shum, and W. Freeman, "Face hallucination: theory and practice," *International Journal of Computer Vision*, vol. 1, no. 1, pp. 115–134, 2007.
- [9] A. Efros and W. Freeman, "Image quilting for texture synthesis and transfer," in *Proceedings of SIGGRAPH*, 2001, pp. 341–346.
- [10] A. Hertzmann, C. Jacobs, N. Oliver, B. Curless, and D. Salesin, "Image analogies," in *Proceedings of SIGGRAPH*, 2001, pp. 327–340.
- [11] D. Lin and X. Tang, "Coupled space learning of image style transformation," in *Proceedings of IEEE International Conference on Computer Vision*, 2005, pp. 1699–1706.
- [12] J. Chen, D. Yi, J. Yang, and G. Zhao, "Learning mappings for face synthesis from near infrared to visual light images," in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, 2009, pp. 156–163.
- [13] B. Klare, Z. Li, and A. Jain, "Matching forensic sketches to mug shot photos," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, pp. 639–646, 2011.
- [14] W. Zhang, X. Wang, and X. Tang, "Coupled information-theoretic encoding for face photo-sketch recognition," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 513–520.
- [15] M. Zhao and S. Zhu, "Portrait painting using active templates," in *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Non-Photorealistic Animation and Rendering*, 2011, pp. 117–124.
- [16] N. Wang, J. Li, D. Tao, X. Li, and X. Gao, "Heterogeneous image transformation," *Pattern Recognition Letters*, vol. 34, no. 1, pp. 77–84, 2013.
- [17] I. Jolliffe, *Principal component analysis*. New York: Springer, 2002.
- [18] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [19] X. Tang and X. Wang, "Face photo recognition using sketch," in *Proceedings of IEEE International Conference on Image Processing*, 2002, pp. 257–260.
- [20] —, "Face sketch synthesis and recognition," in *Proceedings of IEEE International Conference on Computer Vision*, 2003, pp. 687–694.
- [21] Y. Li, M. Savvides, and V. Bhagavatula, "Illumination tolerance face recognition using a novel face from sketch synthesis approach and advanced correlation filters," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2006, pp. 357–360.
- [22] Q. Liu, X. Tang, H. Jin, H. Lu, and S. Ma, "A nonlinear approach for face sketch synthesis and recognition," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2005, pp. 1005–1010.
- [23] W. Liu, X. Tang, and J. Liu, "Bayesian tensor inference for sketch-based face photo hallucination," in *Proceedings of International Joint Conference on Artificial Intelligence*, 2007, pp. 2141–2146.
- [24] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [25] X. Gao, K. Zhang, D. Tao, and X. Li, "Image super-resolution with sparse neighbor embedding," *IEEE Transactions on Image Processing*, vol. 21, no. 7, pp. 3194–3205, 2012.
- [26] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1031–1044, 2010.
- [27] M. Chang, L. Zhou, Y. Han, and X. Deng, "Face sketch synthesis via sparse representation," in *Proceedings of International Conference on Pattern Recognition*, 2010, pp. 2146–2149.
- [28] H. Lee, A. Battle, R. Raina, and A. Ng, "Efficient sparse coding algorithms," in *Proceedings of Advances in Neural Information Processing Systems*, 2007, pp. 801–808.
- [29] X. Gao, N. Wang, D. Tao, and X. Li, "Face sketch-photo synthesis and retrieval using sparse representation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 8, pp. 1213–1226, 2012.
- [30] N. Ji, X. Chai, S. Shan, and X. Chen, "Local regression model for automatic face sketch generation," in *Proceedings of International Conference on Image and Graphics*, 2011, pp. 412–417.
- [31] S. Wang, L. Zhang, Y. Liang, and Q. Pan, "Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2216–2223.
- [32] X. Gao, J. Zhong, D. Tao, and X. Li, "Local face sketch synthesis learning," *Neurocomputing*, vol. 71, no. 10–12, pp. 1921–1930, 2008.
- [33] B. Xiao, X. Gao, D. Tao, Y. Yuan, and J. Li, "Photo-sketch synthesis and recognition based on subspace learning," *Neurocomputing*, vol. 73, no. 4–6, pp. 840–852, 2010.
- [34] B. Xiao, X. Gao, D. Tao, and X. Li, "A new approach for face recognition by sketches in photos," *Signal Processing*, vol. 89, no. 8, pp. 1576–1588, 2009.
- [35] H. Zhou, Z. Kuang, and K. Wong, "Markov weight fields for face sketch synthesis," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 1091–1097.

- [36] A. Schwaighofer and V. Tresp, "Transductive and inductive methods for approximate gaussian process regression," in *Proceedings of Advances in Neural Information Processing Systems*, 2003, pp. 953–960.
- [37] F. Perez-Cruz, A. Navia-Vazquez, A. Figueiras-Vidal, and A. Artes-Rodriguez, "Empirical risk minimization for support vector classifiers," *IEEE Transactions on Neural Networks*, vol. 14, no. 2, pp. 296–303, 2003.
- [38] R. Vanderbei, "Loqo: an interior point code for quadratic programming," *Optimization Methods and Software*, vol. 11, no. 1, pp. 451–484, 1999.
- [39] A. Martinez and R. Benavente, "The AR face database," CVC Technical Report #24, Tech. Rep., 1998.
- [40] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, "Xm2vtsdb: the extended m2vts database," in *Proceedings of International Conference on Audio- and Video-Based Biometric Person Authentication*, 1999, pp. 72–77.
- [41] P. Phillips, H. Moon, P. Rauss, and S. Rizvi, "The FERET evaluation methodology for face recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.
- [42] P. Viola and M. Jones, "Robust real-time object detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.
- [43] P. Belhumeur, J. Hespanha, and D. Kriegeman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.
- [44] L. Chen, H. Liao, M. Ko, J. Lin, and G. Yu, "A new lda-based face recognition system which can solve the small sample size problem," *Pattern Recognition*, vol. 33, no. 10, pp. 1713–1726, 2000.
- [45] X. Wang and X. Tang, "Random sampling for subspace face recognition," *International Journal of Computer Vision*, vol. 70, no. 1, pp. 91–104, 2006.
- [46] M. Song, D. Tao, X. Huang, C. Chen, and J. Bu, "Three-dimensional face reconstruction from a single image by a coupled RBF network," *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2887–2897, 2012.
- [47] D. Tao, M. Song, X. Li, J. Shen, J. Sun, X. Wu, C. Faloutsos, and S. Maybank, "Bayesian tensor approach for 3-d face modeling," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 10, pp. 1397–1410, 2008.
- [48] M. Adankon, M. Cheriet, and A. Biem, "Semisupervised least squares support vector machine," *IEEE Transactions on Neural Networks*, vol. 20, no. 12, pp. 1858–1870, 2009.
- [49] T. Kato, H. Kashima, and M. Sugiyama, "Robust label propagation on multiple networks," *IEEE Transactions on Neural Networks*, vol. 20, no. 1, pp. 35–44, 2009.
- [50] B. Geng, D. Tao, C. Xu, Y. Yang, and X. Hua, "Ensemble manifold regularization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 6, pp. 1227–1233, 2012.



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 Centre 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 and co-authored more than 100 scientific articles at top venues including IEEE T-PAMI, T-KDE, T-NN, T-IP, NIPS, ICML, AISTATS, ICDM, IJCAI, AAAI, CVPR, ECCV; ACM T-KDD, and KDD. Dr. Tao has received many Best Paper Awards.



Xinbo Gao (M'02-SM'07) received the B.Eng., M.Sc. and Ph.D. degrees in signal and information processing from Xidian University, China, in 1994, 1997 and 1999 respectively. From 1997 to 1998, he was a research fellow with the Department of Computer Science, Shizuoka University, Japan. From 2000 to 2001, he was a postdoctoral 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 System, and Director of the VIPs Laboratory, Xidian University. His current research interests are 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 IEEE T-IP, T-CSVT, T-NN, T-SMC, Pattern Recognition and so on. Dr. Gao is on the editorial boards of journals including EURASIP Signal Processing (Elsevier), and Neurocomputing (Elsevier). He served as General Chair/Co-Chair or Program Committee Chair/Co-Chair or a PC member for around 30 major international conferences. Now, he is a Fellow of IET.

Xuelong Li (M'02-SM'07-F'12) is a full professor with the Center for OPTical IMagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, P. R. China.



Nannan Wang received the B.Sc. degree in information and computation science from Xi'an University of Posts and Telecommunications in 2009. Now 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 mainly focus on computer vision specially for heterogeneous image transformation and its applications.



Jie Li received the B.Sc. degree in electronic engineering, the M.Sc. degree in signal and information processing, and the Ph.D. degree in circuit and systems, from Xidian University, Xi'an, China, in 1995, 1998, and 2004, respectively. She is currently a Professor in the school of electronic engineering, Xidian University, China. Her research interests include video processing and machine learning.