Neurocomputing 257 (2017) 214–221

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Data-driven vs. model-driven: Fast face sketch synthesis

Nannan Wang[a,*], Mingrui Zhu[b], Jie Li[b], Bin Song[a], Zan Li[a]

[a] The State Key Laboratory of Integrated Services Networks, School of Telecommunications Engineering, Xidian University, Xi'an city 710071, P.R. China
[b] School of Electronics Engineering, Xidian University, Xi'an City 710071, P.R. China

ARTICLE INFO

Article history:
Received 26 May 2016
Revised 20 June 2016
Accepted 11 July 2016
Available online 4 February 2017

Keywords:
Face sketch synthesis
Model-driven
Data-driven
Face recognition
Image quality assessment

ABSTRACT

Face sketch synthesis refers to the technique generating a sketch from an input photo. Existing methods are data-driven, which synthesize a sketch by linearly combining K candidate sketch patches which are purposely selected from the training data. However, these methods have large computation cost due to neighbor selection process that perform neighbor searching on a large scale of training image patches. Instead of the aforementioned commonly used data-driven strategy, we propose to learn some models from training photos to training sketches which could speed the synthesis process a lot while preserving comparable or even better synthesis performance. Specially, we learn some ridge regressors from training photo patch intensities to training sketch patch intensities. An initial estimation is obtained from these regressors. Simultaneously, a high-frequency image is hallucinated from some ridge regressors which are learned from the high-frequency information of training photos and sketches. The high-frequency image is superimposed to the initial estimation to compensate the filtered details due to the dense average in the initial estimation process. Extensive experiments on public face sketch database illustrate the effectiveness of proposed model-driven strategy.

1. Introduction

Face sketch synthesis refers to generate a sketch from an input photo. It has been applied in digital entertainment (*e.g.* face animation) and law enforcement (sketch based face recognition) [1]. Existing face sketch synthesis methods are data-driven, whose computation complexity is linearly increasing with the scale of training data.

Data-driven methods generally work on the patch level (see Fig. 1) and these methods consists of two parts: neighbor selection model and neighbor fusion model (also known as weight computation model). For each test patch, K nearest photo patches are selected from the training data and simultaneously K sketch patches corresponding to selected photo patches are taken as the candidate sketch patches. Given these candidate sketch patches, they are usually linearly combined to obtain the final sketch patch. Eventually, these synthesized sketches are fused into a whole sketch.

Data-driven face sketch synthesis starts from the work of Tang and Wang [2–4]. Principle component analysis is employed to project the test photo onto the training photos and the projected coefficients are then utilized to reconstruct the target sketch by linearly combining training sketches. Considering that the whole linear assumption on the mapping from photo to sketch cannot always hold in the seminal work [2], Liu et al. [5] proposed to estimate the nonlinear mapping from photo to sketch in a piece-wise manner inspired from the idea of locally linear embedding (LLE) [6]. For each test patch, K-nearest neighbors (K-NN) is conducted to search K nearest photo patches. The corresponding K sketch patches are linearly combined to synthesize the target sketch patch weighted by coefficients learned from a linear least square. Song et al. [7] took the similar piece-wise assumption by casting the face sketch synthesis problem as an image denoising problem. Instead of fixed number of nearest neighbors for all test patches, Gao et al. [8,9] proposed to adaptively determine the number of neighbors by exploring sparse representation. In view of the fact that aforementioned methods neglect the neighboring constraint between adjacent sketch patches during the neighbor selection stage, Wang and Tang [10] deployed Markov random fields (MRF) to describe the dependency between adjacent patches. However, since the MRF-based method [10] adopts the maximum a posterior formulation for neighbor selection, only one "best" candidate sketch patch is selected as the final target sketch patch. This results in deformations around some face details such as mouth, eye and nose. Zhou et al. [11] proposed to introduce the linear combination of K nearest neighbors to the Markov networks and the new network is so called Markov weight fields (MWF).

The most time consuming part for data-driven face sketch synthesis methods lies in the neighbor selection model. It usually

* Corresponding author.
E-mail addresses: nnwang@xidian.edu.cn, nannanwang.xidian@gmail.com (N. Wang).

http://dx.doi.org/10.1016/j.neucom.2016.07.071
0925-2312/

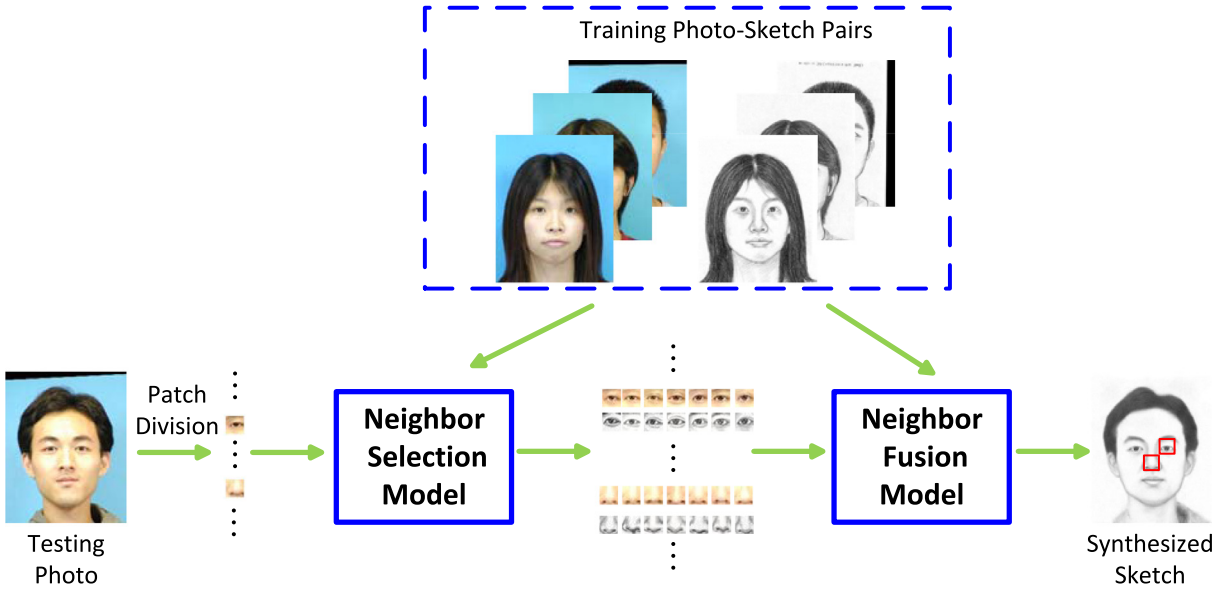

Fig. 1. Graphical outline for data-driven face sketch synthesis methods.

needs to traverse the whole training data. In this paper, we proposed a model-driven face sketch synthesis framework that traversal is not needed. Inspired from the idea of piece-wise learning aforementioned [5,7], we propose to divide the space spanned by images into some subspaces spanned by image patches. Simple linear regressors are learned in these subspaces. Then in the testing, for each test patch, only the nearest subspace it located is used to synthesize the target sketch patch. Since the main computation for test is some multiplications, the computation cost is heavily reduced.

The contributions of this paper are threefold as follows.

1. We proposed a model-driven face sketch synthesis framework; To the best of our knowledge, this is the first model-driven face sketch synthesis method.
2. Within the proposed model-driven face sketch synthesis framework, we further propose a simple yet effective face sketch synthesis method based on ridge regression.
3. By comparing the proposed method with state-of-the-art methods, the proposed method could heavily reduce the computational cost while preserving comparable performance of even better performance.

Excepted as noted, we utilize a bold lowercase letter to represent a column vector, a bold uppercase letter to stand for a matrix, and regular lowercase and uppercase letters to denote a scalar. The rest of this paper is organized as follows. Section 2 introduces the proposed model-driven face sketch synthesis method. Section 3 provides the experimental results and some analysis. Finally, some concluding remarks are given in the section 4.

2. Model-driven face sketch synthesis

Data-driven methods synthesize each target sketch patch by searching all training images around the location as the test patch. This process limits the efficiency for synthesis. In our proposed model-driven framework, instead, we learn a fixed mapping from photo patches to sketch patches locally as shown in Fig. 2.

In the training stage, all training image pairs are divided into patches size of $p \times p$. These patches are then grouped into C clusters according to their locations. In each cluster, suppose the matrix of training sketch patches to be $\mathbf{Y} \in \mathbb{R}^{(d \times n)}$ and the matrix of training photo patches to be $\mathbf{X} \in \mathbb{R}^{(d \times n)}$, where $d = p^2$, and n is the number of training photo-sketch patch pairs. Specially, we employ ridge regressor as the mapping function to learn the transformation from photo patches $\mathbf{X}$ to sketch patches $\mathbf{Y}$. It can be formulated as

$$\mathbf{Y} = \mathbf{\Theta}\mathbf{X} + \mathbf{\Gamma}, \tag{1}$$

where $\mathbf{\Theta}_{(m \times p)}$ is the coefficients matrix, and $\mathbf{\Gamma}$ is the residual error matrix.

To overcome overfitting and seek for the unique solution to the coefficients matrix $\mathbf{\Theta}$, the regularization on the coefficient matrix should be imposed:

$$\min_{\mathbf{\Theta}} ||\mathbf{\Theta}\mathbf{X} - \mathbf{Y}||_F^2 + \lambda ||\mathbf{\Theta}||_F^2, \tag{2}$$

where λ is the regularization parameter. Actually, manifold regularization techniques could also be employed here such as Laplacian regularization and p-Laplacian regularization [12] since manifold regularization techniques take the manifold structure of samples into consideration. In our implementations, we utilize the ridge regression embedded in the scikit-learn package of python. The regularization parameter in our python implementations is empirically set to 10^6.

Given a test photo $\mathbf{t}$, it is first divided into N patches $\mathbf{t}_1, \ldots, \mathbf{t}_N$ and each patch is size of $p \times p$. In the testing phase, each patch is indexed to a subspace and the learned ridge regressor on this subspace is utilized to transform the test patch to a target sketch patch. Given the test photo patch $\mathbf{t}_i$, $1 \le i \le N$, the corresponding target sketch patch $\mathbf{s}_i$ can be estimated as

$$\mathbf{s}_i = \mathbf{\Theta}\mathbf{t}^i. \tag{3}$$

After all target sketch patches generated, they are fused to a whole sketch with overlapping region averaged.

As in [9] illustrated, the linear combination process in Eq. (3) acts like a high-frequency filter and some detail information may be lost in this synthesis process. This would result in some over-smooth or blurring effect. Zhang et al. [13] proposed a support vector regression (SVR) method for further improving the quality of the initial estimate by synthesize the lost high frequency information. But they adopted different strategy to synthesize the initial estimate and the high frequency information, which makes the problem complicate and time-consuming. We propose to synthesize the high frequency information using the same strat-