Signal Processing 130 (2017) 1–11

Contents lists available at ScienceDirect

Signal Processing

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

Unified framework for face sketch synthesis

Nannan Wang [a], Shengchuan Zhang [b], Xinbo Gao [b,1,*], Jie Li [b], Bin Song [a], Zan Li [a]

[a] State Key Laboratory of Integrated Services Networks, School of Telecommunications Engineer, Xidian University, Xi'an 710071, China
[b] School of Electronic Engineering, Xidian University, Xi'an 710071, China

ARTICLE INFO

Article history:
Received 19 November 2015
Received in revised form
23 March 2016
Accepted 13 June 2016
Available online 16 June 2016

Keywords:
Face sketch synthesis
Position constraint
Global search
High frequency
Markov network

ABSTRACT

Face sketch synthesis (FSS) has great significance to sketch based face retrieval or recognition and digital entertainment. Recently, great progress has been made in face sketch synthesis. Most state-of-the-art FSS methods work at patch level. However, these methods only consider either position constraint or global search when selecting candidate image patches. Furthermore, the common used weighting combination in these methods leads to the lost of reasonable high frequency details. We argue that all these factors are necessary for face sketch synthesis. To this end, we propose a simple yet effective unified approach considering position constraint, global search and high frequency compensation to infer pseudosketches from input test photos. Firstly, a nearest-neighbor search is conducted by combining both position constraint and global search. After obtaining the candidate image patches, a Markov network is applied to generate the pseudosketch. Secondly, the residue between an original sketch and the synthesized pseudosketch is modeled by the same Markov network to compensate the high frequency details. The effectiveness of the proposed method is demonstrated on the CUHK face sketch database by comparing with state-of-the-art FSS methods.

1. Introduction

As an important branch of face image style transformation [1–7], face sketch synthesis (FSS) has attracted much more attention in recent years. This problem was first introduced by Tang and Wang [8] to assist the police in building sketch-based photo retrieval system. For law enforcement, the police need to automatically match a sketch of a suspect drawn by an artist against with the mug-shot photo database. However, the great discrepancies between photos and sketches make straightforward matching using generic face retrieval algorithms [9–14] difficult. Therefore, FSS techniques transform photos and sketches into the same modality to facilitate retrieval.

Besides, Face sketch synthesis has many applications in digital entertainment [15]. Some people prefer to use face sketches as their profile pictures in social media. Novices can employ FSS technique to create enjoyable face sketches without much artistic training. Furthermore, face sketch synthesis assists artists to simplify the animation procedure.

We argue that a successful FSS algorithm should meet the following three requirements:

(1) *Style requirement*: the synthesized sketches must be close to the ones drawn by artists.
(2) *Identity requirement*: the synthesized sketches can still be identified as the original persons.
(3) *Quality requirement*: the synthesized sketches should have high visual quality (less over-smoothing and noisy).

The first requirement can be met through data-driven methods. Particularly, since we work at patch-level, we apply position constraint to search candidate sketch patches to meet style requirement. For the second requirement, we select K nearest candidate sketch patches to interpolate the new target sketch patch which may be nonexistent in the training sketch patches. The third requirement can be satisfied through two aspects: intact structure and high definition, which can be realized by global search and high frequency compensation respectively. In other words, style requirement is composed of position constraint. Identity requirement is made up of constraint on using several candidate sketch patches to synthesize the new sketch patch. Quality requirement consists of global search constraint and high frequency compensation constraint.

It is of great importance to fulfill aforementioned three requirements to generate satisfying face sketches. Concretely speaking, without the position constraint, two local patches which have similar appearance in photos may possess very different corresponding sketch patches drawn by the artist [7]. Without constraint on using several candidate sketch patches to synthesize

* Corresponding author.
E-mail address: xbgao@mail.xidian.edu.cn (X. Gao).
[1] Member of Editorial board for EURASIP Signal Processing.

http://dx.doi.org/10.1016/j.sigpro.2016.06.014
0165-1684/

the new sketch patch, it sometimes cannot produce good results for features like eyes and mouths varied over different individuals [16]. Without the global search constraint, the results may lack integrated structure when the test photo owns some unique non-facial factors [17]. Without constraint on the high frequency compensation, the results may be too smooth.

These constraints motivate us to design a unified approach which incorporates all the constraints in a graphical, compositional face model. The graphical model represents a face in a Markov network while the compositional model represents a face with two parts: low resolution portion and high resolution portion. We separately apply the position constraint and the global search constraint to select K candidate photo-sketch patch pairs. Then we simply utilize the Euclidean distance between intensities of test photo patch and $2K$ candidate photo patches to reselect K candidate photo-sketch patch pairs from obtained $2K$ candidate photo-sketch patch pairs. A Markov network with constraint on using K candidate sketch patches to synthesize the new sketch patch is employed to obtain an initial estimate for the target sketch. Taking high frequency compensation constraint into consideration, an optimal residue image which is the difference between the target sketch and the initial estimate is inferred from the test photo by minimizing the energy of the same Markov network. The final synthesized sketch is obtained by the sum of the initial estimate and the residue image. The flowchart of the unified approach is illustrated in Fig. 1.

The contributions of this paper are threefold. (1) It designs a unified approach for face sketch synthesis; (2) It proposes a data-driven based high frequency information compensation strategy for face sketch synthesis; (3) It outperforms state-of-the-art FSS methods.

2. Previous work

Tang and Wang [8] were the first to undertake the problem with principle component analysis (PCA) method [18]. Based on PCA, an eigensketch transformation algorithm was developed to address face sketch synthesis. They assumed that the photo-sketch mapping can be approximated as a linear process. However, it is more reasonable to approximate the sketch drawing process of an artist by a nonlinear process.

Liu et al. [19] presented a sketch generation method, which is based on local linear preserving of geometry between photo and sketch image patches inspired by the idea of locally linear embedding (LLE) [20] in manifold learning. If linear patch combination in [19] is conducted densely, the synthesized sketch contains significant amount of noise. Song et al. [21] proposed a simple yet effective FSS method called spatial sketch denoising (SSD) which takes face sketch synthesis as an image denoising problem. One defect of the above two methods is that they synthesize each local patch independently at a fixed scale and cannot learn face structures in a large scale well.

Wang and Tang [7] developed a new approach to synthesize face sketch using a multiscale maximum a posterior (MAP) based Markov random fields (MRF) model [22] to reduce the drawback of the method proposed in [19]. They then extended this method into a robust algorithm [23] which integrated shape prior, photo-to-photo patch matching, photo-to-sketch patch matching, intensity compatibility, and gradient compatibility and worked well under lighting and pose variations. However, both the MAP-MRF model and the robust algorithm only select a sketch patch from the training set as the final target output and it is NP-hard for the MRF optimization.

In [16,24], both Zhou et al. and Wang et al. applied K nearest neighbors as the candidates to synthesize the final target sketch patch that does not appear in the training set. Moreover, the approach proposed by Wang et al. integrated the given test samples into the learning process to reduce the expected loss and improve the synthesis performance.

Zhou's work [16] selected candidate patches using local search range in the implementation to save computational cost. Wang et al. [7] set a search range around the same position of the query photo patch since the same face components on different images may locate roughly at the similar position. In contrast to Zhou's and Wang's methods, our previous work [17] designed a sparse representation based greedy search (SRGS) strategy to select candidate patches from the whole training set specially which is robust to image backgrounds.

Since the number of nearest neighbors is fixed in most existing methods, which might introduce some deformation and noise into the synthesized sketches, Gao et al. [25] and Wang et al. [26]

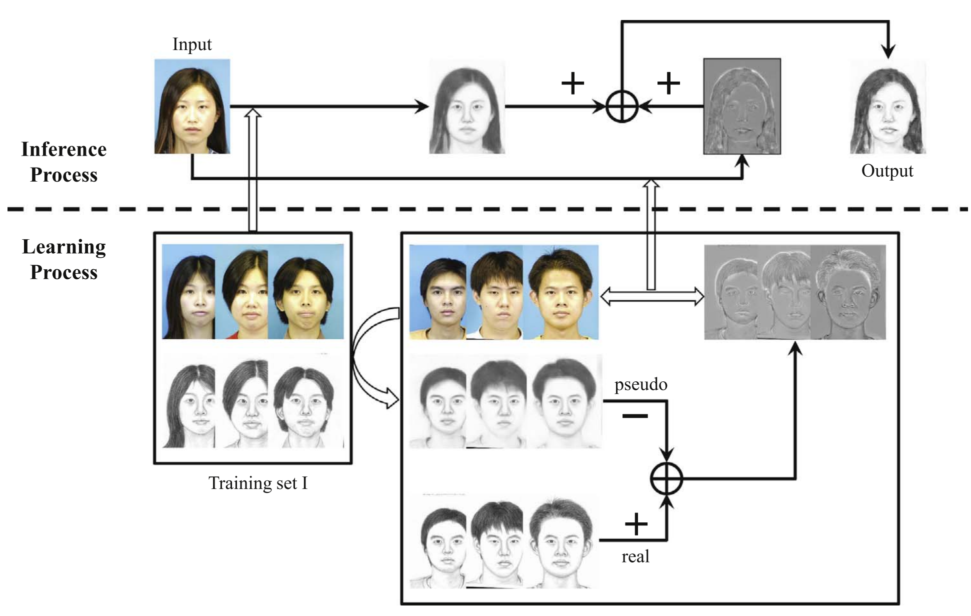

Fig. 1. Framework of the proposed unified approach for face sketch synthesis.