Neurocomputing 212 (2016) 107–120

Contents lists available at ScienceDirect

Neurocomputing

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

Structured detail enhancement for cross-modality face synthesis

Chunwei Song, Feng Li, Yunqi Dang, Huijun Gao, Zifei Yan, Wangmeng Zuo *

Harbin Institute of Technology, 92, West Dazhi Street, 150001 Harbin, China

ARTICLE INFO

Article history:
Received 20 November 2015
Received in revised form
23 January 2016
Accepted 31 January 2016
Available online 28 June 2016

Keywords:
Face sketch synthesis
K-nearest neighbors
Face landmark detection
Augmented Lagrangian method
L_p-norm minimization

ABSTRACT

Transformation, matching, and retrieval of face images in different modalities are often encountered in many applications. Cross-modality face synthesis, which aims to transform face images between modalities, can serve as an intermediate representation for these tasks, and remains an active yet challenging topic in low level vision. Most existing cross-modality face synthesis methods are effective in recovering global features for a given photo, but fail to capture fine-scale details or introduce some extra noise in the synthesis results. In this paper, we propose a two-step algorithm to tackle this problem by enhancing structured details after initial synthesis. In the first step, a combination of quasi-smooth area synthesis and facial component synthesis is proposed to generate more plausible facial components. In the second step, an MAP-based model is constructed to introduce more facial details into the initial synthesis, meanwhile eliminate noise. Experiments on face sketch synthesis and near-infrared face synthesis demonstrate the effectiveness of the proposed method in enhancing fine-scale facial details while reducing noise.

1. Introduction

Cross-modality face synthesis, which aims to generate a synthetic face image in target modality (e.g., near-infrared image) based on a reference image in source modality (e.g., visible image), is an active yet challenging topic for face image manipulation. Cross-modality face synthesis can serve as an intermediate representation for the transformation, matching, and retrieval of face images among different modalities, and has been widely applied to many fields, such as cross-media analysis, law enforcement, and digital entertainment.

Face sketch synthesis and near-infrared (NIR) face synthesis are two typical examples of cross-modality face synthesis. In many forensic scenarios, it is difficult to obtain the high quality facial photo of suspect, while sketch of the suspect can still be drawn by artists according to the description of witnesses. Therefore, face sketch synthesis can provide a possible solution for narrowing down the potential searching group [1–3]. In many surveillance scenarios, face can be captured as either near-infrared (NIR) image or visible image, and thus NIR face synthesis can also contribute to robust face matching and analysis [4,5]. Besides, cross-modality face synthesis can also be used in many digital entertainment applications such as producing cartoon like pictures or movies, or making personal avatars [6–8].

By far, a number of approaches have been proposed for cross-modality face synthesis, especially face sketch synthesis. Fig. 1 shows an example of two face photos with the corresponding sketches. One intrinsic difficulty is that the cross-modality facial images generally are located in different geometric space and there is no explicit mechanism to connect the spaces of different modalities of images. Statistical learning methods [9,10] have been first developed for face sketch synthesis, and subsequently eigen-transformation-based approaches [2,11,12] have been suggested. Unfortunately, all these methods assume that the cross-modality transformation is linear, limiting its applications to high quality synthesis. Therefore, several nonlinear transformation methods have been proposed, e.g., the local geometry preserving method [13] and the Markov random field (MRF) based methods [14–16]. Recently, motivated by the great success of sparse representation and dictionary learning [17,18], several approaches [19–21] have been presented by exploiting the representation coefficients to bridge the gap between the representations across modalities. However, even many linear and nonlinear methods have been suggested to model the complex relationship, as shown in Fig. 2, the existing approaches generally tend to produce over-smooth results or to introduce noise in the synthesis results. Moreover, another limitation of the existing face sketch

* Corresponding author.
E-mail addresses: wolvesandme@gmail.com (C. Song), fenglihit@hotmail.com (F. Li), dyqhitcs@126.com (Y. Dang), hjgao@hit.edu.cn (H. Gao), cszfyan@gmail.com (Z. Yan), cswmzuo@gmail.com (W. Zuo).

http://dx.doi.org/10.1016/j.neucom.2016.01.119
0925-2312/

synthesis methods is they generally perform much poorer in generating shading effect than sketches by artist.

In this paper, we propose a two-step scheme (i.e., initial synthesis construction and detail enhancement optimization) for detail-enhanced cross-modality face synthesis. In comparison to our earlier version of this work [23], we modify both the first and the second steps to improve the synthesis result. In the first step, instead of KNN-based synthesis, a combination of quasi-smooth area synthesis and facial component synthesis is proposed to generate more plausible facial components. In the second step, instead of guided filtering [24], an MAP-based model is constructed to introduce more facial details into the initial synthesis result. A face sketch synthesis result using our method is given in Fig. 2(e), where one can find that the result of our method is more visually pleasant.

Our contributions are of three-fold. First, facial component synthesis is considered separately. Most of the previous methods take the facial component and other areas indiscriminately, and one side-effect of this practice is that facial components are usually over-smoothed. Since facial components play a key role in face recognition and other face analysis tasks, it is crucial to process facial components separately. Second, due to the adoption of component-based synthesis, the synthesized facial components are more similar to the ground truth. Different from previous patch-based methods, our method takes each facial component (such as eyes) as a whole part. By searching for the most similar component in training set, the synthesized facial components tend to have more details, as shown in the experiment section. Last but not least, a newly proposed MAP-based model can further transfer facial details from test image to the final synthesis. Most of the previous methods ignored the synthesis of hair or other small scale facial details, while in our method a simple but effective optimization framework is proposed to better preserve them, as shown in Fig. 2 and other results in the experiment section.

The remainder of this paper is organized as follows. In Section 2, we review the related work on cross-modality face synthesis, especially face sketch synthesis. Section 3 describes the algorithm and Section 4 presents our experimental results. Finally, the whole paper is concluded in Section 5.

Fig. 1. Face photos and the corresponding sketches. Photos are in the left column, and the corresponding sketches are in the right column.

2. Related work

In this section, we briefly review the related work on cross-modality face synthesis, especially face sketch synthesis. Eigen-transformation [2,11,12,25] is one representative category of methods based on the assumption that photo and sketch patches lie on low-dimensional principal component analysis (PCA) space. Eigen-transformation further requires that the representation coefficients between the photo and the sketch patches should be equal, which indicates that the transformation from photo to sketch is linear. Apparently, this linearity assumption is too strong to be satisfied. To relax this constraint, Lin et al. [26] introduced a coupled transformation matrix among the photo and sketch representation coefficients, and Liu et al. [13] adopted a nonlinear generation method based on local linear preservation of geometry between photo and sketch patches.

Another representative category of methods is based on MRF [27], which can be used to enforce the smoothness among the neighbored patches of the synthesized sketch. In [14], Wang et al. proposed a multi-scale MRF model for face sketch synthesis and recognition, where local patches in different regions and scales can be jointly learned. Zhang et al. [15] further extended the work in [14] by introducing specific shape priors to facial components and using SIFT patch descriptors to improve the robustness to lighting variations. Considering that MRF can only produce an approximate solution and cannot synthesize new patches that do not exist in training sets, Markov Weight Fields (MWF) was proposed in [16], where the sketch layer was composed of weight vectors. MWF essentially combined MRF with the linear transformation method, and thus can synthesize new sketch patches with smooth overlapping areas. In [28], Peng et al. proposed a multiple representations-based face sketch synthesis method, where multiple features were extracted from multiple filters and Markov networks were deployed to exploit the interacting relationships

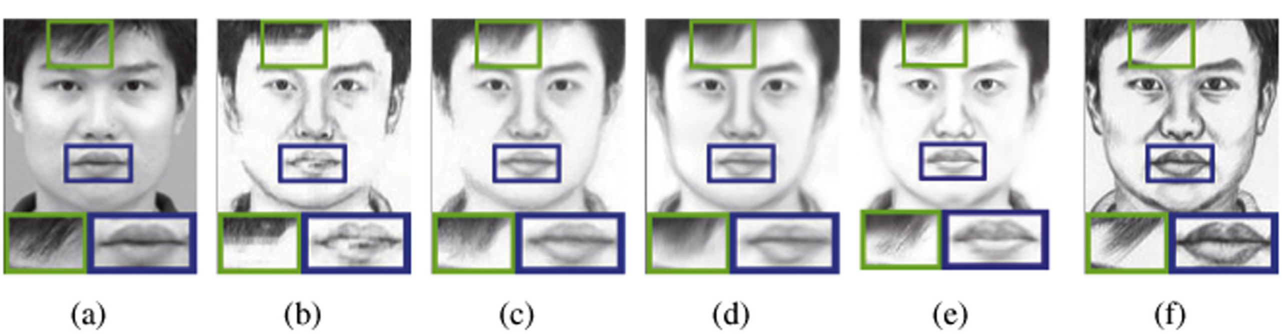

Fig. 2. Illustrative example. (a) The original photo; (b) Result by MRF [14]; (c) Result by MWF [16]; (d) Result by SSD [22]; (e) Result by the proposed method; (f) The sketch by artist.