Image and Vision Computing 44 (2015) 59–72

Contents lists available at ScienceDirect

Image and Vision Computing

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

Editor's Choice Article

Robust face hallucination using ensemble of feature-based regression functions and classifiers☆

Ching-Ting Tu *, Jang-Ren Luo

Department of Computer Science and Information Engineering, Tamkang University, Taiwan, ROC

ARTICLE INFO

Article history:
Received 20 January 2015
Received in revised form 18 July 2015
Accepted 12 October 2015
Available online 24 October 2015

Keywords:
Principal component analysis (PCA)
Canonical correlation analysis (CCA)
Eigenfaces
Facial hallucination

ABSTRACT

An example-based face hallucination system is proposed, in which given a low-resolution facial image, a corresponding high-resolution image is automatically obtained. In practice, such a problem is extremely challenging since it is often the case that two discriminative high-resolution images may have similar low-resolution inputs. To address this issue, this study proposes an ensemble of image feature representations, including various local patch- or block-based representations, a one-dimensional vector image representation, a two-dimensional matrix image representation, and a global matrix image representation. Notably, some of these representations are designed to preserve the global facial geometry of the low-resolution input, while others are designed to preserve the local detailed texture. For each feature representation, a regression function is constructed to synthesize a high-resolution image from the low-resolution input image. The synthesis process is conducted in a layer-by-layer fashion, with the output from one layer serving as the input to the following layer. Importantly, each regression function is associated with a classifier in order to determine which regression functions are required in the synthesis procedure in accordance with the particular characteristics of the input image. Furthermore, these classifiers can also help to deal with the individual ambiguity of system low-resolution inputs. The experimental results show that the proposed framework is capable of synthesizing high-resolution images from low-resolution input images with a wide variety of facial poses, geometry misalignments and facial expressions even when such images are not included within the original training dataset.

1. Introduction

The images captured by a camera in the real-world are often affected by such factors as unfavorable occlusion, poor lighting conditions, low resolution, and so on. Consequently, the images are unsuitable for such applications as facial recognition, human–computer interaction, or image retrieval. Accordingly, the problem of super-resolution has emerged as a major research topic in the computer vision field. Among the various applications reliant on super-resolution technology, that of face hallucination, in which a low-resolution facial image is converted into an equivalent high-resolution image, is of particular interest since it serves as an important preprocessing step for many face-related applications, including face recognition [4,38] and facial animation [15,18,35].

Broadly speaking, super-resolution methods can be classified as either single-image-based super-resolution schemes or example-based schemes. Methods of the former type require just one low-resolution input image, and include such methods as the interpolation-based method [1], the high-frequency enhancement-based method [6,22], and the edge enhancement method [26,34]. By contrast, example-based methods require multiple training examples in order to construct the knowledge required to synthesize a high-resolution image for an unseen input. Typical methods of example-based approach include those proposed in [8,10,13]. Facial images have a specific structure, and consequently most existing face hallucination methods adopt an example-based approach. More specifically, through the assistance of multiple training examples, the synthesis system learns the global and local facial geometry and local detailed texture which characterize the human face and then uses this information to predict (synthesize) the high-resolution facial image for any unknown low-resolution image.

The proposed framework is the example-based facial hallucination method, where the aim is to synthesize a high-resolution image which resembles the original low-resolution image to the greatest extent possible. In practice, such a process is extremely challenging since it is often the case that two discriminative high-resolution images may have similar low-resolution inputs. To address this issue, our framework composes an ensemble of image feature representations to include both the global facial geometry and the local geometry/detailed texture information. Notably, that the global facial geometry describes the relative positions and sizes of the individual facial features and facial contours, while the local geometry/detailed texture information describes the skin texture or specific characteristics of the facial features, such as

☆ Editor's Choice Articles are invited and handled by a select rotating 12 member Editorial Board committee. This paper has been recommended for acceptance by Stefanos Zafeiriou.

* Corresponding author. Tel.: +886 2 26215656x2745; fax: +886 2 26209749.
E-mail address: cttu@mail.tku.edu.tw (C.-T. Tu).

http://dx.doi.org/10.1016/j.imavis.2015.10.001

large eyes, double-fold eyelids, and so on. The proposed synthesis process is conducted in a layer-by-layer fashion, with the output from one layer serving as the input to the following layer. Here, each layer composes several regression functions, and each regression function is constructed based on a specified feature representation. Importantly, in order to deal with the individual ambiguity of system synthesized results, each regression function is associated with a classifier that determines if such regression function is required in the synthesis procedure in accordance with the particular characteristics of the input image.

1.1. Related works

Our approach is related to several example-based methods proposed recently in computer vision. For instance, statistical techniques [31] and [30] use a number of training data to learn the relationship between facial image and sketch pairs, and then apply the Bayesian inference or the regression process to synthesize new sketches. Such approaches have also been applied to human faces in face hallucination [32]. We briefly review related work in example-based face hallucination before introducing our approach.

The correlation between the low-resolution image and the high-resolution image is typically highly complex. Consequently, the mapping function from the low-resolution image to the high-resolution image tends to have a one-to-many characteristic; that is, visually similar low-resolution input images can often be mapped to many different discriminative high-resolution outputs. Accordingly, multiple-step, multiple-layer, or iterative synthesis approaches are typically required to solve the super-resolution problem. The face hallucination method proposed in [17] applies a parametric-based PCA approach to construct the global facial geometry first, and then uses a non-parametric MRF method to complement the local detailed texture. Similarly, the method presented in [36] uses cubic B-spline interpolation to construct the facial geometry first, and then applies a simplified MRF model to obtain the correlation between the AC coefficients of the high-resolution and low-resolution images in the discrete cosine transform (DCT) domain. In all of the aforementioned studies, the aim is to synthesize images containing the global facial geometry first, and then to add the facial high-frequency detailed textures in order to obtain a more realistic synthesis outcome. However, the methods presented in [24,29] apply an eigen-transformation method and an error back-projection method, respectively, to iteratively complement the difference between the high-resolution and low-resolution images directly.

In addition to adopting multiple-layer synthesis frameworks in order to preserve both the global features of the input image and the local texture details, existing face hallucination methods also adopt different facial area positions or sizes to maintain the global or local facial geometry details or the local texture properties. For example, works [20] and [14] provided patch-based face hallucination methods; the method proposed in [24] processes both the entire facial region and the particular facial regions corresponding to the local facial features, respectively, and then combines the two synthesized images to obtain the final result. Similarly, the method presented in [33] processes the facial image using three specific areas, namely the facial feature (component) area, the edge area, and the smooth area. Furthermore, various facial features have been also considered for synthesis purposes, including the patch gray value [11,16,21] and the residual gray value (i.e., the gray value difference between the low-resolution image and the high-resolution image) [19,23,37]. Moreover, the methods proposed in [23,36] adopt the DCT coefficient for image representation purposes, while that in [33] uses the image gradient feature.

1.2. Key contributions

The face hallucination framework proposed in this study adopts an iterative (layer-by-layer) approach to incrementally reduce the complex correlation between the low-resolution image and the high-resolution image. The proposed framework has three main advantages over existing methods, including (i) the objective of proposed synthesis framework is to minimize the difference between the synthesized results and the ground-truth high-resolution images, and simultaneously decrease individual ambiguity of these synthesized system synthesized results. Accordingly, the resolutions of synthesized results are close to those of ground-truth images, and these results are easy to be recognized as their corresponding input subjects. (ii) The framework operates on ensembles of various image (or patch) representations, patch positions, and patch sizes. Accordingly, the robustness of the synthesis process in capturing the unique facial geometry, personal characteristics, and local detailed textures of the input image is significantly improved. (iii) For each feature representation, a corresponding regression function is reconstructed for transferring the low-resolution image into the high-resolution image. Especially, the proposed method uses an ensemble of classifiers to select suitable regression functions for synthesis purposes based on the particular characteristics of the system input. Thus, the proposed framework is capable of synthesizing high-resolution images from low-resolution input images with a wide variety of facial poses, geometry misalignments, and facial expressions even when such images do not appear within the original training dataset.

2. Feature representations

In the present study, the synthesis framework is trained by several pairwise interpolated low-resolution image and high-resolution image, and the quality of the synthesis results is enhanced through the use of an ensemble of image features. Fig. 1 shows the various feature representations for each pairwise image sample.

1) *Image representation types* (Fig. 1(a)): two different image types are considered, namely the gray value-to-gray value type (Fig.1(a-1)) and the gray value-to-residual value type (Fig.1(a-2)). The gray value-to-gray value type is used to learn the gray value correlation of the pairwise sample, and this correlation is then used to synthesize the gray values of the high-resolution image for an unknown low-resolution input. Meanwhile, for the gray value-to-residual value type, the correlation between the gray values of the low-resolution image and the residual image (i.e., the difference between the high-resolution and low-resolution images) is learned, and this correlation is then applied to estimate the residual image for the unknown low-resolution input. Combining the estimated residual image with the low-resolution image, we have the synthesized high-resolution image.
2) *Image patch representations with various sizes and arrangements* (Fig. 1(b)–Fig. 1(c)): the gray value and residual value feature representations constitute the backbone of the proposed synthesis framework. However, as described in the following, several image block arrangements are also used for synthesis proposes. In order to retain both the global properties and the local facial properties of the input image, the framework proposed in the present study uses the following ensemble of image patch sizes and arrangements for feature representation purposes.

(i) *Equally-divided image patches of various sizes*: the interpolated low-resolution/high-resolution image is divided into a regular grid of image patches. The proposed framework considers three different patch sizes, namely 20×20, 10×10, and 5×5 pixels. Furthermore, each patch is represented as either a one-dimensional vector or a two-dimensional matrix utilizing either the gray value-to-gray value type or gray value-to-residual value type image representation. Consequently, each pairwise image sample yields $2 \times 3 \times 2 = 12$ features, where the numbers 2, 3 and 2 respectively denote the one-dimensional vector or two-dimensional matrix, three patch sizes, and gray value-to-gray value or gray value-to-residual value image type, respectively.