Optik 127 (2016) 525–534

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Hallucinating faces based on adaptive neighborhood selection and dual tree complex wavelet transform

Yu Deng[a,*], Fang Liu[b]

[a] Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
[b] China Ship Development and Design Center, Wuhan 430064, China

ARTICLE INFO

Article history:
Received 22 July 2014
Accepted 18 August 2015

Keywords:
Face hallucination
Super-resolution
Face alignment
Dual tree complex wavelet transform
Adaptive neighborhood selection

ABSTRACT

In this paper, a novel face hallucination method is proposed. Instead of utilizing the gray values of pixels directly, the feature vector extracted using dual tree complex wavelet transform (DTCWT) is exploited to describe the directional detailed information of facial image. Then, the adaptive neighborhood selection (ANS) algorithm is used to preserve neighborhood relationship between the low-resolution and high-resolution face by selecting the optimal neighbors of each facial patch from the feature space. The selected neighborhood can well reflect the local geometrical structure of face manifold, so that the linear subspace determined by the optimal linear fitting can approximate the local facial geometry with a higher accuracy. In order to reduce the effect of misalignment, a face alignment method based on three facial keypoints is also used. Experimental results show that the proposed face hallucination method outperforms other state-of-the-art methods in terms of visual inspection and objective evaluations, especially when the facial images are compressed and have low-resolution.

1. Introduction

Face hallucination or Face super-resolution is a technique to reconstruct a high-resolution (HR) facial image from its low-resolution (LR) inputs. It is very useful in many real applications such as video surveillance, in which the facial images captured by the live cameras could have a low resolution due to the limitations of hardware or a high compression ratio.

Baker and Kanade [1] first developed a pixel-wise face hallucination method, which used the Laplacian pyramid and the Gaussian pyramid to decompose an image into a pyramid of features in order to produce a HR facial image. Liu et al. [2] proposed a two-step approach consisting of a global parametric model with Gaussian assumption and a local nonparametric model based on Markov random fields (MRF). Inspired by Liu's suggestions, many researchers treated the face hallucination as a two-step problem [2–9]. Usually, the two-step approach contains two components: (1) a global facial image, which held the main characteristics of the original HR face but lacked some detailed features; (2) a residue facial image, which contains the high-frequency image information and could be synthesized and added to the global facial image to produce the final results. For example, Zhuang et al. [3] proposed the locality preserving hallucination algorithm which used locality preserving projection (LPP) and radial basis function (RBF) regression together to produce a global high-resolution facial image. Huang et al. [7] adopted canonical correlation analysis (CCA) to extract the coherent subspaces of HR and LR facial images. In their strategies, neighbor reconstruction was used for the input LR faces to determine the PCA coefficients of the corresponding HR images during the global face reconstruction step, and then CCA and neighbor reconstruction were used to add further detailed information by considering the residual images divided into patches. Hu et al. [8] presented a face hallucination framework termed from local pixel structure to global image super-resolution (LPS–GIS). Their framework utilized the input low-resolution (LR) facial image to search a face database for similar example high-resolution (HR) faces in order to learn the local pixel structures for the target HR face. And most recently, a two-step face hallucination approach was proposed to generate a HR facial image from several LR facial images [9]. It used a sparse representation and the approximate nearest neighbors (ANN) search method to enhance both global face shape and local high frequency information, which can improve the processing speed greatly. However, the global reconstruction procedure of two-step face hallucination algorithm usually results in a low reconstruction precision [10].

* Corresponding author. Tel.: +86 0 1013552936968.
E-mail address: dengyuee@mail.tsinghua.edu.cn (Y. Deng).

http://dx.doi.org/10.1016/j.ijleo.2015.08.041
0030-4026/© 2015 Published by Elsevier GmbH.

In contract to the two-step method, a number of one-step face hallucination methods have been proposed in the past decades. In [11], principal component analysis (PCA) method was used for representing the structural similarity of facial images. However, the results can hardly maintain the global smoothness and visual rationality, especially at those locations around the face contour and margin of the mouth. Ma et al. [10] reconstructed the HR image patch using the same position image patches of each training image. The optimal weights of the training image position-patches were estimated and the hallucinated HR patches were produced by using the same weights. In the following work, Ma et al. [12] further presented a sparse representation and position prior based face hallucination method for single facial image super-resolution. To retain the holistic structure of facial images, an efficient mapping model based on singular value decomposition (SVD) was proposed for reconstructing the HR face [13]. Li et al. [14] employed the prior models learned using sparse representation to guide the reconstruction process and achieved a superior performance in terms of both reconstruction error and visual quality. Motivated by manifold learning method such as local linear embedding (LLE), Chang et al. [15] developed Neighbor Embedding based on the assumption that the low- and high-resolution training images form manifolds with similar local geometry in two distinct feature spaces. Enlighten by Chang's work, a lot of face hallucination approaches [16–18] were developed based on this assumption. However, Su's experiments [19] showed that neighborhood preservation for low-resolution and high-resolution patches rarely holds. According to Su's suggestion and Zhang's study [20], there are two ways to increase the neighborhood preservation. One is to select a good feature to represent image patch, which can preserve neighborhood well. The other is to adaptively select the optimal neighbors to accurately reflect the local geometrical relationship of manifold.

In this paper, we propose an efficient one-step face hallucination method. Our method is mainly composed of four steps: First, the dual tree complex wavelet transform (DTCWT) is applied on the patches of the LR facial image to extract the features. Second, the adaptive neighborhood selection (ANS) algorithm is utilized to select the optimal neighbors of each LR image patch from the feature space. Third, the reconstruction weight vector for each input LR patch can be obtained by solving a constrained least squares problem. At last, the HR patch can be reconstructed using the same weights, and then the final HR facial image is synthesized by concatenating all the hallucinated HR patches. Additionally, the position of image patch is also used as one of the features in order to improve the precision of face hallucination.

The rest of this paper is organized as follows: In Section 2, a brief review of the dual tree complex wavelet transform (DTCWT) and adaptive neighborhood selection (ANS) algorithm is provided. In Section 3, the proposed face hallucination approach is proposed and explained in details. The experimental results are given in Section 4, followed by the conclusion in Section 5.

2. Review of DTCWT and ANS algorithm

2.1. Dual tree complex wavelet transform

It is well known that the discrete wavelet transform (DWT) is widely used in the field of digital signal/image processing [21–23]. However, the DWT often suffers from two main disadvantages [24]:

(1) Lack of shift invariance, which means that small shifts in the input signal will cause major variations in the distribution of energy between DWT coefficients at different scales.
(2) Poor direction selectivity for diagonal features, due to the wavelet filters are separable and real.

To overcome these limitations, the dual tree complex wavelet transform (DTCWT) was developed by Kingsbury [24,25] to provide a multi-scale decomposition of digital signal (and other multi-dimension data). Compared with DWT, the DTCWT has following advantages:

(1) Approximate shift invariance;
(2) Good directional selectivity in 2-dimensions (2-D) with Gabor-like filters (also true for higher dimensionality data);
(3) Perfect reconstruction (PR) using short linear-phase filters;
(4) Limited redundancy, independent of the number of scales, 2:1 for 1-D ($2m$:1 for m-D);
(5) Efficient N-order computation – only twice the simple DWT for 1-D ($2m$ times for m-D).

In the case of 2-D data, the DTCWT can generate six directional complex coefficient subbands (i.e. all of the coefficients are complex) at each decomposition level, as shown in Fig. 1, where $Z_{\pm}^{i,1}$, $Z_{\pm}^{i,2}$ and $Z_{\pm}^{i,3}$ represent the six-directional high frequency complex subbands orientated at ±15°, ±45° and ±75° at ith decomposition level, respectively.

Take the lena image as an example, Fig. 2 illustrates its 2-level DTCWT decomposition. From Fig. 2, it is obvious that much more directional information can be revealed by the DTCWT decomposition. That is, the high frequency details of the image can be clearly described by means of the complex wavelet coefficients, which is very crucial to image enhancement or restoration [26].

According to the superior properties of DTCWT, we will exploit it to extract the features from the image patches instead of directly using the gray values of pixels in the existing algorithms [1,2,4–6,8–12,14,15,17,18]. The feature extraction method will be discussed in Section 3 in details.

2.2. Adaptive neighborhood selection algorithm

Many real applications of machine learning, including signal processing, computer vision and financial data mining etc., inevitably produce a great number of high-dimensional data.

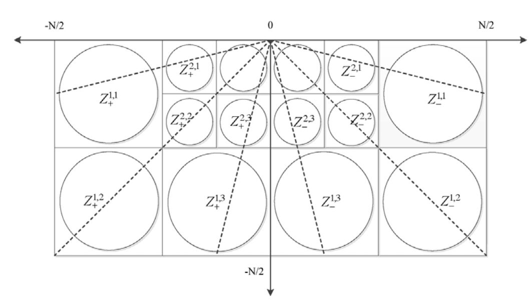

Fig. 1. Six directional subbands of 2-level 2D-DTCWT decomposition.

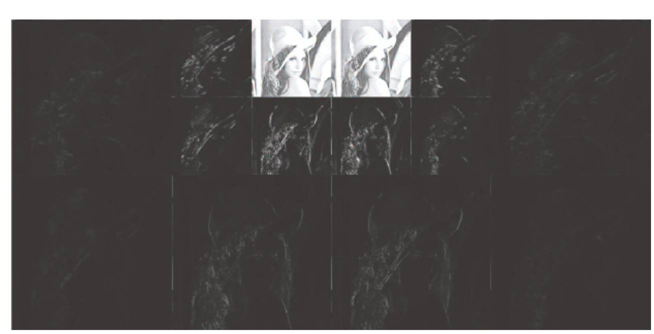

Fig. 2. The 2-level 2D-DTCWT decomposition of lena image.