Signal Processing 147 (2018) 233–246

Contents lists available at ScienceDirect

Signal Processing

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

Noise robust face hallucination algorithm using local content prior based error shrunk nearest neighbors representation

Shyam Singh Rajput[a,*], Ankur Singh[a], K.V. Arya[a,c], Junjun Jiang[b]

[a] Multimedia and Information Security Research Group, ABV-Indian Institute of Information Technology and Management, Gwalior, 474015, India
[b] School of Computer Science, China University of Geosciences, Wuhan, 430074, China
[c] Department of Computer Science & Engineering, Institute of Engineering and Technology, Lucknow 226021, India

ARTICLE INFO

Article history:
Received 14 September 2017
Revised 30 December 2017
Accepted 29 January 2018
Available online 31 January 2018

Keywords:
Super-resolution
Face hallucination
Learning and position-patch based method

ABSTRACT

In recent years face hallucination or super-resolution (SR) is getting much attention due to its wide applicability in real world scenarios. The existing SR methods and models perform well for noise free or small camera/atmospheric noisy faces. However, when suffering from mixed Impulse-Gaussian (MIG) noise, face hallucination becomes a challenging task. To address this problem, a novel error shrunk nearest neighbors representation (ESNNR) based face hallucination algorithm is proposed in this paper. Here, local content prior is incorporated to identify the high variance content (HVC) in the input images. The proposed algorithm suppresses the identified HVC in the input face to minimize the squared error. Moreover, the similarity matching between the input and training images is improved to achieve the locality and sparsity in the presence of MIG noise. Simulation results performed on public FEI, CAS-PEAL, CMU+MIT face databases, and locally captured surveillance video frames show that the proposed algorithm is computationally efficient, suitable for practical applications and give better performance than the existing face SR methods.

1. Introduction

In most of the real world scenario the images captured by the camera or sensor under uncontrolled environment are low-resolution (LR) images which may be degraded by various effects including atmospheric blur, camera blur [1], low-resolution imaging system, distance of the object from the imaging system, noise [2], etc. (see Fig. 1). High frequency information in the image is also lost during image storage, transmission and processing (analog to digital conversion, read-write operations, etc.) [3].

To ameliorate the quality and resolution of input LR images, SR methods have been developed and can be categorized as: generic image SR [1,2,4–12] and domain specific SR i.e., face SR [13–22] and text SR [23,24].

As a domain specific SR technique, the existing face SR can broadly be classified in two categories [17]: (i) reconstruction-based [25–27] and (ii) learning-based methods. The learning-based SR methods can further be classified as: Holistic (global face) [28–32] and patch based methods [13–20]. In holistic face based methods, input LR face is represented as a weighted linear combination of training LR faces by using various representation techniques [28,29,31,33,34] and then the resultant high-resolution (HR) face is synthesized from the corresponding training HR faces with same weight coefficients. In local patch based methods, global face is divided into small patches and each patch is processed in the same manner as for holistic methods. The review of some of the representative methods in this field is given below.

In 2000, the idea of the "Face Hallucination" was introduced by Baker and Kanade [35] to enhance the resolution of input LR face images. This was the first work on face SR that introduced the concept of learning and nearest neighbor representation techniques for resolution improvement. Here, gradient prior is predicted from training samples using Laplacian, Gaussian and feature pyramids to improve the resolution of input frontal face. A patch based learning SR algorithm called neighbor embedding (NE) has been developed by Chang et al. [13] which utilizes the concept of local linear embedding (LLE) [36]. The idea of NE is further explored by Zhang and Cham [37] where the DCT coefficient estimation in frequency domain is used to formulate an objective function instead of pixel intensity estimation in spatial domain. Huang et al. in [31], analyzed the nonlinear mapping between LR and HR in the coherent space by using canonical correlation analysis (CCA). The work of CCA is further extended to 2D CCA by An and Bhanu [38]. Hu et al. [39] developed the face SR framework based on the concept of the

* Corresponding author.
E-mail addresses: ershyamrajput@gmail.com, shyam@iiitm.ac.in (S.S. Rajput), kvarya@iiitm.ac.in (K.V. Arya).

https://doi.org/10.1016/j.sigpro.2018.01.030
0165-1684/

Fig. 1. Left and right are the two video frames recorded by the surveillance camera with 640 × 424 pixels resolution. In middle two low-resolution faces extracted from these two frames.

local pixel structure to global image SR. Another extension of the local pixel structure to global image SR was carried out by Shi et al. [40] by merging the local sparsity, global reconstruction, and local pixel correlation model into a unified regularization framework. Jiang et al. [41] proposed face SR method based on manifold alignment with the assumption that there is a common hidden manifold space between LR and HR manifold.

In this paper, local patch-based methods are primarily considered. They work with the assumption that if two LR face patches are similar then their counterpart HR patches are also similar. This is used to recover the lost high frequency details of LR patch from the training patches in HR space. The face SR model proposed by Ma et al. [14] differs from many of the previous manifold learning and probabilistic based models. The main task in this model is the computation of optimal reconstruction weight vector for a given patch of the test LR image which is computed from the same position patches in training LR space via least square representation (LSR). Further, it utilizes the same weight vector with counterpart patches in training HR space for synthesizing the super-resolved face. However, in this work the scheme of position-patch is used instead of neighbor patches. The position-patch based models work with some prior knowledge about the face such as common local geometric structure and alignment. The concept of least square estimation used in [14] may lead to biased results when the dimension of the patch is smaller as compared to the size of the training set. Consequently, it makes the solution of the least square problem unstable. Inspired by numerous sparse representation techniques [42], sparse coding (SC) based SR model was proposed by Yang et al. [8]. It was further explored by Wang et al. [43] by introducing the concept of weighted adaptive sparse regularization (WASR). Different from many of the previous methods, in [40] local sparsity, global consistency and pixel correlation are combined into a unified framework for face SR. This model performs well for noise free face image but with increase in noise level, the sparse coding coefficients may be unstable.

To overcome the limitations of above methods and models, the locality-constrained representation (LcR) based face hallucination model was proposed by Jiang et al. [16]. Here, locality between input and training LR patches is computed by pairwise Euclidean distance. This work mainly emphasizes on locality constraint which plays an important role in face hallucination. To achieve locality constraint the objective function is modified by adding the diagonal matrix of distance between the input and training patches with the L_2 norm (i.e., neighbor representation term). Consequently, during the reconstruction process the larger weights are assigned to more similar patches and smaller weights to less similar patches. The performance of LcR method can be improved by inclusion of locality constraint in the formulation. However, the concept of sparsity (another important constraint) has not been fully explored. To achieve these two constraints the work of [16] is further extended by introducing Tikhonov regularization term with neighbor embedding called Tikhonov regularized neighbor representation (TRNR) [19]. The use of Tikhonov matrix makes this model capable of selecting most similar patches from the training set to obtain discriminable facial information, especially when the LR face is noisy. It utilizes only K nearest neighbors from whole training which allow it to achieve the sparsity constraint also. Hence, TRNR [19] achieve the locality and sparsity simultaneously.

The method like LcR [16] and TRNR [19] consider the locality along with sparsity but the smoothness of the reconstruction weights are not taken into the account. In LcR [16], very different weights may be assigned to two similar patches, which is undesirable. To alleviate this problem Jiang et al. [18], incorporated fused least absolute shrinkage and selection operator (LASSO) regression with locality regularization to achieve smoothness in reconstruction weights.

As seen above, there are two major requirements for good face SR i.e., locality and sparsity. Both of these constraints utilizes the similarity matching between the input image and training images. In all of the methods discussed above, however, the similarity matching based on Euclidean distance will lead to unstable solution in the presence of heavy MIG noise. Therefore, the performance of these Euclidean distance similarity matching based face SR methods will drop sharply when encountered with heavy MIG noise. How to robustly super-resolve the LR face images with heavy MIG noise is the urgent problem to be solved.

1.1. Motivation and contribution

General strategy to handle MIG noise in the field of image processing is the median filter. In particular, the pixels from the input image are used to compensate the noisy content, which is not helpful when the noise level is very high (see Section 4.4.1 more discussion). Inspired by de-noising using outliers (noisy pixels) detection and estimation of new value via non-impulsive nearest neighbor searching and index distance weighted mean filtering method [44], this paper firstly incorporates local content prior (LCP) (see Section 3.1) to identify the high variance content (outliers) in the input and then the proposed ESNNR technique shrinks the identified content towards the guided LR face (see Fig. 2). Consequently, the use of error suppressed manifold in the data fitting efficiently reduce the squared residual. Moreover, it also helps in achieving the true locality and sparsity simultaneously in the presence of MIG noise. The main contributions of this work are as follows:

- This paper introduces a new local content prior (LCP) based thresholding technique to identify the high variance content (HVC), e.g., outliers or noise, in the input LR face. This will help the proposed algorithm (ESNNR) to shrink the identified HVC towards the calculated guided face (GF). Consequently, the proposed algorithm is able to abolish most of the noise from the input.
- Instead of treating all the pixel equally, the proposed SR algorithm uses the above error suppression in data fitting term of its objective function to reduce the influence of noisy pixels and obtain optimal reconstruction weights for synthesizing the resultant HR face.
- To test the real world applicability of the proposed algorithm, we additionally performed some experiments on LR faces extracted from CMU+MIT face database and locally captured surveillance video frames.

The rest of the paper is organized as follows: In Section 2, notations used in this paper are given. In Section 3, local content prior, objective function, optimization and pseudo-code of the proposed face hallucination algorithm are presented. Section 4 shows the experimental results and analysis performed on the public FEI, CAS-