Neurocomputing 133 (2014) 153–160

Contents lists available at ScienceDirect

Neurocomputing

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

Letters

Noise robust face hallucination employing Gaussian–Laplacian mixture model

Zhong-Yuan Wang, Zhen Han*, Rui-Min Hu, Jun-Jun Jiang

NERCMS, School of Computer, Wuhan University, Wuhan 430072, China

ARTICLE INFO

Article history:
Received 22 April 2013
Received in revised form
29 August 2013
Accepted 4 November 2013
Communicated by Shiguang Shan
Available online 10 January 2014

Keywords:
Face hallucination
Sparse representation
Gaussian–Laplacian mixture
Elastic net

ABSTRACT

Because of the excellent ability to characterize the sparsity of natural images, ℓ_1-norm sparse representation (SR) is widely used to formulate the linear combination relationship in dictionary-learning-based face hallucination. However, due to inherently less sparse nature of noisy images, Laplacian prior assumed for ℓ_1-norm seems aggressive in terms of sparsity, which ultimately leads to significant degradation of hallucination performance in the presence of noise. To this end, we suggest a moderately sparse prior model referred to as a Gaussian–Laplacian mixture (GLM) distribution and employ it to infer the optimal solution under the Bayesian framework. The resulting regularization method known elastic net (EN) not only maintains same hallucination performance as SR under noise free scenarios but also outperforms the latter remarkably in the presence of noise. The experimental results on simulation and real-world noisy images show its superiority over some state-of-the-art methods.

1. Introduction

Face super-resolution, or face hallucination, refers to the technique of estimating a high-resolution (HR) face image from low-resolution (LR) face image sequences or a single LR one. Face hallucination is extensively used for pre- and/or post-processing in video applications, such as video surveillance. It is universally acknowledged that current face hallucination methods fall into three categories: interpolation, reconstruction-based and learning-based methods. Among them, learning-based methods have attracted much attention since they can provide high magnifying factors.

Learning-based methods can date back to the early work proposed by Freeman et al. [1] who employed a patch-wise Markov network to model the relationship between LR images and the HR counterparts. Afterwards, Baker and Kanade [2] developed a Bayesian approach to infer the missing high-frequency components from a parent structure with the help of training samples, and first coined the term "face hallucination". Following their pioneering work, Liu et al. [3] presented a two-step statistical modeling approach which integrates global structure reconstruction with local detail refinement. Chang et al. [4] proposed a neighbor embedding (NE) method, which merely uses K nearest neighbors instead of the entire training set for reconstruction. Ma et al. [5] introduced a position-patch based method to estimate a HR image patch using the same position patches of all training face images. As far as accuracy and stability are concerned, the above-mentioned representation methods are unsatisfactory. Specifically, ridge regression (RR) employed by Ref. [3] is hard to capture salient properties of natural images, while least squares (LS) estimator in Refs. [4,5] fails to guarantee the stability of solution.

Recently, considerable effort has been spent in designing alternative sparse representation (SR) based face hallucination. Yang et al. [6] are the first to introduce ℓ_1-norm SR to face hallucination, who proposed a local patch method with respect to coupled over-complete patch dictionaries to enhance the detailed facial information. Lately, Zeyde et al. [7] came up with a modified version by using a different training approach for the dictionary-pair and gained improved results and high efficiency as well. In Ref. [8], authors presented a dual-dictionary learning scheme to recover more image details, in which not only main dictionary but also residual dictionary is learned by sparse representation.

In addition to the generic sparsity prior in the above work [6–8], some specific image priors can be further exploited to boost the performance of SR based image restoration. Considering facial positions, Jung et al. [9] advanced a position-patch face hallucination method with LS [5] replaced by SR algorithm. Motivated by nonlocal similarity of image statistics, Lu et al. [10] put forward geometry constrained sparse coding (GCSR) for single image super-resolution. Very recently, Dong et al. [11] proposed nonlocally centralized sparse representation (NCSR) to explore the image nonlocal self-similarity. Methods in Refs. [10,11] can obtain good estimates of the sparse coding coefficients, whereas, without

* Corresponding author. Tel./fax: +86 27 876 482 33.
E-mail address: hanzhen_2003@hotmail.com (Z. Han).

0925-2312/$ - see front matter Crown Copyright © 2014 Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.11.021

the help of high frequency components from face database examples, they may fail to recover visual details under large magnification. In a multi-scale dictionary method [12], local and non-local priors are integrated simultaneously, with the local prior suppressing artifacts and the non-local prior enriching visual details. Such similarity-guided methods [10–12] promote the learning performance of SR algorithm and demonstrate impressive results for generic images. However, when the self-similarity assumption does not hold well (e.g., human face images), their performance will be considerably restricted, especially for noisy images since noise damages the image self-similarity.

Essentially, SR based methods incorporates sparsity prior knowledge as a constraint on the solution to obtain a global unique and stable one. In our empirical observations, however, noisy images exhibit less sparsity than noise free images in solution space and Laplacian prior assumed for ℓ_1-norm does not quite agree with the actual distribution. Consequently, just as pointed out by Ref. [13], ℓ_1-norm is aggressive in terms of sparsity, leading to considerable degradation of hallucination performance in the presence of noise. In this paper, we manage to seek a more fitted prior model for hallucinating noisy face images, which is referred to as Gauss–Laplacian mixture (GLM) distribution. GLM is a discrete mixture of a Gaussian distribution and a Laplacian one, which was introduced by Kanji [14] as a model for wind shear data. By placing a GLM prior on solution, we can derive a moderately sparse regularization method under Bayesian framework, which turns out to be elastic net (EN) [15]. Our method maintains the same performance as ℓ_1-norm SR based methods under noise free conditions while outperforms the latter in the presence of noise. Experimental results on simulated noisy images as well as real-world images validate its effectiveness. The main contributions are as follows:

(1) Our study on statistical properties of face hallucination behavior reveals that noisy images are less sparse than noise free images, and thus a relatively conservative GLM distribution is introduced to model coefficient prior.
(2) We are the first to introduce EN to face hallucination and achieve satisfactory hallucinated results in the presence of noise.
(3) An automatic estimation way for regularization parameters is devised to speed up coefficient training.

The remainder of this paper is organized as follows. Section 2 particularly introduces the proposed method. Various experimental results are shown in Section 3. Section 4 concludes the paper.

2. Proposed method

Face hallucination is typically a linear inverse problem. In this part, following the well-known maximum *a posteriori* (MAP) estimation framework, we employ GLM to infer the optimal solution of this linear problem:

$$x = Yw + \varepsilon \tag{1}$$

where $\mathbf{x} \in R^{N \times 1}$ is a N-dimensional observation vector, $\mathbf{Y} \in R^{N \times M}$ is a training set (also called dictionary) consisting of M basis images with each column being one, and $\boldsymbol{\varepsilon}$ stands for reconstruction error. The solution denoted by M-dimensional vector $\mathbf{w} = [w_1, w_2, \ldots, w_M]^T$ consists of a set of linear combination coefficients. Each entry in $\mathbf{w}$ is associated with an individual base in the training dictionary.

2.1. Modeling prior with GLM

A variety of image prior models have been suggested to formulate the appropriate prior knowledge of natural images. For example, widespread sparsity prior assumes that coefficient vector $\mathbf{w}$ is governed by i.i.d. zero-mean multivariate Laplacian distribution, namely,

$$P_L(\mathbf{w}) = \frac{1}{(2\mu)^M} \exp\left(-\frac{\|\mathbf{w}\|_1}{\mu}\right) \tag{2}$$

which just corresponds to classic ℓ_1-norm sparse representation. Scale parameter $\mu = \frac{\sigma_w}{\sqrt{2}}$ indicates the diversity and σ_w is standard variance of coefficients.

Another often adopted prior for coefficient vector $\mathbf{w}$ is Gaussian distribution in the following form

$$P_G(\mathbf{w}) = \frac{1}{(2\pi\sigma_w^2)^{M/2}} \exp\left(-\frac{\|\mathbf{w}\|_2^2}{2\sigma_w^2}\right) \tag{3}$$

where $\|\ \|_2^2$ denotes squared ℓ_2-norm and is actually associated with ridge regression.

In contrast to the sharp peak at zero in Laplacian prior, Gaussian prior is relatively conservative in the sense of sparseness. Nevertheless, as stated in Ref. [13], due to its non-singularity at origin, the resulting squared ℓ_2-norm is non-sparse. To enforce the prediction accuracy for noisy images while not completely lose variable selection functionality of the model, we therefore employ GLM distribution to represent the latent prior in coefficient space, which is formulated as follows:

$$P(\mathbf{w}) = \alpha P_L(\mathbf{w}) + (1-\alpha) P_G(\mathbf{w}) \tag{4}$$

This mixture is characterized by the mixing proportion α, $0 \le \alpha \le 1$, which denotes the fraction of Laplacian in the distribution. Because this fraction is generally unknown *a priori*, it has to be estimated from the data.

For simplification, we optimize α by minimizing error between actual and assumed distributions instead of using complex EM algorithm [16]. Suppose $\{H(i) | 1 \le i \le L\}$ is discrete actual distribution obtained by histogram analysis method, where L is sampling number; $\{P_L(i) | 1 \le i \le L\}$ and $\{P_G(i) | 1 \le i \le L\}$ are fitted Laplacian and Gaussian distributions by maximum likelihood (ML) method, the cost function in minimum squared error is expressed as

$$\alpha^* = \arg\min_{\alpha} \sum_{i=1}^{L} \{\alpha P_L(i) + (1-\alpha) P_G(i) - H(i)\}^2 \tag{5}$$

whose close-form solution is given by

$$\alpha^* = \frac{\sum_{i=1}^{L} \{P_G(i) - H(i)\}\{P_G(i) - P_L(i)\}}{\sum_{i=1}^{L} \{P_L(i) - P_G(i)\}^2} \tag{6}$$

Note that, the solution should be clipped to the range of [0, 1], inclusively.

The plausibility of GLM model can be verified by an experiment, where original noiseless images and noisy images corrupted by Gaussian noise with two different standard deviations of $\sigma = 5$ and 10 get involved. We carry out ordinary linear regression analysis to acquire regression coefficients and then use histogram method to get their actual distribution. The fitted distribution is generated with ML estimator from available coefficient samples. Finally, the probability density curves of the actual distribution and three kinds of fitted ones (Laplacian, Gaussian and GLM) are depicted in Fig. 1.

By briefly examining the actual distribution histograms under different noise levels in Fig. 1, we can find that their peaks at zero indeed decline with respect to noise interference. In other words, the coefficients become less sparse with the increasing amount of noise (a signal is sparse if most entries of the coefficient vector are zero or close to zero). Fig. 1(a) shows Laplacian distribution is well fitted to the actual one under noise free conditions. As shown in Fig. 1(b) and (c), with the reduced sparsity resulting from noise, sharply-peaked Laplacian turns far over-sparse relative to actual