Signal Processing 103 (2014) 142–154

Contents lists available at ScienceDirect

Signal Processing

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

Generalized joint kernel regression and adaptive dictionary learning for single-image super-resolution

Chen Huang*, Yicong Liang, Xiaoqing Ding, Chi Fang*

State Key Laboratory of Intelligent Technology and Systems, Tsinghua National Laboratory for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

ARTICLE INFO

Article history:
Received 4 June 2013
Received in revised form
30 September 2013
Accepted 18 November 2013
Available online 27 December 2013

Keywords:
Single-image super-resolution
Face hallucination
Face recognition
Joint kernel regression
Dictionary learning

ABSTRACT

This paper proposes a new approach to single-image super-resolution (SR) based on generalized adaptive joint kernel regression (G-AJKR) and adaptive dictionary learning. The joint regression prior aims to regularize the ill-posed reconstruction problem by exploiting local structural regularity and nonlocal self-similarity of images. It is composed of multiple locally generalized kernel regressors defined over similar patches found in the nonlocal range which are combined, thus simultaneously exploiting both image statistics in a natural manner. Each regression group is then weighted by a regional redundancy measure we propose to control their relative effects of regularization adaptively. This joint regression prior is further generalized to the range of multi-scales and rotations. For robustness, adaptive dictionary learning and dictionary-based sparsity prior are introduced to interact with this prior. We apply the proposed method to both general natural images and human face images (face hallucination), and for the latter we incorporate a new global face prior into SR reconstruction while preserving face discriminativity. In both cases, our method outperforms other related state-of-the-art methods qualitatively and quantitatively. Besides, our face hallucination method also outperforms the others when applied to face recognition applications.

1. Introduction

Single-image super-resolution (SR) refers to the task of estimating a high resolution (HR) image $\mathbf{X} \in \mathbb{R}^n$ from a *single* low resolution (LR) image $\mathbf{Y} \in \mathbb{R}^m$ (lexicographically ordered vector and $m < n$). SR techniques are central to various applications, such as medical imaging, satellite imaging and video surveillance. They are especially necessary for face recognition applications in video surveillance systems because the face resolution is normally low in surveillance video, causing the loss of essential facial features for recognition purposes. The SR of face images is also called face hallucination [1–5].

The imaging model in the SR problem is generally expressed as

$$\mathbf{Y} = \mathbf{DHX} + \mathbf{V}, \tag{1}$$

where $\mathbf{D} \in \mathbb{R}^{m \times n}$ and $\mathbf{H} \in \mathbb{R}^{n \times n}$ are the downsampling matrix and blurring matrix respectively, and $\mathbf{V} \in \mathbb{R}^m$ is assumed to be an additive Gaussian white noise vector. Then recovering an HR $\mathbf{X}$ from the input LR $\mathbf{Y}$ is an ill-posed problem, and the optimal HR image $\mathbf{X}$ can be found by maximizing the posterior probability $p(\mathbf{X}|\mathbf{Y})$ based on the *maximum a posteriori* (MAP) criterion and Bayesian rule

$$\hat{X} = \arg\max_{\mathbf{X}} p(\mathbf{X}|\mathbf{Y}) = \arg\max_{\mathbf{X}} \frac{p(\mathbf{Y}|\mathbf{X})p(\mathbf{X})}{p(\mathbf{Y})}. \tag{2}$$

* Corresponding authors. Tel.: +86 10 62772369 645.
E-mail addresses: yach23@gmail.com (C. Huang), liangyicong@ocrserv.ee.tsinghua.edu.cn (Y. Liang), dxq@ocrserv.ee.tsinghua.edu.cn (X. Ding), fangchi@ocrserv.ee.tsinghua.edu.cn (C. Fang).

0165-1684/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.sigpro.2013.11.042

Generally, $p(\mathbf{Y}|\mathbf{X})$ is modeled by the Gaussian distribution, thus maximizing $p(\mathbf{Y}|\mathbf{X})$ boils down to minimizing the data constraint [2] $\|\mathbf{Y}-\mathbf{DHX}\|_2^2$. On the other hand, $p(\mathbf{X})$ codes the prior knowledge that we want to impose in the HR space. Typically, the task of SR reconstruction is formulated as a regularized least-square optimization problem as follows:

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \|\mathbf{Y}-\mathbf{DHX}\|_2^2 + \lambda \mathbf{C}(\mathbf{X}), \quad (3)$$

where λ is the parameter balancing the effects of the data constraint and the regularization term $\mathbf{C}(\mathbf{X})$. Most of the past works focus on designing different formulations of $\mathbf{C}(\cdot)$ to regularize the ill-posed reconstruction problem.

Currently, the single-image SR methods can be mainly categorized into three classes: interpolation-based methods, reconstruction-based methods, and example-based methods. Interpolation techniques (e.g. [6]) are simple and fast but tend to blur the fine details. The reconstruction-based methods (e.g. [7–10]) follow the form of Eq. (3) and how to design a good *image prior* is always an essential issue; $\mathbf{C}(\cdot)$ is usually a smoothness constraint. The example-based methods (e.g. [9,11–15]) hallucinate detailed textures from a training set of LR/HR image or patch pairs. However, such methods strongly rely on the chosen dataset for satisfactory results.

Many example-based methods directly or implicity use a co-occurrence prior to constrain the correspondence between LR and HR patches. For example, Yang et al. [11] explored the sparse representation of LR patches over an LR dictionary, and used the same representation coefficients to generate the HR output, but the result usually suffers from inconsistency between neighboring patches. Other natural image priors are also studied in the literature. The gradient profile prior is developed in [8] to preserve sharp edges, but is limited in modeling the visual complexity of real images. Later, priors of image *self-similarities* and *local/nonlocal regularities* have been exploited for more robust estimation. In [9], the nonlocal self-similarity properties both within and across spatial scales are fully exploited, but the local regularities are neglected. Zhang et al. [10] improved by assembling the Steering Kernel Regression [16] (SKR)-based local prior and Nonlocal Means [17] (NLM)-based nonlocal prior, whose connection, however, remains loose.

Another trend in SR is to combine the reconstruction- and example-based methods (usually dictionary induced) into a unified framework to produce more compelling results. In fact, SR can be viewed as a *regression* problem aiming to map LR images to target HR images. Then in this sense, dictionary-based methods do local regression using bases learned from an external database or the input image itself, while regression models directly estimate HR pixels (kernel learning) or regularize the estimator. As for the regression models, examples include SKR [16], Gaussian Process Regression (GPR) [12], Kernel Ridge Regression (KRR) [13] and Non-Local Kernel Regression (NLKR) [14], and they can all be effectively exploited as a prior for SR reconstruction. Among them, NLKR overcomes the drawbacks of the literature [10] by unifying the local and nonlocal priors into a single model in a complimentary way, but it discards the further potential enabled by the higher-order statistics. Besides, it needs a separate deblurring process which is ill-posed by itself. In our previous work [18], we proposed an Adaptive Joint Kernel Regression (AJKR) method, combining a set of coherent NLM-generalized local regressors in the nonlocal range with higher-order information (i.e. *regional redundancy*) injected in. By further integrating adaptive dictionary learning under the MAP framework, this algorithm produces superior results than NLKR and excludes the necessity of separate deblurring. However, it only builds kernel regressors at the same scale and rotation. To exploit the full potential offered by such joint regression, the core algorithm should be generalized.

For face hallucination, the above methods dealing with general natural images cannot be readily applied due to the ignorance of the special properties of face images. This problem was first addressed in the pioneering work of Baker and Kanade [1]. They adopted an image pyramid to learn a prior on the gradient distribution of frontal face images using the Bayesian theory. However, the HR image prediction is pixelwise which causes discontinuities and artifacts. To generate high-quality HR face images, the current face hallucination methods usually involve two steps. The first step reconstructs global faces in the face subspace using MAP criterion [11,2] or manifold learning methods. Principal Component Analysis (PCA) [2,3] is widely used for face modeling. Classic manifold learning methods include Locality Preserving Projections (LPPs) [5], Canonical Correlation Analysis (CCA) [4] and so on. While neighborhood preservation and correlation maximization are the only concerns, *discriminativity* is often lost and the frequently used PCA, for example, yields results like mean face; the second step produces a residue image to recover details [2,4,5,11].

This paper focuses on SR from a given LR version of general natural image or face image. Similar to our previous AJKR method [18], we address this problem from the viewpoints of learning good regression priors and robust dictionaries. However, we generalize AJKR in two ways: (1) to extend the regression range to multi-scales and rotations, obtaining a Generalized AJKR (G-AJKR) method, and (2) to incorporate a new global structure prior of human faces into the G-AJKR method for face hallucination, while preserving individual discriminativity based on Partial Least Squares (PLS) [19], which is very important when applied to face recognition applications.

The remainder of the paper is organized as follows. Section 2 reviews related works on dictionary and manifold learning as the development that follows relies on them. Section 3 details our G-AJKR framework and its extension to face hallucination. Experimental results of SR on generic and face images with applications in face recognition are provided in Section 4. We conclude the paper in Section 5.

2. Related works

2.1. Dictionary learning

Learning a good dictionary is important for example-based methods to do local "regression" using the learned