Pattern Recognition Letters 31 (2010) 1–10

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Resolution enhancement based on learning the sparse association of image patches

Jinjun Wang *, Shenghuo Zhu, Yihong Gong

NEC Laboratories America Inc., Cupertino, CA 95014, USA

ARTICLE INFO

Article history:
Received 26 September 2008
Available online 8 September 2009

Communicated by M. Ferretti

Keywords:
Image super-resolution
Image representation
Sparse-coding

ABSTRACT

Example-based image super-resolution techniques model the co-occurrence patterns between the middle and high frequency layers of example images to estimate the missing high frequency component for low resolution input. However, many existing approaches seek to estimate the optimal solution within a small set of candidates by using empirical criteria. Hence their representational performance is limited by the quality of the candidate set, and the generated super-resolution image is unstable, with noticeable artifacts. In this paper, we propose a novel image super-resolution method based on learning the sparse association between input image patches and the example image patches. We improve an existing sparse-coding algorithm to find sparse association between image patches. We also propose an iterative training strategy to learn a redundancy reduced basis set to speed up the super-resolution process. Comparing to existing example-based approaches, the proposed method significantly improves image quality, and the produced super-resolution images are sharp and natural, with no obvious artifact.

1. Introduction

Images with high pixel density are desirable in many applications, such as high resolution (HR) medical images for medical diagnosis, high quality video conference, high definition television broadcasting, Blu-ray movies, etc. While people can use higher resolution camera for the purpose, there is an increasing demand to shoot HR image/video from low-resolution (LR) cameras such as cell phone camera or webcam, or converting existing standard definition footage into high definition video material. Hence software resolution enhancement techniques are very desirable for these applications.

The task of software image resolution enhancement is to estimate more pixel values to generate an image with higher resolution. Simple interpolation methods cannot recover the missing high frequency components and often blur the discontinuities of magnified images. The super-resolution (SR) idea was introduced by Tsai and Huang (1984) who reconstructed the HR image by recovering additional information from multiple LR images using signal processing techniques. The SR approach has attracted increasing research attentions in recent years, and many improved applications such as single image SR and real-time SR, have been reported (Borman and Stevenson, 1998b; Park et al., 2003). Existing SR methods can be divided into three major categories, specifically the functional-interpolation, the reconstruction-based and the learning-based methods, as reviewed in the following paragraphes, respectively.

The functional-interpolation methods apply an existing function on the LR input to obtain a processed image (Irani and Peleg, 1991, 1993; Dai et al., 2007; Wang and Gong, 2008). Unlike those simple interpolation methods which cannot recover the missing high frequency information, the functional-interpolation based method encodes certain prior knowledge into the applied functions, such that the produced SR images can have some high frequency components, and hence are more perceptual pleasing. For instance, Irani and Peleg (1991) applied a simulated image down-grading process to project the interpolated image into an LR image, and then projected the difference between the original LR input and the projected LR image back into the interpolated image to obtain a processed result. Dai et al. (2007) applied the alpha-matting model to obtain a maximum a posteriori (MAP) decomposition of any local image patch into background and foreground components, and reconstruct the discontinuity between the two. The method generates SR images with good perceptual quality, but is very computational expensive. Hence Wang and Gong (2008) applied the connected-components-analysis technique to speed up the decomposition process, and designed a dedicated spatial filter to reconstruct the discontinuity and allow the system to perform near real-time image/video SR. However, the limitation with functional-interpolation methods is that, since the parameters of the true image down-grading process are usually unknown, the applied functions are not stable and sometimes generate serious artifact for certain types of images.

In the second category, the reconstruction-based methods build an HR image from a sequence of LR images (Tsai and Huang, 1984; Borman and Stevenson, 1998b; Park et al., 2003). It forms the largest body of SR research (Jiji and Subhasis, 2006). The

* Corresponding author. Fax: +1 408 863 6099.
E-mail address: jjwang@sv.nec-labs.com (J. Wang).

0167-8655/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2009.09.004

reconstruction-based methods assume that, the observed LR images result from shifting, warping, blurring, and subsampling operators performed on the HR image, corrupted by additive noise (Nguyen et al., 2001b). Since the down-grading process can be regarded as a linear transform, Ng et al. (2002) regarded the SR resolving process as a constrained least squares fitting problem by solving a transform-based preconditioned system (Ng and Bose, 2002; Nguyen et al., 2001a) applied Tikhonov-regularization and proposed circulant block preconditioners to accelerate the reconstruction process; Schultz and Stevenson (1994) and Rajan and Chaudhuri (2002)) used the Huber–Markov Random Field prior and described a MAP estimator; and Elad and Feuer (1997) proposed a unified restoration methodology by combining maximum likelihood (ML), MAP, and Projection Onto Convex Sets approaches. However, the limitation of reconstruction-based methods is that, as the image magnification factor increases, the reconstruction constraints provide less and less useful information (Borman and Stevenson, 1998a; Lin and Shum, 2004). In addition, it requires that the image pairs are related by global parametric transformations, and/or the parameter of the camera's Point-Spread-Function (PSF) is known in advance, which restrict its usage in various real scenarios. Besides, moving the camera/object to obtain a set of displaced LR images while minimizing the motion blur requires expensive hardware devices (Ben-Ezra et al., 2004).

The third category of current SR techniques is the learning-based methods which is very popular in recent years. It solves the SR problem by learning the co-occurrence prior between HR and LR image patches or coefficients (Sun et al., 2003; Jiji et al., 2004), and processing the LR input along with appropriate smoothness constraint (Freeman et al., 2000; Capel and Zisserman, 2001) to generate HR image. Hence learning-based SR is also called "example-based" SR (Freeman et al., 2000; Wang et al., 2005; Chang et al., 2004; Kong et al., 2006) or "recognition-based" SR (Borman and Stevenson, 1998a). To list some examples, Freeman et al. (2000) introduced a parametric Markov Network to learn both the co-occurrence prior between local HR/LR patches and global dependence. Wang et al. (2005) extended Freeman's framework by using Conditional Random Field to inference both the matched patches and the PSF parameters. Chang et al. (2004) applied Locally Linear Embedding (LLE) to weighted combine rank-N best matched LR patches to estimate one local HR patches. To utilize domain-specific knowledge, learning for certain types of images, such as face (Wang and Tang, 2004), fingerprint (Jiji and Chaudhuri, 2004) and text (Dalley et al., 2004), using visual intensity, wavelet coefficients (Jiji et al., 2004), image contourlet (Jiji and Subhasis, 2006), or edge profile (Sun et al., 2008) features, have also been examined.

In general, since the introduction of the SR idea by Tsai and Huang (1984), image/video SR has become one of the most spotlighted research areas, because it can overcome the inherent resolution limitation of the imaging system and improve the performance of most digital image processing applications (Park et al., 2003). Comparing to other two categories, the learning-based methods are able to process single image input, and are more stable by using statistical learning technique, hence they have become a promising direction for the SR problem. However, the capacity of current learning-based SR methods in modeling image patterns is low due to the lack of effective ways to represent the image co-occurrence prior knowledge, hence their performance dependents on the training set. To cope with this limitation, this paper presents a novel example-based image SR method based on discovering the sparse association between input and example patches, enabling precise modeling of distinct image patterns for better SR quality. Our major contributions include:

(1) We propose to apply sparse-coding technique to discover the sparse associations between input image patches and a large example image patch set. The method greatly extends the capacity of modeling the distinct image patterns for better estimation performance. Comparinge to existing example-based approaches, the proposed method significantly improves the image quality, and the produced SR images are sharp and natural, with no obvious artifact.
(2) We adopt and modify an existing sparse-coding algorithm used in text-mining domain to learn the sparse associations between image patches. Our modification solves two critical problems in the original algorithm when applied to SR problem. The advantage of the algorithm is that, by adjusting the sparsity setting, the proposed SR method can generate different SR quality with different processing time to meet various applications' requirement.
(3) We present a learning strategy to reduce the redundancy in extracted image patch examples. We obtain a small set of basis vectors by iteratively searching for the optimal coding and solving a bounded least square fitting problem with Tikhonov-regularization. Experimental results show that the size-reduced patch example set leads to minimal loss of image quality while speed up the overall SR processing.

The remainder of the article is organized as follows: Section 2 describes some background about existing example-based image SR techniques. Section 3 presents our proposed SR method based on an improved sparse-coding search algorithm and a basis set redundancy reduction method. Section 4 lists the experimental results, and Section 5 summarizes the proposed methods and discusses future works.

2. Example-based SR approaches

Example-based SR approaches (Freeman et al., 2000) assume that, an HR image (Fig. 1a) consist of three frequency layers, specifically the high frequency layer (denoted as h, Fig. 1b), the middle frequency layer (denoted as m, Fig. 1c), and the low frequency layer (denoted as l, Fig. 1d). The LR image results from discarding the high frequency components from the original HR version. Hence the example-based approach solves the SR problem by maximizing $Pr(h|m,l)$ for any input LR image. In addition, the high frequency component h is independent of l, hence it is only required to maximize $Pr(h|m)$, which greatly reduces the variability to be stored in the example set.

A typical example-based SR resolving process works as follows: From the example images, N patch pairs $\{\boldsymbol{p}_i^m, \boldsymbol{p}_i^h\}_{i=1}^N$ are extracted from the middle frequency layer m and the corresponding high frequency layer h, respectively. Both $\boldsymbol{p}_i^m$ and $\boldsymbol{p}_i^h$ are the column expansion of a square image region. Their dimensions are $D^m \times 1$ and $D^h \times 1$, respectively, and often $D^m > D^h$. In reconstruction, for an LR input image, L middle frequency patches are extracted in a similar way, and denoted as $\left\{\boldsymbol{y}_j^m\right\}_{j=1}^L$. The missing high frequency components $\left\{\boldsymbol{y}_j^h\right\}$ are estimated based on the co-occurrence patterns stored in $\{\boldsymbol{p}_i^m, \boldsymbol{p}_i^h\}_{i=1}^N$. The following subsections review three different models for the estimation process.

2.1. Nearest neighbor

Assuming that an image patch is of Normal distribution, i.e., $Pr(\boldsymbol{y}^m) \sim \mathcal{N}(\boldsymbol{\mu}^m, \boldsymbol{\Sigma}^2)$, and $Pr(\boldsymbol{y}^h|\boldsymbol{y}^m) \sim \mathcal{N}(\boldsymbol{\mu}^h, \boldsymbol{\Sigma}^2)$, it can be easily verified that, for any observed patch $\boldsymbol{y}_j^m$ from the input LR image, the ML estimation of $\boldsymbol{\mu}_j^m$ minimizes the following objective function