Neurocomputing 82 (2012) 21–28

Contents lists available at SciVerse ScienceDirect

Neurocomputing

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

Resolution-invariant coding for continuous image super-resolution

Jinjun Wang [a,*], Shenghuo Zhu [b]

[a] Epson Research and Development, Inc., 2580 Orchard Parkway, San Jose, CA 95131, United States
[b] NEC Laboratories America, Inc., 10080 N. Wolfe Road, Cupertino, CA 95014, United States

ARTICLE INFO

Article history:
Received 21 June 2011
Received in revised form
30 August 2011
Accepted 21 September 2011
Communicated by Qingshan Liu
Available online 24 December 2011

Keywords:
Image representation
Sparse-coding
Image super-resolution

ABSTRACT

The paper presents the resolution-invariant image representation ($\mathcal{R}$IIR) framework. It applies sparse-coding with multi-resolution codebook to learn resolution-invariant sparse representations of local patches. An input image can be reconstructed to higher resolution at not only discrete integer scales, as that in many existing super-resolution works, but also continuous scales, which functions similar to 2-D image interpolation. The $\mathcal{R}$IIR framework includes the methods of building a multi-resolution bases set from training images, learning the optimal sparse resolution-invariant representation of an image, and reconstructing the missing high-frequency information at continuous resolution level. Both theoretical and experimental validations of the resolution invariance property are presented in the paper. Objective comparison and subjective evaluation show that the $\mathcal{R}$IIR framework based image resolution enhancement method outperforms existing methods in various aspects.

1. Introduction

Most digital imaging devices produce a rectangular grid of pixels to represent the photographic visual data. This is called the *raster image*. The human perceptual clarity of a raster image is decided by its spatial resolution which measures how closely the grid can be resolved. Raster images with higher pixel density are desirable in many applications, such as *high resolution* (HR) medical images for cancer diagnosis, high quality video conference, HD television, Blu-ray movies, etc. There is an increasing demand to acquire HR raster images from *low resolution* (LR) inputs such as images taken by cell phone cameras, or converting existing standard definition footage into high definition image/video materials. However, raster images are resolution dependent, and thus cannot scale to arbitrary resolution without loss of apparent quality.

Another generally used image representations is the *vector image*. It represents the visual data using geometrical primitives such as points, lines, curves, and shapes or polygon. The vector image is totally scalable, which largely contrasts the deficiency of raster representation. Hence the idea of vectorizing raster image for resolution enhancement has long been studied. Recently, Ramanarayanan et al. [1] added the vectorized region boundaries to the original raster images to improve sharpness in scaled results; Dai et al. [2] represented the local image patches using the background/foreground descriptors and reconstructed the sharp discontinuity between the two; to allow efficient vector representation for multi-colored region with smooth transitions, gradient mesh technique has also been attempted [3]. In addition, commercial softwares such as [4] are already available. However, vector-based techniques are limited in the visual complexity and robustness. For real photographic images with fine texture or smooth shading, these approaches tend to produce over-segmented vector representations using a large number of irregular regions with flat colors. To illustrate, Fig. 1(a) and (b) is vectorized and grown up to $\times 3$ scale using methods in [2,4]. The discontinuity artifacts in region boundaries can be easily observed, and the over-smoothed texture regions make the scaled image watercolor like.

Alternatively, researchers have proposed to vectorize raster image with the aids of a bases set to achieve higher modeling capacity than simple geometrical primitives. For example, in image/video compression domain, pre-fixed bases, such as the DCT/DWT bases adopted in JPEG/JPEG-2000 standard, and the anisotropic bases such as countourlets [5], have already been explicitly proposed to capture different 2-D edge/texture patterns, because they lead to sparse representation which is very preferable for compression [6]. In addition to pre-fixed bases, adaptive mixture model representations were also reported. For example, the Bandelets model [7] partitions an image into squared regions according to local geometric flows, and represents each region by warped wavelet bases; the primal sketch model [8] detects the high entropy regions in the image through a sketching pursuit process, and encodes them with multiple Markov random fields. These adaptive representations capture the stochastic image generating process, therefore they are suited for image parsing, recognition and synthesis.

In the large body of example-based image resolution enhancement literature, or called "Single Frame Super-Resolution

* Corresponding author.
E-mail address: wangjinjun@gmail.com (J. Wang).

0925-2312/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2011.09.027

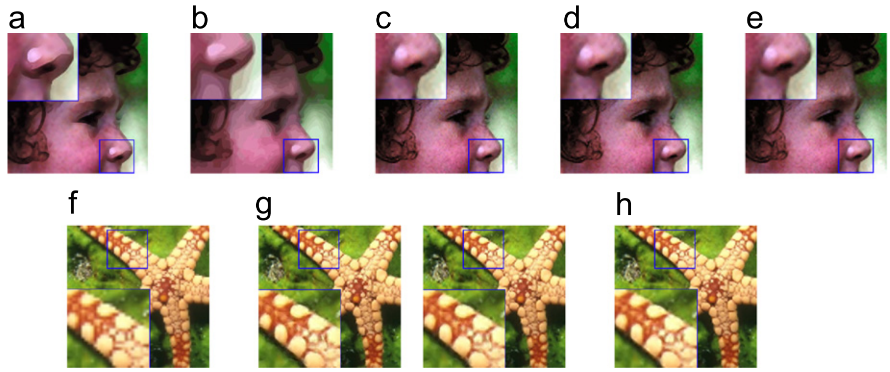

Fig. 1. Image SR quality by our $\mathcal{R}$IIR framework. Top: comparison to image vectorization. Bottom: comparison to different example-based methods.

(SR in short)", researchers utilize the co-occurrence prior between LR and HR representations in an over-completed bases set to "infer" the HR image. For example, Freeman et al. [9] represented each local region in the LR image using one example LR patch, and applied the co-occurrence prior and global smoothness dependence through a parametric Markov Network to estimate the HR image representation. Qiang et al. [10] adopted Conditional Random Field to infer both the missing HR patches and the point-spread-function parameters. Chang et al. [11] utilized locally linear embedding (LLE) to learn the optimal combination weights of multiple LR base elements to estimate the optimal HR representations. In our previous work [12] and in Yang et al.'s work [13], the sparse-coding model is applied to obtain the optimal reconstruction weight using the whole bases set. In addition to example patches, representing images in transferred domain, such as edge profile [14], wavelet coefficients [15], image contourlet [16], etc., has also been examined.

However, although the example-based SR methods significantly improve image quality over 2-D image interpolation, the bases used by existing approaches have only single scale capacity. E.g., the base used for $\times 2$ up-sizing cannot be used for $\times 3$ up-sizing. Hence these existing methods are not capable for multi-scale image SR. To cope with these limitations, this paper presents a novel method that uses example bases set yet is capable for multi-scale and even continuous-scale image resolution enhancement. The contribution includes:

- The paper introduces a novel resolution-invariant image representation ($\mathcal{R}$IIR) framework that models the inter-dependency between example base sets of different scales. The paper shows that, an image can be encoded into a resolution-invariant representation, such that by applying different bases set, the LR input can be enhanced to multiple HRs. This capability has obviously the importance in many novel resolution enhancement applications that existing SR method cannot handel well.
- The key components of the $\mathcal{R}$IIR framework include constructing a $\mathcal{R}$IIR bases set and coding the image into $\mathcal{R}$IIR. In addition to our previous work [12,17], this paper introduces several coding schemes that all possess the resolution-invariant property, as illustrated in Fig. 1(f)–(h). A comprehensive evaluation was conducted to evaluate the advantages of different coding scheme over different aspects.
- The paper further extends the proposed $\mathcal{R}$IIR framework to support continuous scale image SR. A new base for any arbitrary resolution level can be synthesized using existing $\mathcal{R}$IIR set on the fly. In this way the input image can be enhanced to continuous scales using only matrix–vector multiplication, which can be implemented very efficiently by modern computers.

The rest of the paper is organized as follows: Section 2 introduces the image decomposition model and generalizes the invariant property between different image frequency layers. Section 3 introduces our key $\mathcal{R}$IIR framework based on the invariant property between base sets of different scales. Section 4 applies the $\mathcal{R}$IIR framework for continuous image SR. Section 5 lists the experimental results, and Section 6 summarizes the proposed methods and discusses future works.

2. Resolution invariant property between frequency layers

2.1. Image model

Example-based SR approaches assume that [9] an HR image $\mathbf{I}=\mathbf{I}^h+\mathbf{I}^m+\mathbf{I}^l$ consists of a high frequency layer (denoted as $\mathbf{I}^h$), a middle frequency layer ($\mathbf{I}^m$), and a low frequency layer ($\mathbf{I}^l$). The down-graded LR image $\bar{\mathbf{I}}=\mathbf{I}^m+\mathbf{I}^l$ results from discarding the high frequency components from the original HR version. Hence the image super-resolving process strives to estimate the missing high frequency layer $\mathbf{I}^h$ by maximizing $\Pr(\mathbf{I}^h|\mathbf{I}^m,\mathbf{I}^l)$ for any LR input. In addition, since the high frequency layer $\mathbf{I}^h$ is independent of $\mathbf{I}^l$ [9], it is only required to maximize $\Pr(\mathbf{I}^h|\mathbf{I}^m)$, which greatly reduces the variability to be stored in the example set.

A typical example-based SR process works as follows: Given an HR image $\mathbf{I}$ and the corresponding LR image $\bar{\mathbf{I}}'$, $\bar{\mathbf{I}}'$ is interpolated to the same size as $\mathbf{I}$ and denoted as $\bar{\mathbf{I}}$. The missing high frequency layer $\mathbf{I}^h$ can be obtained by $\mathbf{I}^h=\mathbf{I}-\bar{\mathbf{I}}$. A Gaussian filter $\mathbf{G}^l$ is properly defined to obtain the middle frequency layer $\mathbf{I}^m$ by $\mathbf{I}^m=\bar{\mathbf{I}}-\bar{\mathbf{I}}\otimes\mathbf{G}^l$. Now from $\mathbf{I}^h$ and $\mathbf{I}^m$, a patch pair set $\mathcal{S}=\{\mathcal{S}^m,\mathcal{S}^h\}$ can be extracted as the example bases set. $\mathcal{S}^m=\{\mathbf{p}_i^m\}_{i=1}^N$ and $\mathcal{S}^h=\{\mathbf{p}_i^h\}_{i=1}^N$ represent the middle frequency and the high frequency bases respectively. Each element pair in $\{\mathbf{p}_i^m,\mathbf{p}_i^h\}$ is the column expansion of a square image patch from the middle frequency layer $\mathbf{I}^m$ and the corresponding high frequency layer $\mathbf{I}^h$. The dimensions of $\mathbf{p}_i^m$ and $\mathbf{p}_i^h$ are $D^m\times 1$ and $D^h\times 1$ respectively, and often $D^m\neq D^h$. Now from a given LR input, the middle frequency patches can be extracted accordingly and denoted as $\{\mathbf{y}_j^m\}$. The missing high frequency components $\{\mathbf{y}_j^h\}$ are estimated based on the co-occurrence patterns stored in $\mathcal{S}$. The following subsections review three different models for the estimation process.

2.2. Nearest neighbor

Assuming that image patches follow Gaussian distribution, i.e., $\Pr(\mathbf{y}^m)\sim\mathcal{N}(\boldsymbol{\mu}^m,\boldsymbol{\Sigma}^2)$, and $\Pr(\mathbf{y}^h|\mathbf{y}^m)\sim\mathcal{N}(\boldsymbol{\mu}^h,\boldsymbol{\Sigma}^2)$, it can be easily verified that, for any observed patch $\mathbf{y}_j^m$ from the LR input, the maximum likelihood (ML) estimation of $\boldsymbol{\mu}_j^m$ minimizes the