Neural Networks 66 (2015) 64–78

Contents lists available at ScienceDirect

Neural Networks

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

Multi-frame image super resolution based on sparse coding

Toshiyuki Kato [a], Hideitsu Hino [b,*], Noboru Murata [a]

[a] School of Science and Engineering, Waseda University, 3-4-1 Ohkubo, Shinjuku, Tokyo, 169-8555, Japan

[b] Department of Computer Science, University of Tsukuba, 1-1-1 Tennoudai, Tsukuba, Ibaraki, 305-8573, Japan

ARTICLE INFO

Article history:
Received 10 September 2014
Received in revised form 21 January 2015
Accepted 18 February 2015
Available online 9 March 2015

Keywords:
Image super resolution
Multi-frame super-resolution
Sparse coding

ABSTRACT

An image super-resolution method from multiple observation of low-resolution images is proposed. The method is based on sub-pixel accuracy block matching for estimating relative displacements of observed images, and sparse signal representation for estimating the corresponding high-resolution image, where correspondence between high- and low-resolution images are modeled by a certain degradation process. Relative displacements of small patches of observed low-resolution images are accurately estimated by a computationally efficient block matching method. The matching scores of the block matching are used to select a subset of low-resolution patches for reconstructing a high-resolution patch, that is, an adaptive selection of informative low-resolution images is realized. The proposed method is shown to perform comparable or superior to conventional super-resolution methods through experiments using various images.

1. Introduction

Super-resolution (SR) have been receiving a large amount of attention for creating clear images from low-resolution images. In stark contrast to simple picture interpolation techniques, SR methods utilize prior knowledges or assumptions on the structure and relationship of high- and low-resolution images. Development and spread of video equipments drive a growing need for SR techniques, which enable us to create high-resolution (HR) images from low-resolution (LR) images. See, e.g., Borman and Stevenson (1998) and Tian and Ma (2011) for comprehensive surveys.

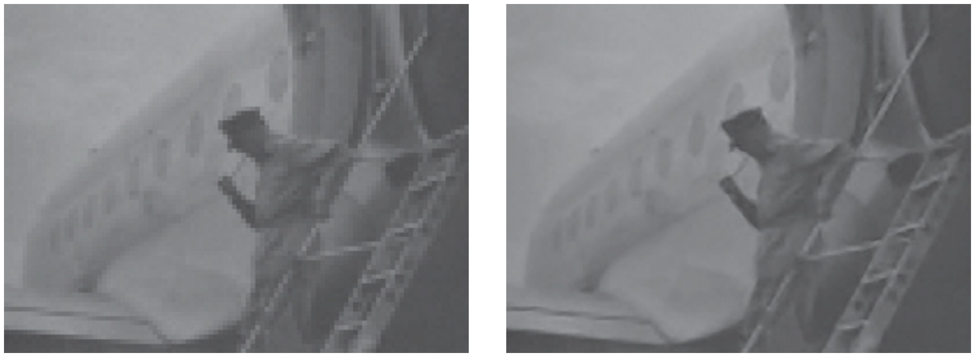
Super-resolution techniques can be divided into two categories, reconstruction-based SR (Farsiu, Robinson, Elad, & Milanfar, 2004; Hardie, Barnard, & Armstrong, 1997; Kanemura, Maeda, & Ishii, 2009), and example-based SR (Chang, Yeung, & Xiong, 2004; Freeman, Jones, & Pasztor, 2002; Kamimura, Tsumura, Nakaguchi, Miyake, & Motomura, 2007; Sun, Zheng, Tao, & Shum, 2003). In the former approach, we compute HR images by simulating the image formation process. It is often used for multi-frame SR, where multiple LR images are used to obtain an HR image. A random Markov field model is usually adopted to represent the relationship between pixels of LR and HR images. This approach is intuitive and natural for SR. It is shown to give favorable results with an appropriate prior (Katsuki, Torii, & Inoue, 2012), but it often requires vast amounts of computation.

Example-based SR aims at inferring HR images based on small image segments extracted from training HR images. In many cases, it is adopted when we can use only one LR image. One of the representative works of example-based SR is the method by Chang et al. (2004), which is based on Neighborhood Embedding (Roweis & Saul, 2000), and actively studied in recent years (Chen & Qi, 2014). Recently, Yang, Wang, Lin, Cohen, and Huang (2012); Yang, Wright, Huang, and Ma (2008, 2010) improved (Chang et al., 2004) and proposed an SR method based on sparse coding (Olshausen & Field, 1996). Sparse coding is a methodology to represent observed signals with combinations of only a small number of basis vectors chosen from a large number of candidates. A set of basis vectors is called *dictionary*. A lot of single frame SR methods based on sparse coding are proposed, and they are experimentally shown to offer favorable HR images. In SR based on sparse coding, correspondence between a dictionary for HR images and a dictionary for LR images is established based on certain assumptions. An LR image is represented by a combination of LR bases, then the coefficients for LR bases are used for combination coefficients of HR bases to obtain the sharpened image. In Yang et al. (2012, 2008, 2010), sparse coding is ingeniously utilized to obtain an HR image from only one LR observation. However, it should be possible to gain further improvements when we have multiple LR images for reconstructing an HR image.

In this paper, we propose a multi-frame SR method based on sparse coding. In general, multi-frame SR methods require

* Corresponding author. Tel.: +81 298 53 5538; fax: +81 298 53 5538.
E-mail address: hinohide@cs.tsukuba.ac.jp (H. Hino).

http://dx.doi.org/10.1016/j.neunet.2015.02.009
0893-6080/



(a) LR frame. (b) HR image.

Fig. 1. (a) An LR frame captured from a movie. (b) The HR image obtained by the proposed SR method.

registration of LR images, that is, we have to estimate relative displacements of LR images. Our proposed method employs a sub-pixel block matching method for image registration. Most of conventional single-frame SR methods based on sparse coding use the joint dictionary learning technique for combining HR and LR image dictionaries. For multi-frame SR with sub-pixel level block matching, it is unrealistic to prepare all correspondences between HR and LR dictionaries for all possible displacements. Hence, in our method, a dictionary for HR images is prepared in advance, and taking into account the image degradation process, a dictionary for LR images is generated from the HR dictionary. When degrading the HR dictionary to LR dictionary, we apply sub-pixel translation. Another contribution is that the proposed method can adaptively select informative LR images for reconstructing HR image, which is also a beneficial side-effect of sub-pixel accuracy block matching. In that sense, the proposed method *adaptively* selects the patches to be used for SR. By thresholding the matching score, the number of LR images used for reconstructing HR image varies in each small patch. When we deal with movies, it is difficult to estimate relative displacements for the regions where objects move quickly, and it is easy to estimate the relative displacements for the regions without such objects. In general, for improving the quality of movies, it is significant to sharpen those objects slowly moving or being at rest, and this property is advantageous to the proposed method. In Fig. 1, we show an experimental result obtained by applying the proposed method to SR from movie frames. We capture five consecutive images from each movie, then their blurred and down-sampled version is used as LR images. Fig. 1(a) is the target LR frame, (b) is the HR image obtained by the proposed method. It is seen that the proposed method sharpens LR movie frame. Details of experimental settings and results will be shown in Section 7.

The rest of this paper is organized as follows. The image observation model is introduced in Section 2, and the sub-pixel accuracy block matching method is briefly explained without technical details in Section 3. The notion of sparse coding is introduced in Section 4. In Section 5, conventional super-resolution methods based on sparse coding is explained, and in Section 6, a novel multi-frame SR method based on sparse coding is proposed. Section 7 shows experimental results, and the last section is devoted to concluding remarks.

2. Image observation model

In this section, we describe the image observation model. Following the idea of Farsiu et al. (2004), we assume a continuous image $\tilde{X}(x, y)$ where $(x, y) \in \mathbb{R}^2$ are coordinate values. Then, we assume that an ideal discrete HR image X and an LR image Y are sampled from the continuous image $\tilde{X}$ according to the following models:

$$X[m, n] = \left[\mathcal{W} \left(\tilde{X}(x, y) \right) \right] \downarrow_X \tag{1}$$

$$Y[m, n] = \left[\mathcal{H} * \mathcal{W} \left(\tilde{X}(x, y) \right) \right] \downarrow_Y + \mathcal{E}[m, n], \tag{2}$$

where $\mathcal{W}$ and $\mathcal{H}$ are warp and blur operators, $\downarrow_X$ and $\downarrow_Y$ are quantization operators to generate HR and LR images, and $\mathcal{E}$ is an additive noise. In this paper, we denote coordinates in the continuous space and the discrete space by (x, y) and $[m, n]$, respectively. The blurring is expressed by the convolution operator $*$.

Following the conventional formulation of super resolution, we treat the HR and LR images as vectors $\mathbf{X} \in \mathbb{R}^{p_h}$, $\mathbf{Y} \in \mathbb{R}^{p_l}$, and the LR observation is assumed to be related with the HR image by

$$\mathbf{Y} = \bar{S}\bar{H}\bar{W}\mathbf{X} + \bar{\boldsymbol{\varepsilon}}, \tag{3}$$

where $\mathbf{Y}$ and $\mathbf{X}$ are vectorized images, the matrix $\bar{W} \in \mathbb{R}^{p_h \times p_h}$ encodes the warping or spacial distortion, the matrix $\bar{H} \in \mathbb{R}^{p_h \times p_h}$ models the blurring effect, $\bar{S} \in \mathbb{R}^{p_l \times p_h}$ is the down-sampling operator, and $\bar{\boldsymbol{\varepsilon}} \in \mathbb{R}^{p_l}$ is the Gaussian noise.

Example-based SR approaches usually extract small patches $\mathbf{x} \in \mathbb{R}^{q_h}$ from the HR image and $\mathbf{y} \in \mathbb{R}^{q_l}$ from the LR image $\mathbf{Y}$. The whole HR image $\mathbf{X}$ is obtained by integrating HR patches $\mathbf{x}$. Each patch pair $(\mathbf{x}, \mathbf{y})$ is connected by the observation model

$$\mathbf{y} = SHW\mathbf{x} + \boldsymbol{\varepsilon}, \tag{4}$$

where $W \in \mathbb{R}^{q_h \times q_h}$, $H \in \mathbb{R}^{q_h \times q_h}$, $S \in \mathbb{R}^{q_l \times q_h}$, and $\boldsymbol{\varepsilon} \in \mathbb{R}^{q_l}$. Hereafter, operators on a whole image are denoted with over-line, and operators on a patch are denoted without over-line. The blurring effects are often modeled by convolution with a point spread function. In this paper, we use a Gaussian filter for the point spread function. The down-sampling process S is assumed to be an impulse sampling. The down-sampled image is further affected by the sensor noise and color filtering noise, and they are assumed to be an additive Gaussian noise. In general, the spacial distortion W includes translation, rotation, deformation and other possible distortions. In this paper, we restrict W to simple translations, which is reasonable when we treat small patches instead of the whole image. In multi-frame SR, we assume that a set of N observed LR images $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_N$ are given. We first choose a *target* LR image $\mathbf{Y}_t$ out of N observed images, and the final output of an SR method is the HR version of $\mathbf{Y}_t$. We refer other LR images as *auxiliary* LR images henceforth. Without loss of generality, we let $t = 1$, i.e., $\mathbf{Y}_1$ is the target image and $\mathbf{Y}_j$, $j = 2, \ldots, N$ are the auxiliary images. The LR observations are related with the HR image $\mathbf{X}$ by

$$\mathbf{Y}_j = \bar{S}\bar{H}\bar{W}_j\mathbf{X} + \bar{\boldsymbol{\varepsilon}}, \quad j = 1, \ldots, N, \tag{5}$$