Signal Processing: *Image Communication* 29 (2014) 514–529

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

Video super resolution based on non-local regularization and reliable motion estimation

Jian Lu, HongRan Zhang, Yi Sun*

Department of Electronic Engineering, Dalian University of Technology, Dalian, Liaoning 116024, China

A R T I C L E I N F O

Article history:
Received 4 August 2013
Received in revised form
14 November 2013
Accepted 6 January 2014
Available online 28 January 2014

Keywords:
Super-resolution
Non-local regularization
Motion estimation
Blur kernel estimation

A B S T R A C T

Video super-resolution (SR) is a process for reconstructing high-resolution (HR) images by utilizing complementary information among multiple low-resolution (LR) images. Accurate estimation of the motion among the LR images significantly affects the quality of the reconstructed HR image. In this paper, we analyze the possible reasons for the inaccuracy of motion estimation and then propose a multi-lateral filter to regularize the process of motion estimation. This filter can adaptively correct motion estimation according to the estimation reliability, image intensity discontinuity, and motion dissimilarity. Furthermore, we introduce a non-local prior to solve the ill-posed problem of HR image reconstruction. This prior can fully utilize the self-similarities existing in natural images to regularize the HR image reconstruction. Finally, we employ a Bayesian formulation to incorporate the two regularizations into one *Maximum a Posteriori* (MAP) estimation model, where the HR image and the motion estimation can be refined progressively in an alternative and iterative manner. In addition, an algorithm that estimates the blur kernel by analyzing edges in an image is also presented in this paper. Experimental results demonstrate that the proposed approaches are highly effective and compare favorably to state-of-the-art SR algorithms.

1. Introduction

In many imaging applications, acquiring an image with high resolution (HR) is hardly possible because of a number of theoretical and practical limitations. To overcome these limitations, super resolution (SR) technology has recently gained widespread interest. The aim of SR is to reconstruct the HR frames from a low resolution (LR) sequence. Although SR has been extensively studied over the past few decades, this method remains an open and challenging topic. In existing systems, SR algorithms generally comprise two parts: (i) registration, where the motion between LR images is estimated, and (ii) image restoration, where the HR image is recovered from LR images using information on motion and blurring.

Both registration and image restoration are generally highly ill-posed problems, especially when the motion parameters are estimated along with the HR image solely from LR images. Therefore, estimation errors are unavoidable in many practical systems. When inaccurately estimated motion is used, it often leads to disturbing artifacts that cause the output to be inferior even when compared to the simple interpolated versions of the given LR observations.

A number of approaches have been proposed to address this problem, which can be classified into two major categories based on the stage where registration is performed. The first class of methods [1–5] starts with a registration step to estimate the motion parameters from the observed LR images, which are used then in a separate HR image restoration process. Given that the motion

* Corresponding author.
E-mail addresses: lu.jian.lu@gmail.com (J. Lu), zhr1988@hotmail.com (H. Zhang), lslwf@dlut.edu.cn (Y. Sun).

0923-5965/$ - see front matter
http://dx.doi.org/10.1016/j.image.2014.01.002

parameters estimated using only the LR images can be unreliable, an algorithm that is robust to outliers and errors in motion estimates is desired. Zomet et al. [2] proposed a robust back-projection method based on median estimators. Farsiuet et al. [3] proposed an observation model based on l_1-norms and image priors grounded on bilateral total-variation functions, the combination of which makes the algorithm robust to motion outliers. Belekos et al. [5] proposed a class of SR algorithms based on the multichannel image prior. Other methods employed regularization by modeling the registration errors as Gaussian noise [6,7]. All methods in this category attempt to reduce the effects of estimation errors and noise by decreasing the weight of unreliable observations in the restoration process. However, these methods did not attempt to correct the errors in the motion estimation process.

Another class of SR methods couples motion estimation with the image recovery process as a joint-estimation task. The most common approach in this category is alternating minimization, where estimates of the HR image and the motion parameters are improved progressively in an alternating manner at each iteration [8–20]. Several methods in this category also employ explicit models of errors in motion estimates. In [14,15], the errors in motion and blur parameters are assumed to follow Gaussian distributions. In [14], the HR image is marginalized from the joint distribution, and the motion and blur parameters are jointly estimated from this marginal distribution. A major disadvantage of this method is that it uses the Gaussian prior as image prior, which over-penalizes strong image edges and therefore reduces the quality of the estimated HR image. In [15], this problem is mitigated by employing a Huber prior to model the HR image. Joint identification methods have been recently proposed in [21–24], where the optimization problem is solved simultaneously for both the HR image and motion parameters. In addition, methods that avoid explicit motion estimation have also been studied in [25–28].

A major issue that most SR methods encounter is that they have to be regularized by some prior knowledge because of their ill-posed nature. The most popular priors are based on the statistical properties of the gradients of images, including Gaussian prior, total variation (TV) prior, and natural image gradient prior [29,30]. These statistical priors are obtained from the statistics of a large number of sample images and do not fully consider the characteristics of the image to be recovered, which usually results in undesirable artifacts in the output result. For instance, the Gaussian prior tends to make the final result over-smoothed, whereas TV prior induces a stair-casing effect, that may produce a disturbing blocky image.

In this paper, we propose a novel SR algorithm within a MAP inference framework, which addresses both of the aforementioned issues. By incorporating a multi-lateral filter in our method, the proposed algorithm can use reliable motion estimates to correct unreliable estimates, thus providing more reliable information on motion fields for HR image reconstruction. The HR image reconstruction is then regularized by using a non-local (NL) similarity prior instead of the popular statistical priors. This NL prior not only reduces undesirable artifacts but also mitigates the effect of inaccurate motion estimation. Finally, the blur kernel estimated by analyzing sharp edges is used to deblur the recovered blurry HR image. The most important advantage of the proposed algorithm is that all priors (for image, motion, and blur kernel) used in the system are obtained directly from the given LR observations and the recovered HR image. In this manner, our algorithm exhibited good self-adaptive capability. Experimental results demonstrate that the proposed methods provide HR images with high quality and accurate motion information compared with the existing SR methods.

The remainder of this paper is organized as follows: Section 2 provides a mathematical model for the LR image acquisition process. In Section 3, we take advantage of a MAP framework for modeling the unknown variables. The inference procedures for the development of the proposed methods, such as image reconstruction with NL prior, reliable motion estimation, and deblurring process, are presented in Sections 4 and 5. Section 6 demonstrates the effectiveness of the proposed methods with the experimental results, and Section 7 concludes the paper.

2. Problem formulation

Super resolution proposes a fusion of several LR images $\{\mathbf{y}_t\}_{t=1}^{T}$ into one higher quality result $\mathbf{x}$ with better optical resolution. The generative model of the LR sequence is illustrated in Fig. 1. This model assumes that $\{\mathbf{y}_t\}_{t=1}^{T}$ is generated from $\mathbf{x}$ through a sequence of operations that includes (i) geometrical warps $\mathbf{F}_t$, (ii) a linear space-invariant blur $\mathbf{H}$, (iii) a decimation step represented by $\mathbf{D}$, (iv) an additive zero-mean white and Gaussian noise $\mathbf{n}_t$, and (v) a match weight matrix $\mathbf{W}_t$ that penalizes outliers (we need to model outliers because the geometrical warps $\mathbf{F}_t$ alone cannot perfectly explain the correspondence between the two frames. An example is a scenario in which one of the LR images in the set is an outlier. In this case, the weights assigned to the pixels in this image will be zeros because the image does not match the reference frame. Therefore, those pixels will effectively be disregarded in the reconstruction process). These are all linear operators, represented by a matrix multiplying the image they operate on. We assume hereafter that $\mathbf{H}$ and $\mathbf{D}$ are identical for all images in the sequence. This model leads to the following equation:

$$\mathbf{W}_t\mathbf{y}_t = \mathbf{W}_t\mathbf{D}\mathbf{H}\mathbf{F}_t\mathbf{x}+\mathbf{n}_t \quad \text{for } t=1,2,\ldots,T. \tag{1}$$

If both $\mathbf{H}$ and $\mathbf{F}_t$ are space-invariant operators, they can be assumed to have a block-circulant structure (assuming a cyclic boundary treatment). In this case, these operators can commute. Thus, by defining $\mathbf{z}=\mathbf{H}\mathbf{x}$, we can convert the SR problem into two sub-problems. First, we estimate the “blurry” HR image $\mathbf{z}$, which is the *reconstruction step*. We then apply a *deblurring step* on $\mathbf{z}$ to obtain the final “clear” HR image $\mathbf{x}$. Although $\mathbf{F}_t$ is not space-invariant in the whole space domain in practice, this operator can still be assumed to be locally invariant. The match weight matrix $\mathbf{W}_t$ will help us to avoid both outliers and regions that are not locally invariant. Thus, according to the derivation in [28], we can commute $\mathbf{H}$ and $\mathbf{F}_t$ in the proposed model