J. Vis. Commun. Image R. 30 (2015) 181–190

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Video super-resolution with registration-reliability regulation and adaptive total variation ☆

Xinfeng Zhang [a], Ruiqin Xiong [b,*], Siwei Ma [b], Ge Li [c], Wen Gao [b]

[a] Key Lab of Intelligent Information Processing, Institute of Computing technology, Chinese Academy of Sciences, Beijing 100190, China
[b] Institute of Digital Media, Peking University, Beijing 100871, China
[c] School of Electronic and Computer Engineering, Peking University Shenzhen Graduate School, Shenzhen 518005, China

ARTICLE INFO

Article history:
Received 28 July 2014
Accepted 5 April 2015
Available online 11 April 2015

Keywords:
Super-resolution
Registration reliability
Regularization
Total variation
Nonlocal similarity
Structure tensor
Interpolation
Optical flow

ABSTRACT

In super-resolution that constructs a high-resolution (HR) image from a set of low-resolution (LR) reference images, it is crucial to align the LR reference images in order to efficiently exploit the pixels therein. However, due to the existence of complex local motion, ideal registration is difficult to acquire. In this paper, we present a robust video super-resolution scheme with registration-reliability regulation and content adaptive total variation regularization, which make the scheme resilient to registration failures. In order to handle ill-registered pixels, we propose a registration-reliability regulated data-fidelity term, which assigns smaller weights to the pixels with larger locally-averaged registration residuals. In addition, a content adaptive total variation based on structure tensor, which is used to estimate image local structures, is proposed to regularize the super-resolved images. The structure tensor is derived not only from the gradients of local patches but also the nonlocal similar patches. Experimental results show that the proposed scheme can remarkably improve both the objective and subjective quality of the video super-resolution results.

1. Introduction

Super-resolution (SR) is a technique to generate high resolution (HR) images from one or multiple low resolution (LR) observation(s). Numerous algorithms have been proposed in the literatures [1–13]. They can be broadly classified into three categories, i.e., interpolation-based, learning-based and reconstruction-based methods. The interpolation-based methods, e.g., in [1–3], usually generate the HR image from only one LR image, utilizing the continuity or smoothness within a small neighborhood of any pixel. However, such assumption of local smoothness tends to blur the edges or high contrast textures in images. The learning-based methods, e.g., in [4–6], utilize a set of example images, organized in a form of low- and high- resolution pairs, to derive the missing high frequency information from the high resolution parts, whose corresponding low resolution parts are similar with input image. This kind of methods is efficient only if the input images are similar to the examples in terms of image structures. However, the performance of these methods usually deteriorates drastically when the input images are not similar to the examples. The reconstruction-based super-resolution algorithms, e.g., in [7–9,13], compute HR images by formulating and inversing the image formation process, using prior knowledge to regularize the solution. Most of reconstruction-based SR methods proposed in the literature consist of the three stages, i.e., image registration, interpolation and restoration (i.e., inverse procedure), where the last two steps are usually implemented jointly. In [11,12], Izadpanahi et al. jointly utilize the super-resolution and edge-directed interpolation to reconstruct high resolution images.

To recover high-frequency information reliably, it is crucial to exploit the relevant pixels in reference images to increase the effective sampling rate. For this purpose, accurate registration (e.g., with sub-pel accuracy) is required to determine the object displacement in a sequence of images, i.e., optical flow, to map the pixels from reference images to the current image. Although many image registration methods [14–18] have been proposed in the literature, video sequence registration is still a very challenging problem, because complex local motion usually exists in real scenes. Traditional SR schemes generally regard all the reference LR images as equally reliable and the different magnitude of registration errors are not taken into consideration. Therefore, undesirable visual artifacts may still appear at the region of complex motion in reconstructed images.

☆ This paper has been recommended for acceptance by Yehoshua Zeevi.

* Corresponding author.
E-mail addresses: xfzhang@jdl.ac.cn (X. Zhang), rqxiong@pku.edu.cn (R. Xiong), swma@pku.edu.cn (S. Ma), gli@pkusz.edu.cn (G. Li), wgao@pku.edu.cn (W. Gao).

http://dx.doi.org/10.1016/j.jvcir.2015.04.002
1047-3203/

To cope with the inaccurate registration problem, the cost function measuring the conformance between the estimated image and the observations is extended to incorporate a set of weights which reflect the registration reliability of reference pixels in the SR process [8,19–22]. In [19], Kondi et al. employ a frame-wise weighting scheme in which LR frame with larger registration residuals is assigned smaller weight. Unfortunately, frame-wise weights cannot reflect the variation of the registration reliabilities within a frame. To address the issue, in [8], pixel-wise weights are proposed in super-resolution reconstruction. However, since the weights are calculated based on the registration residual of only one pixel, they are sensitive to noises or interpolation errors in registration. In [20], Kanaev and Miller apply a Gaussian filter to the single pixel residual based weights to depress the influence of random noise. In the work [21,22], they propose the region-based weight to improve its robustness, in which a weight is computed using all the registration errors in a region and is assigned to each pixel in the local region. The main problem of [21,22] is that the weights cannot reflect the motion difference in the same region, especially when parts of regions are occluded. In addition, it is also a difficult problem of image region segment. In [7], Farsiu et al. take L1 norm to constrain the registered LR images instead of L2 norm to improve the robustness of the SR reconstruction to outliers. In [23], Takeda et al. take a spatial–temporal steering kernel based on the local structures in image and motions between images to estimate high resolution image.

Due to the ill-condition of the super-resolution reconstruction, regularization methods are widely used in solving SR problem. Total variation (TV) [24] is a widely used regularization in image processing, which implicitly assumes the pixel differences follow identical and independent Laplacian distribution. However, the distribution of pixel difference changes for image with different structures, and the spatial invariant TV is unable to adapt different image structures. In [7], Farsiu et al. proposed a bilateral total variation (BTV) by assigning weights to pixel differences according to the pixel similarity. In [25], Yuan et al. designed the weight according to the image difference curvature. This method assigns smaller weights to the pixels around edges and larger weights to the pixels in smooth areas.

In this paper, we propose a robust video super-resolution scheme with a new reconstruction objective function consisting of registration-reliability regulated data-fidelity and content adaptive total variation (CATV) regularization. Firstly, to tackle pixels with different registration accuracy, we propose a registration-reliability regulated data-fidelity to differentiate reference pixels by assigning different weights to them according to their registration residuals. Instead of considering the "lack of fit" of a single pixel, we utilize the weighted registration residuals in a neighborhood to compute the registration reliability of a reference pixel. For pixels with large locally-averaged registration residuals, they may be ill-registered and are assigned to small weights, vice versa. Secondly, to regularize the super-resolved HR image, we propose a content adaptive total variation regularization which penalizes image pixel difference (or image variation) differentially based on the anisotropic local structure of reconstructed image. We take both the local and nonlocal similar patches to calculate structure tensor, which is used to reflect the image local structure. Based on the structure tensor, we assign different weights to pixel differences. Generally speaking, larger weights are assigned to image pixels along edges or in smooth areas, and smaller weights are assigned to those across edges or in texture areas, which is equal to penalizing pixel differences crossing edges weakly.

The remainder paper is organized as follows. In Section 2, we give the formulation of the video super-resolution problem. The proposed registration-reliability regulated data-fidelity is introduced in the Section 3. The proposed content adaptive total variation regularization is elaborated in Section 4. The proposed super-resolution reconstruction method is presented in Section 5. Experimental results on real video sequences are reported in Section 6 and some concluding remarks are made in Section 7.

2. Problem formulation

Super-resolution reconstruction is the inverse problem of low resolution observation formation. Many LR video observation models have been proposed in the literature e.g. [7,26,27]. In this paper, we utilize the observation model that assumes the neighboring HR frames in temporal domain describing the same scene and having complementary information to each other. The LR frames are acquired from the corresponding HR frames through blurring and down-sampling. In this process, the LR frames may be distorted by noise. Fig. 1 shows a LR video observation model, in which the HR frames of size $L_1N_1 \times L_2N_2$ written in lexicographical notation as the vector $\mathbf{X}_t = [x_{t,1}, x_{t,2}, x_{t,3}, \ldots, x_{t,N}]^T$ and the corresponding

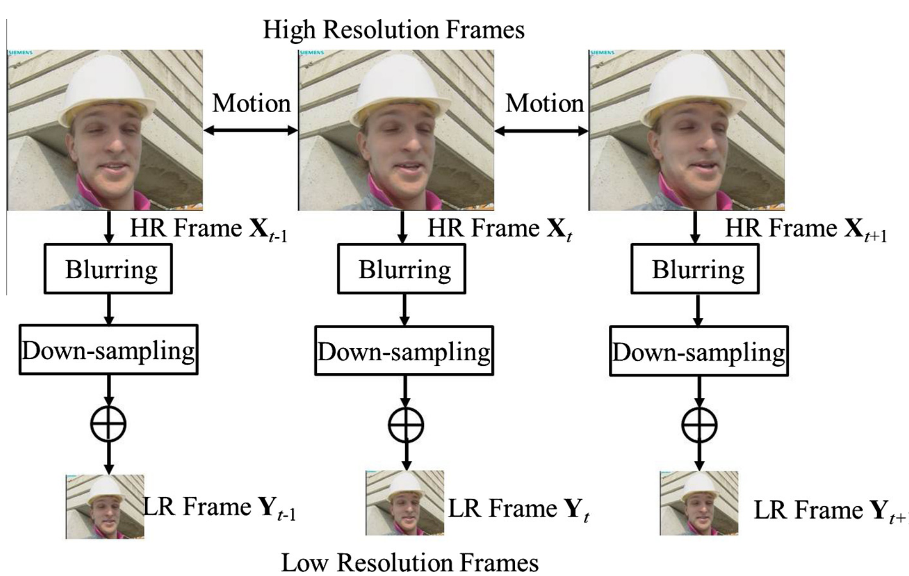

Fig. 1. Video sequence observation model.