Signal Processing 93 (2013) 2076–2086

Contents lists available at SciVerse ScienceDirect

Signal Processing

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

Fast communication

Motion based video super resolution using edge directed interpolation and complex wavelet transform

S. Izadpanahi, H. Demirel *

Department of Electrical and Electronic Engineering, Eastern Mediterranean University, Gazimağusa, via Mersin-10, Turkey

ARTICLE INFO

Article history:
Received 6 July 2012
Received in revised form
17 November 2012
Accepted 10 January 2013
Available online 18 January 2013

Keywords:
Wavelet transform
Interpolation
Super resolution

ABSTRACT

This paper proposes a novel super resolution technique using dual tree complex wavelet transform (DT-CWT) and new edge directional interpolation (NEDI) based on localizing motion blocks in consecutive frames. A super resolution process is applied on the extracted motion blocks and the static region of the frame to generate the low frequency subbands of DT-CWT. DT-CWT decomposition followed by NEDI generates the high frequency subbands in order to obtain the components required for inverse DT-CWT (IDT-CWT). Finally, the super resolved output frame is generated by composing the obtained high and low frequency subbands using IDT-CWT. Experimental results based on PSNR measures illustrate the advantage of proposed technique over the state of the art video resolution enhancement methods.

1. Introduction

High resolution image/video is required in most of the electronic imaging application, since it contains more details that can be critical for that application. An Image processing approach attempts to generate a high resolution (HR) Image from one or more low resolution (LR) version of it. The image/video observation model is employed to relate the desired referenced HR image/frame to all the observed LR images/frames. Usually, the imaging process involves warping, followed by blurring and down-sampling to generate LR images from the HR image. The detailed observation model for video HR reconstruction model is illustrated in Fig. 1. Let the underlying HR image be denoted in the vector form by $z=[z_1, z_2,\ldots, z_{l1N1\times L2N2}]^T$, where $L_1N_1 \times L_2N_2$ is the HR image size. Letting L_1 and L_2 denote the down-sampling factors in the horizontal and vertical directions, respectively, each observed LR image has the size $N_1 \times N_2$. Thus, the LR image can be represented as $y_k=[y_{k,1},y_{k,2},\ldots,y_{k,N1\times N2}]^T$, where $k=1,2,\ldots,P$, with P being the number of LR images. Assuming that each observed image is contaminated by additive noise, the observation model can be represented as

$$y_k = DB_kM_kz + n_k \tag{1}$$

where M_k is the warp matrix with the size of $L_1N_1L_2N_2 \times L_1N_1L_2N_2$, B_k represents the camera blur matrix also of size $L_1N_1L_2N_2 \times L_1N_1L_2N_2$, D is a $N_1N_2 \times L_1N_1L_2N_2$ down-sampling matrix, and n_k represents the $N_1N_{2\times 1}$ noise vector. It should be noted that all the images are assumed to have the same blurring function [1].

There are different approaches to increase the resolution of the blurred, down sampled images such as linear interpolator, adaptive image interpolation algorithms and edge directed interpolation techniques. However, a reasonable approach is to use signal processing techniques to obtain a high-resolution (HR) image from observed multiple low-resolution (LR) images. Such a resolution enhancement approach has been one of the most dynamic study areas, called multi-frame super resolution (SR) [1–4]. Multi-frame super-resolution presents a way out to produce a high-resolution image with finer details, by combining the information in a series of low-resolution frames, with relative sub-pixel shifts. It consists of two main

* Corresponding author. Tel.: +90 533 854 6387;
fax: +90 3923650240.
E-mail address: hasan.demirel@emu.edu.tr (H. Demirel).

0165-1684/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.sigpro.2013.01.006

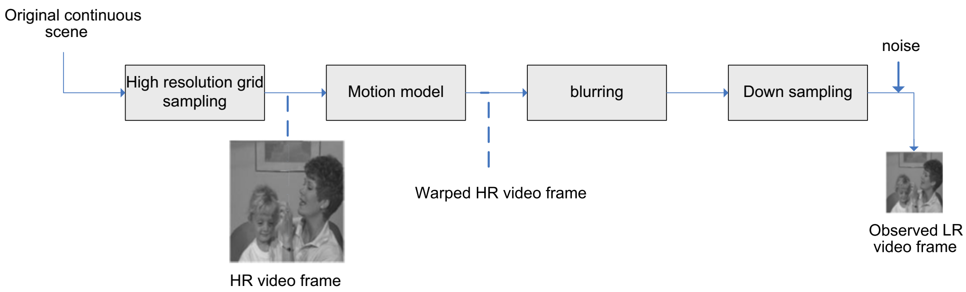

Fig. 1. Observation model for video HR reconstruction [1].

phases: first estimating motion parameters (registration) and then projecting the low resolution images onto a high resolution pattern (reconstruction). However, some algorithms attempt to perform the super resolution, without any explicit motion estimation [5,6]. They are based on the fact that inaccurate motion estimation results in an output image with disturbing artifacts. An example of these kinds of SR is Non-Local-Means SR (NLM-SR) proposed by Protter et al. [5] in which the Non-Local-Means denoising method is generalized in order to achieve an SR algorithm. This is done by modifying an energy function which is defined to clarify the NLM. In [5] the similarity of patches in the image is calculated across space and time that ends up with probabilistic motion estimation. Assigning larger weights to more similar patches in generating the high resolution image is the key of this SR approach. Another SR method based on the development of locally adaptive 3-D filters with no need for motion estimation is demonstrated in [6]. The coefficients of 3-D Local series are approximated by solving a local weighted least-squares problem. In this method the pixel values are estimated using the local behavior of the pixel in its spatiotemporal neighborhood. The comparison of the pixels with their neighborhood produces information about the local motion of the pixels across time. This 3-D kernel regression method neglects the non-local-self similarity and utilizes the local spatial-temporal formation by expanding their 2D spatial kernel regression. The work presented in [7] is also a SR algorithm with probabilistic motion estimation. Under some assumption this method leads to the same method as in NLM-SR [7]. The difference is that, [7] is based on the classic SR and imaging models. Similar to [5] this SR method is formed by employing local image structures and repeating them across the image. The self similarity property obtains a redundancy which is important for many image processing issues. A Non-Local Kernel Regression (NL-KR) based SR is explained in [8]. The method employs local structural regularity and nonlocal similarity for image and video SR. The reason is the possibility of regularizing nonlocal similar pattern fusion using the structural regularity since the redundancy from similar pattern causes more precise approximation of structural regression. The multi-frame SR presented in [9] attempts to recover both edges and flat regions of the LR image. They propose to apply locally adaptive bilateral total variation (LABTV) and consistency of gradient for regularization. LABTV is calculated using the fuzzy-entropy based neighborhood homogeneous measurement to impose the smoothness of the generated image. In parallel, the gradient error increases the resolution of the output image.

Wavelet transform is a tool that divides data or functions or operators into different frequency components, and then studies each component with a resolution matched to its scale. Recently, various wavelet transforms has been applied to image/video super resolution methods in order to improve their performance [10–12]. Estimating the preserved high frequency components from the given images is the strategy of the wavelet-based methods to enhance the resolution of the reference image. Stationary Wavelet Transform (SWT) used in [12] and (Discrete Wavelet Transform) DWT applied in [11] provide three high frequency and one low frequency subbands in each decomposition level. These subbands are constructed by filters with impulse responses corresponding to the directions along 0° (LH), $+45^\circ$ (HH) and 90° (HL) orientations. A one-level dual tree complex wavelet transform (DT-CWT) of a single frame of a video sequence produces complex-valued low frequency subbands and complex-valued high frequency subbands with more directivity (six directions: $+75^\circ$, $+45^\circ$, $+15^\circ$, -15°, -45°, and -75° orientations) in comparison to DWT and SWT [13]. Using DT-CWT which generates multiple frequency subbands with more directions can preserve the integrity of the high frequency components such as edges throughout the proposed SR process. The proposed technique uses DT-CWT to decompose low resolution static and motion regions of the video sequences into different frequency subbands. All the high frequency subbands go through interpolation where low frequency subband is discarded. The input frames are super resolved and used as the low frequency subband in the IDT-CWT procedure of the technique in order to generate the output high resolution image. In this work, Vandewalle et al. [14] registration and Structure Adaptive Normalized Convolution (SANC) [15] reconstruction methods are used to enhance the resolution of different subbands determined by the selected wavelet transform applied in each algorithm. The results of different algorithms applied on six benchmark video sequences (“Akiyo”, “Mother & daughter”, “Container”, “Miss America”, “Suzie” and “Foreman”) in terms of PSNR and SSIM are compared, to confirm the superiority of the proposed method over the conventional resolution enhancement methods.

2. Detection and extraction of the motion blocks

A global displacement occurs when the camera moves and the scene are stationary. In contrast, the scene contains