Signal Processing: Image Communication 68 (2018) 1–12

Contents lists available at ScienceDirect

Signal Processing: Image Communication

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

Spatio-temporal super-resolution for multi-videos based on belief propagation

Tinghua Zhang [a,b], Kun Gao [a,*], Guoqiang Ni [a], Guihua Fan [b], Yan Lu [a]

[a] Key Lab of Photoelectronic Imaging Technology and System, Ministry of Education of China, Beijing Institute of Technology, Beijing, China
[b] Department of Optical and Electronic Equipment, Space Engineering University, Beijing, China

ARTICLE INFO

Keywords:
Super resolution
Spatio-temporal
MAP-MRF
SIFT flow

ABSTRACT

Aiming at improving spatio-temporal resolution of video for real-world applications, this paper focuses on the spatio-temporal super-resolution reconstruction algorithm for multiple asynchronous video sequences. The current multi-video super-resolution algorithms still have some weaknesses in application such as oversimplification of motion, unknown blurring and noise level. For further improvement, a Maximum Posterior Likelihood-Markov Random Field (MAP-MRF) based super-resolution reconstruction method is proposed and proved robust in achieving the real-world super-resolution imaging. The proposed method adopts weighted 3D neighborhood system (WNS) MAP-MRF model to accurately describe the spatial and temporal correlations between multiple video sequences. In order to improve the estimation accuracy of the motions for complex scenes, the improved Scale-Invariant Feature Transform (SIFT) Flow algorithm based on sparsity in wavelet domain is proposed, which can afford the large displacement, rotational movement and other complexities in asynchronous multi-video sequences. The Belief Propagation (BP) algorithm is applied to estimate the parameters of MAP-MRF model, such as motion vector and the super-resolution images in an iterative coarse-to-fine scheme. With the proposed algorithms mentioned above, MAP-MRF based super-resolution reconstruction method has better capabilities of edge sharpness and detailed texture preserving, and robustness of noise suppressing. The experimental result has confirmed the effectiveness of the proposed method under the practical conditions.

1. Introduction

The super-resolution reconstruction (SR) has become the critical aspect in researches about image processing and computer vision, and widely applied in traffic monitoring, remote sensing, live recording, etc. [1–4]. Based on the number of available LR observations, SR algorithms can be generally divided into single image based and multiple-image based algorithms [1].

Multiple image SR algorithms are mostly reconstruction-based algorithms, including single video sequence obtained by one camera and multiple-video sequences obtained by multiple synchronous or asynchronous cameras. The accuracy of multiple image based SR algorithms is highly dependent on the estimation accuracy of the motions between the LR observations, which gets more unstable in real-world applications where different objects in the same scene can have different and complex motions. The single image based SR algorithms can use more meaningful and class wise priors like, e.g., the class of face images [4]. In situations like these, single-image based SR algorithms may work better. However, most existing complex models either generalize hard to diverse natural images or require a lot of time for model training, while simple models have limited representation capability [5]. Aiming at improving spatio-temporal resolution simultaneously, this paper focuses on the temporal–spatial super-resolution reconstruction algorithm of multiple asynchronous video sequences.

Multi video super-resolution algorithms reconstruct high spatio-temporal resolution video by exploiting complementary information in multiple low-resolution video sequences [1], [6]. Its advantage is that it can improve both temporal and spatial resolution simultaneously. In general, the spatio-temporal resolution of images or videos is affected by the imaging hardware and the disturbance of dynamic scenes. With the techniques of video super-resolution reconstruction, a video sequence of higher resolution and better subjective and objective evaluations can be reconstructed in time and space based on the non-redundant information and relativity among the multiple-frame images [7–9].

To deal with the visual artifacts in video sequence, Shechtman et al. proposed a spatial–temporal SR reconstruction algorithm [7], which can suppress the motion blur and aliasing after introducing the concept of

* Corresponding author.
E-mail address: gaokun@bit.edu.cn (K. Gao).

https://doi.org/10.1016/j.image.2018.07.002
Received 5 September 2017; Received in revised form 27 April 2018; Accepted 2 July 2018
Available online 5 July 2018

spatiotemporal kernel and imposing the regular constraints on spatial and temporal dimensions. Furthermore, Oded Shahar [10] and Jordi [11] respectively developed the Shechtman approach and suggested a reconstruction method based on the similarity of spatial–temporal patches to effectively recover the detailed information missing in the video sequence. Yet the algorithm should rely on the self-similarity that exists in the spatio-temporal domain of videos.

At present, the super resolution reconstruction algorithm with regularization is the most applied method for super-resolution multi-video sequences. However, some factors have rendered video super-resolution reconstruction essentially ill-posed, such as the limited observed values, unknown blur kernel and noise. Thus, it requires the regularization term incorporating prior knowledge to obtain stable solution. Köhler T et al. proposed the algorithm based on spatial adaptive Bayes model to reconstruct the images by Bilateral Total Variation (BTV) regularization constraint [12]. The algorithm is robust and remains effective in edge preserving. However, it assumes the blurred kernel already known and regards that some parameters are non-adaptive. An adaptive spatiotemporal regularization method suggests the adaptive selection of the regularization operator and parameters according to the spatiotemporal gradient and local variance [13]. The method can help suppress the noise, preserve the edge, and partially reduce the ghosting effect due to regularization. Another approach [14] of patch group cuts prior depends on the 3D neighborhood to involve more spatial and temporal information in reconstruction. It can afford the target edge and motion trajectory, but weak texture enhancement.

Mudenagudi et al. interpreted the spatiotemporal SR reconstruction as a MAP-MRF question [15], and tried to find solutions by the graph-cut algorithm. However, the algorithm is only applicable in cases with small spatiotemporal deviation. Moreover, to make the cost function satisfactory to the regularity constraint of graph-cut optimization, sophisticated prior cannot be adopted. For the unknown noise and blur level, another approach [16] under MAP-MRF framework has achieved the joint estimation of model parameters in accordance with the gradient sparsity of image, motion vector, noise and blur kernel. This algorithm suits the SR reconstruction of single video sequence. The IRLS (Iteratively Re-weighted Least Squares, IRLS) algorithm is adopted to calculate the parameters, but not very effective for some parameters about motion vector.

The SR reconstruction accuracy of multiple video sequences is highly dependent on the estimation accuracy of the temporal mismatch and inter-frame motions between the LR observations [17]. In practice, there may exist simultaneously the overall mismatch and complex local motions such as multiple targets, large displacement and rotation, posing a high requirement on the robustness of spatiotemporal registration algorithm. Caspi and Irani et al. proposed the two alignment algorithms based on the tracks of the characteristic points [18]. The feature-based alignment algorithm adopts the KLT feature tracker to follow the characteristic point, or detects and traces the mass center of moving objects. The parameters of spatiotemporal transformation can be worked out among the video sequences according to the trajectories of the characteristic points [19]. The effectiveness of the algorithm depends on the detecting and tracking of the spatiotemporal features. It is probable to generate spatiotemporal ambiguities.

The direct alignment algorithm can cope with complicated dynamic scenes such as non-rigid transformation, large displacement and illumination changes. However, the algorithm may cause time aliasing and space resolution lowering, when dealing with the temporal and spatial values separately. The joint spatiotemporal registration algorithm [20] converts the video registration into a multi-parameter extremum problem characterizing time correspondence and geometric transformation. It involves the considerations on the similarity of temporal and spatial dimensions, and the precondition that the videos share the same scene and only have visual differences from one another. To avoid high precision motion estimation, the SKR (Steering Kernel Regression) algorithm has been mentioned to preserve and rebuild the details [21,22]. Given the heavy computation about the patch matching and spatial blur kernel, this algorithm still needs the optical flow method for initial motion estimation and compensation in case of large displacement.

The above analysis shows that the main challenge is the spatiotemporal registration and the ill-posed nature of reconstruction algorithms for real-world video super-resolution reconstruction [16,23]. The existing spatiotemporal registration algorithms are based on the concept of spatiotemporal blur kernel and can only contain the relatively simple motions of the target scene. They hardly adapt to the complicated real-world video with motion aliasing, large displacement, rotation, significant illumination changes, etc. Moreover the effects of noise and blur kernel are always omitted in current algorithms; or the kernel and noise level are assumed already known in the image degradation model. Therefore, the reconstruction of real scene often fails the expectation.

To solve the problem, a MAP-MRF SR reconstruction model for multiple videos is proposed in this paper. It is based on WNS and adopts the multi-scale hierarchical iteration and the coordinate descent method to perform joint estimation on model parameters, so as to achieve the adaptive spatiotemporal SR reconstruction of real-world scenes. Fig. 1 shows the algorithm framework.

In the MAP-MRF model, the weighted spatiotemporal neighborhood system (WNS) accurately describes the spatial and temporal correlations among video sequences, and thus simplifies the complex spatial–temporal registration into the question about spatial registration and motion estimation. Hence it is possible to apply the established algorithms for motion estimation and image registration to the spatiotemporal SR reconstruction of real scenes. The image gradient regularization is adopted to keep the sharp edges, and the algorithm of belief propagation to prevent ringing or jagged artifacts.

Considering SR reconstruction as a MAP-MRF problem, the model conducts the iterative estimation of model parameters such as motion vector and the SR image sequences by belief propagation under the iterative coarse-to-fine scheme. The motion-vector weighting method is applied to reflect the contribution of different spatiotemporal correlations to SR reconstruction, and the method of message length estimation is used to promote the convergence rate.

A modified SIFT Flow algorithm for motion estimation is suggested herein to overcome the difficulties with multiple asynchronous videos, such as large displacement, rotational movement and significant lighting changes, on the basis of unchanged image sparsity for rotation and shift in wavelet domain.

In this paper, we propose a MAP-MRF super-resolution reconstruction based on BP and proved robust in achieving the real-world super-resolution imaging. The proposed weighted 3D neighborhood system MAP-MRF model adopts the weighted total-variation regular terms to describe the spatial and temporal correlations between multi-video sequences. For the estimation on motion vector, the modified SIFT Flow algorithm based on wavelet domain sparsity model can afford the large displacement, rotational movement and other complexities in asynchronous multi-video sequences. The experimental result has confirmed the effectiveness of the proposed method under the practical conditions.

The paper has been organized into the following sections. Section 2 introduces the forward model and Bayesian model for multiple videos super-reconstruction; Section 3 introduces the proposed MAP-MRF based spatio-temporal super-resolution reconstruction method utilizing inter-frame redundant information; Section 4 describes the implementation of the proposed SR algorithm; and Section 5 involves the comparison and discussion of the proposed approach. Finally we conclude this paper in Section 6.

2. Bayesian model of multiple video super resolution reconstruction

In the case of multiple-video SR model, the observation process of each LR video sequence from a dynamic scene contains space–time