Registration-Reliability based Strategy to Enhance Multi-Frame Super-Resolution Algorithms

Qiang Song[1], Ruiqin Xiong[1], Xiaopeng Fan[2], Siwei Ma[1], Wen Gao[1]

[1]Institute of Digital Media, Peking University, Beijing 100871, China
[2]Department of Computer Science, Harbin Institute of Technology, Harbin 150001, China
Email: {songqiang, rqxiong, swma, wgao}@pku.edu.cn, fxp@hit.edu.cn

Abstract— **Image registration plays an important role in most of multi-frame super-resolution methods. As far as we know, the accuracy of most registration algorithms is not enough for super-resolution, which will lead to annoying artifacts. This paper proposes a simple but effective strategy that aims to enhance the performance of existing super-resolution methods. The idea is to measure the reliability of the estimated shifts and only choose the reliable frames to reconstruct a coarse high-resolution (HR) image. This coarse HR image helps to refine all the shifts to finally get a refined HR image. An iterative contour smoothing filter is proposed to improve the accuracy of this refining process. Experimental results demonstrate that the proposed algorithm can help to improve the performance of the existing super-resolution methods with fewer artifacts.**

Index Terms ── **Super-resolution, registration-reliability, coarse HR image, prior model, iterative contour smoothing filter**

I. INTRODUCTION

Multi-frame super-resolution (SR) has been studied widely over the past two decades. It aims to produce one high-resolution (HR) image from multiple low-resolution (LR) frames. SR methods are generally based on the assumption that LR frames are sampled with shifts from the same HR scene and thus can complement each other with missing HR scene information. The degrade model can be formulated as:

$$\mathbf{y}_k = \mathbf{D}_k \mathbf{H}_k \mathbf{S}_k \mathbf{x} + \mathbf{n}_k \quad , \quad k = 1, 2 \ldots, N \qquad (1)$$

where $\mathbf{y}_k$ and $\mathbf{x}$ are the lexicographically rearranged matrix of the k*th* LR image and the original HR image, respectively. $\mathbf{S}_k$ stands for the warp matrix. $\mathbf{H}_k$ is the blurring matrix. $\mathbf{D}_k$ is the down-sampling matrix, and $\mathbf{n}_k$ is the Gaussian noise vector.

Reconstruction-based super-resolution methods include non-uniform interpolation approach [1], projection onto convex sets approach [2], maximum a posteriori (MAP) approach [3] [4], joint MAP approach [5]. These methods rely on the assumption that $\mathbf{D}_k$ and $\mathbf{H}_k$ are known while $\mathbf{S}_k$ should be estimated from the LR images. So image registration techniques are employed firstly to obtain reliable sub-pixel shift information. However, due to the precision of the registration methods, unavoidable shift estimation error will lead to annoying artifacts in super resolved HR frame as shown in Fig. 1. To reduce the artifacts, [6] proposed a

motion-estimation-free algorithm based on Nonlocal-means. But this method can't always get good results because its performance is largely related to the number and quality of the similar blocks that found and its computation complexity is very high.
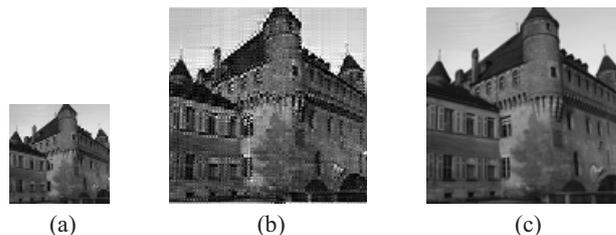

(a)   (b)   (c)

Fig. 1. Artifacts caused by shift estimation error: (a) LR frame (b) reconstructed HR frame using badly estimated shifts (c) reconstructed HR frame using true shifts

In multi-frame super-resolution, a group of images should be registered with respect to each other. However, most of the existing methods are designed to estimate the relative shift between two images at a time. So the alternative approach is to align all the frames with respect to a single reference frame. This methodology destroys the intrinsic group structure and the choice of reference frame can have a severe effect on the overall accuracy of the estimates.

This paper proposes a strategy to enhance multi-frame super-resolution algorithms. As the choice of the reference frame can influence the accuracy of estimation, all the frames will have the chance be selected as reference frame to finally form a shift set. The prior information for the multi-frame shift estimation problem is exploited to project the shift set to a unified coordinate system. Every frame in such coordinate system has several estimating samples consequently. Based on these samples, reliability of the estimations can be computed. A coarse HR frame which is reconstructed only by the reliable frames will then help to improve the accuracy of all the shifts and a refined HR frame is obtained finally. Experimental results show that the proposed strategy can improve the performance of existing super-resolution method significantly.

This paper is organized as follows. In the next section, we briefly review the general multi-frame super-resolution methods and the prior models for the multi-frame registration problem. In section 3, we will describe our registration-reliability based scheme in detail: the theoretical analysis and the framework of the method will be given. Experimental results are shown in Section 4 and Section 5 ends the paper with a few remarks.

II. PREVIOUS WORKS

A. General Multi-Frame Super-Resolution Methods

Assuming that the noise of each LR image is zero-mean Gaussian noise, and each LR frame is independent, using the model in (1), the maximum a posteriori probability (MAP) estimation proposes the following regularization-based least-squares problem:

$$\hat{\mathbf{x}} = \arg\min \sum_{k=1}^{N} \left\| \mathbf{D}_k \mathbf{H}_k \mathbf{S}_k \mathbf{x} - \mathbf{y}_k \right\|_p^p + \lambda \cdot R(\mathbf{x}) \qquad (2)$$

where $\sum \left\| \mathbf{D}_k \mathbf{H}_k \mathbf{S}_k \mathbf{x} - \mathbf{y}_k \right\|_p^p$ is the data fidelity item, $R(\mathbf{x})$ is the regularization, which gives a prior model of the HR image such as spatial smoothness and sparsity. λ is the regularization parameter and $p = 1, 2$. In practice, $\mathbf{D}_k$ and $\mathbf{H}_k$ are identical for all the LR images and are already known in advance.

B. Prior Models for The Multi-Frame Shift Estimation

Assuming we have a sequence of N frames, there are some intrinsic relationships between their shifts as described in [7]. Here we call them multi-frame shift estimation priors.

The first prior dictates that the relative shifts between any pair of frames must be the composition of the relative shifts between two other pairs of frames. More clearly, we have the following description:

$$M_{i,j} = M_{i,k} + M_{k,j}, \quad \forall i, j, k \in \{1,\dots N\} \qquad (3)$$

$M_{i,j}$ means the relative shift between the ith frame and the jth frame. For global shift, $M_{i,j}$ is $\left[\Delta x, \Delta y \right]^T$. In the following section, we will use this definition if there is no special statement.

The second consistency condition prior states that the composition of the relative shifts $M_{i,j}$ and $M_{j,i}$ should be opposite number:

$$M_{i,j} = -M_{j,i}, \quad \forall i, j \in \{1,\dots N\} \qquad (4)$$

We should note that $M_{i,i} = 0$. In case of several frames to be registered, the two intrinsic relationships should be taken into consideration.

III. REGISTRATION-RELIABILITY BASED MULTI-FRAME SUPER-RESOLUTION

As we originally know, more frames can complement more missing information in multi-frame super-resolution. Consequently, general super-resolution methods prefer to use all the LR frames in reconstruction. However, this statement is tenability only when the estimated shifts are true or reliable.
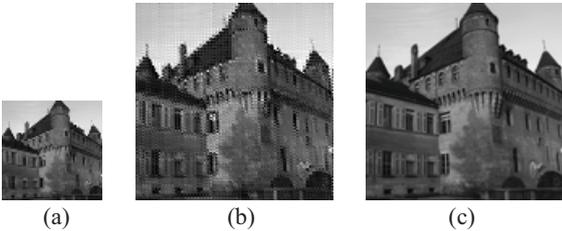

(a)   (b)   (c)

Fig. 2 Example: (a) LR frame (b) reconstructed HR frame using five LR frames (c) reconstructed HR frame using three LR frames

As illustrated in Fig. 2, five LR images (three with true shifts and two with badly estimated shifts) are used in the SR reconstruction (scale=2). Fig. 2(b) shows the result using all of the five LR frames and Fig. 2(c) is the result reconstructed only by the three LR frames with true shifts. We can see that the frames with unreliable estimated shifts might not improve the quality of the HR image but even bring about annoying artifacts.

A. Reliability of The Estimated Shifts

Given a set of N frames, there will be N^2 pair-wise relative shifts for this set. Every frame in the set will be aligned with other N frames (including itself):

$$\left\{ M_{i,j} \mid \forall i, j \in 1, \dots N \right\} \qquad (5)$$

From another point of view, the above expression is equal to that every frame in the set can be a reference frame and other frames are aligned to it. Then we have the following equal expression:

$$\textit{Reference frame } i \in \{1, \dots N\} : \left\{ M_{i,j} \mid \forall j \in 1, \dots N \right\} \qquad (6)$$

i means that the ith frame is chosen as the reference frame and j means that the jth frame is registered with respect to the reference frame.

Putting all these N^2 relative shifts into one matrix, we have:

$$\mathbf{M} = \begin{bmatrix} M_{1,1} & M_{1,2} & \cdots & M_{1,N} \\ M_{2,1} & M_{2,2} & \cdots & M_{2,N} \\ \vdots & \vdots & \vdots & \vdots \\ M_{N,1} & M_{N,2} & \cdots & M_{N,N} \end{bmatrix} \qquad (7)$$

The ith row in the matrix contains all the relative shifts with respect to the reference frame i. The ith column means that frame i is registered to all the frames (including itself). If we take the multi-frame shift estimation prior model into consideration, $M_{i,j} = -M_{j,i}$ and $M_{i,i} = 0$, (7) can be rewritten as:

$$\mathbf{M} = \begin{bmatrix} 0 & M_{1,2} & \cdots & M_{1,N} \\ -M_{1,2} & 0 & \cdots & M_{2,N} \\ \vdots & \vdots & \vdots & \vdots \\ -M_{1,N} & -M_{2,N} & \cdots & 0 \end{bmatrix} \qquad (8)$$

The matrix $\mathbf{M}$ is an anti-symmetric matrix and it has some particular properties such as $\mathbf{M}^T = -\mathbf{M}$. This can be a new global prior for the multi-frame shift estimation problem. For this paper, we will use this prior in another way.

We use elementary transformation on $\mathbf{M}$ so as to transform all the relative shifts into a unified coordinate system. Here we use the coordinate system of 1st frame in the set. We combine this operation with the earlier mentioned prior $M_{i,j} = M_{i,k} + M_{k,j}$ and get the following formation:

$$\mathbf{M}_t = \begin{bmatrix} 0 & M_{1,2} & \cdots & M_{1,N} \\ -M_{1,2}+M_{1,2} & M_{1,2} & \cdots & M_{2,N}+M_{1,2} \\ \vdots & \vdots & \vdots & \vdots \\ -M_{1,N}+M_{1,N} & -M_{2,N}+M_{1,N} & \cdots & M_{1,N} \end{bmatrix} = \begin{bmatrix} 0 & M_{1,2} & \cdots & M_{1,N} \\ 0 & M_{1,2} & \cdots & M_{1,N} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & M_{1,2} & \cdots & M_{1,N} \end{bmatrix} \qquad (9)$$

As illustrated in (9), every row in the shifts matrix $\mathbf{M}$ is an estimated version of the same relative shifts registered with

respect to the first frame in the set. In other words, we get N estimated samples of the same relative shifts.

Based on the above theoretical analysis, for each column, we can estimate the reliability of the pair-wise shift $M_{1,j}$. Here we use the standard deviation as a metric of the reliability:

$$\Upsilon(M_{1,j}) = 1/STD(\mathbf{M}_t(:,j)) \qquad (10)$$

$\Upsilon(\cdot)$ means the reliability of the estimated shift. $STD(\cdot)$ is an standard deviation operator and $\mathbf{M}_t(:,j)$ means the jth column of the matrix $\mathbf{M}_t$. From (10), we can see that frames with larger standard deviation correspond to lower reliability.

As described above, we can easily know which relative shifts are reliable so as to decide which frames can be used in the computation. A ratio is set to determine the number of included frames (empirically 70 percent of the most reliable frames). For the selected frames, the median value of all the estimated samples is the final estimated shifts.

B. Coarse-to-Fine Multi-Frame Super-Resolution

In this section, we will describe the coarse-to-fine enhancement strategy. For the first step, we use the selected reliable frames to construct a coarse version of the HR image based on the methodology of (2):

$$\hat{\mathbf{x}}_{coarse} = \arg\min \sum_{k=1}^{S} \|\mathbf{DHS}_k \mathbf{x} - \mathbf{y}_k\|_p^p + \lambda \cdot R(\mathbf{x}) \qquad (11)$$

S means there are S frames included in the computation.
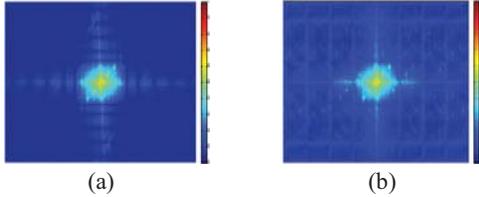

(a)   (b)

Fig. 3. Comparison of Frequency spectrum: (a) HR frame (ground truth) (b) reconstructed HR frame with shift estimation error

As soon as we get $\hat{\mathbf{x}}_{coarse}$, we use it to refine all the shifts. Although most of the artifacts have been controlled, the remaining artifacts will reduce the accuracy of the refining operation. As illustrated in Fig.1 and Fig.3, these artifacts can be treated as abnormal high-frequency and cause what should be smooth contours to be jagged in spatial domain. To keep the contours smooth, a directional filter is desired to filter the image along the direction of the contours while preserving most of the textures. Motivated by the Level-set reconstruction method [8], we propose an iterative contour smoothing filter F_β. For each iteration, we have:

$$\mathbf{x}^{(i+1)} = F_\beta(\mathbf{x}^{(i)}) = \mathbf{x}^{(i)} + \beta \cdot \mathbf{x}_e^{(i)} \qquad (12)$$

β controls the smoothness of the result and $\mathbf{x}_e^{(i)}$ is the curve flow as described in [8]:

$$\mathbf{x}_e^{(i)} = \frac{((\mathbf{x}^{(i)})_x)^2 (\mathbf{x}^{(i)})_{yy} - 2(\mathbf{x}^{(i)})_x (\mathbf{x}^{(i)})_y (\mathbf{x}^{(i)})_{xy} + ((\mathbf{x}^{(i)})_y)^2 (\mathbf{x}^{(i)})_{xx}}{((\mathbf{x}^{(i)})_x)^2 + ((\mathbf{x}^{(i)})_y)^2}$$

As we have the hypothesis that the LR image is sampled from the integer pixel location of HR image, the refinement can be simplified using fast matching method such as exhausting all possible shifts near the location of the previous estimation. More accurate, the relative shifts can be computed by solving the following object function:

$$\hat{M}_{1,k} = \arg\min \|\mathbf{D} \cdot round(scale \cdot M_{1,k}) \cdot \hat{\mathbf{x}}_{coarse} - \mathbf{y}_k\|_2^2 \qquad (13)$$

Then we accept all the frames in the final computation and the final refined HR image $\hat{\mathbf{x}}$ is obtained.

IV. EXPERIMENTS

In this section, experiments are performed on test images to show the performance of the proposed method. For every experiment, the test image is firstly shifted, blurred, and then down-sampled to generate several LR frames. We employ the Gaussian blur kernel with width of 3*3 and variance of 0.5. The shifts are randomly generated using MATLAB function *randn*. In this paper, we focus on the decimation factor of 4 and use 18 LR frames to reconstruct the HR image.

A. Intermediate Results

In this section, we will show the results step by step to demonstrate the effectiveness of the proposed method.

After the initial shift estimation is obtained (method [9] is used), the reliability of the estimates is calculated. Based on the reconstruction method in [10], we then use 70 percent of the most reliable LR frames to reconstruct the coarse HR image (See Fig. 4(b)). This coarse HR image is filtered (See Fig. 4(c)) and used to refine all the estimated shifts. β is empirically set to 0.05 and the total number of iterations is 10. The refined HR image is obtained finally using all the LR frames (See Fig. 4(d)). In addition, comparable result without using the proposed strategy is also shown in Fig. 4(e).

As illustrated, most of the artifacts in the coarse HR image have been controlled compared with Fig. 4(e). The contours of the filtered HR image become much smoother after being filtered. The refined HR image has much fewer artifacts compared with Fig. 4(e) and the quality is better than Fig. 4(b) and Fig. 4(e). All the results prove that the proposed scheme can improve the performance of the multi-frame super-resolution algorithms.

B. Results with/without The Proposed Scheme

In this part, experiments based on different reconstruction methods are performed to show the universality of the scheme. The five original test images are named: 'a', 'b', 'c', 'd', 'e' and of the same size 248*248. Three reconstruction methods [10] [11] and [12] are used and we choose the empirical parameter settings, which give the best performance for the images. For every test image, the shifts are the same for all the reconstruction method. Table I shows the PSNR results of the three reconstructed methods with or without the proposed scheme and Fig. 6 and Fig. 7 shows the results of image 'a' and 'd' based on method [10] and [12] respectively. Due to the limited space, only these images are shown here. We can see that the proposed approach offers better HR images.

V. CONCLUSIONS

This paper proposes a novel registration-reliability based super-resolution method. The proposed strategy discusses the
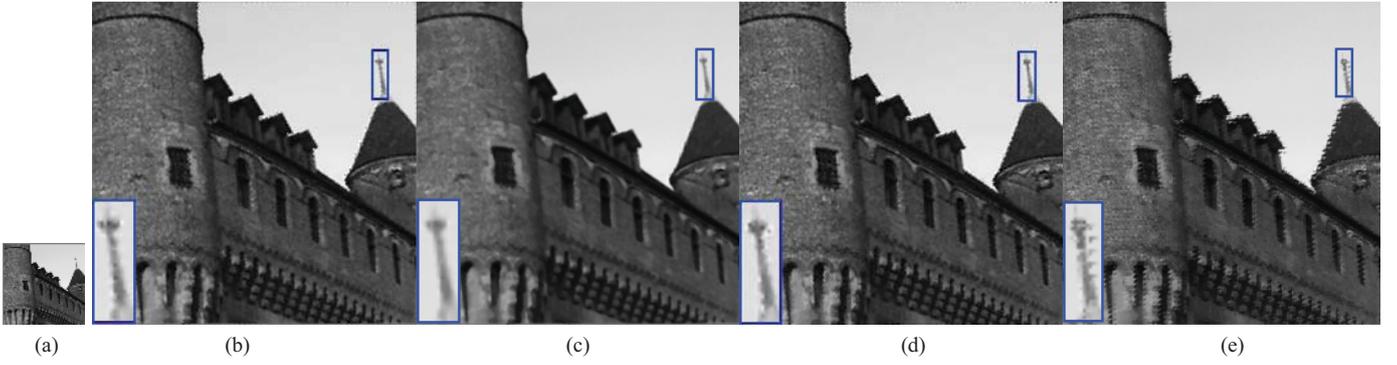
Fig. 4 Intermediate results: (a) LR frame (b) Coarse HR image (22.04dB) (c) Filtered HR image (d) Refined HR image (22.41dB) (e) method[10] (20.92dB)
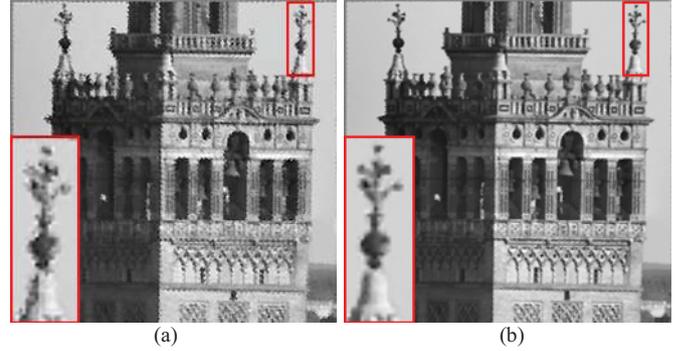

Fig. 5 Test images of Experiment B

Fig. 7 Results of image 'd':(a)method [12](16.762dB) (b)proposed (18.075dB)

TABLE I PSNR RESULTS OF THE RECONSTRUCTED HR IMAGES BY DIFFERENT METHODS

	[10]	Proposed	[11]	Proposed	[12]	Proposed
a	18.918	**21.027**	19.102	**21.114**	19.459	**20.926**
b	18.414	**18.516**	18.480	**18.522**	18.471	**18.524**
c	16.401	**16.468**	16.531	**16.599**	16.453	**16.578**
d	16.196	**17.918**	16.758	**17.858**	16.762	**18.075**
e	15.390	**16.173**	15.632	**16.156**	15.182	**16.212**

multi-frame registration problem in a global view and proposes a new method to measure the reliability of the relative shifts. Instead of using all the LR frames in hand, we use part of the frames that are more reliable to reconstruct a coarse HR image. The coarse HR image is helpful to refine the accuracy of all the shifts and a refined HR image is got finally. Experimental results show that the proposed approach can help to improve the performance of the existing super-resolution method significantly.
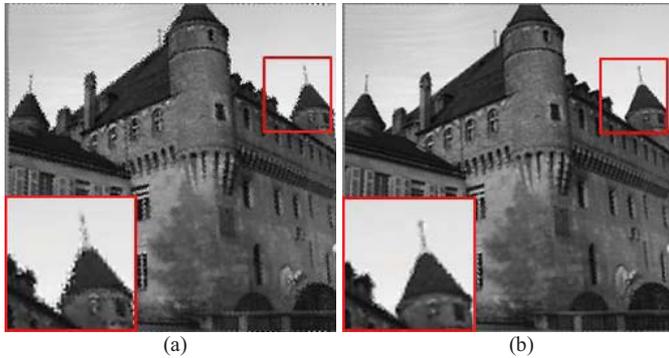

Fig. 6 Results of image 'a':(a)method[10] (18.918dB) (b)proposed (21.027dB)

REFERENCES

[1] M. Alam, et al. "Infrared image registration and high-resolution reconstruction using multiple translationally shifted aliased video frames," IEEE Trans. Instrum. Meas, vol. 49, pp. 915–923, Oct. 2000.

[2] A. Patti, M. Sezan, and A. Tekalp, "Super resolution video reconstruction with arbitrary sampling lattices and nonzero aperture time," IEEE Trans. Image Process, vol. 6, no. 8, pp. 1064–1076, 1997.

[3] S. Belekos, N. Galatsanos, and A. Katsaggelos, "Maximum a posteriori video super-resolution using a new multichannel image prior," IEEE Trans. Image Process., vol. 19, no. 6, pp. 1451–1464, Jun. 2010.

[4] Xinfeng Zhang, et al. "Video super-resolution with registration-reliability regulation and adaptive total variation." Journal of Visual Communication and Image Representation 30 (2015): 181-190.

[5] Wang Qing, and Xiaoli Song. "Joint image registration and super-resolution reconstruction based on regularized total least norm." Image Processing (ICIP), IEEE International Conference on, 2009.

[6] M. Potter, M. Elad, H. Takeda, and P. Milanfar. Generalizing the Nonlocal-Means to Super-Resolution Reconstruction, IEEE Transactions on Image Processing, vol. 19, no. 1, pp. 36-51, 2009.

[7] Farsiu, Sina, Michael Elad, and Peyman Milanfar. "Constrained, globally optimal, multi-frame motion estimation." Statistical Signal Processing, IEEE/SP 13th Workshop on. IEEE, 2005.

[8] B. Morse and D. Schwartzwald, "Image magnification using level-set reconstruction," in Proc. IEEE Conf. Computer Vision Pattern Recognition, vol. 1, 2001, pp. 333–340.

[9] D. Keren, S. Peleg, and R. Brada, "Image sequence enhancement using sub-pixel displacements," IEEE Conference on Computer Vision and Pattern Recognition, pp. 742–746, Ann Arbor, Mich, USA,June 1988.

[10] S. Farsiu, M. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super-resolution," IEEE Trans. Image Process., vol. 13, no. 10, pp. 1327–1344, Oct. 2004.

[11] Q. Yuan, L. Zhang, and H. Shen, "Multiframe super-resolution employing a spatially weighted total variation model," IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 3, pp. 379 –392, march 2012.

[12] J. Zhang, D. Zhao, R. Xiong, S. Ma, and W. Gao, "Image restoration using joint statistical modeling in a space-transform domain," IEEE Transactions on Circuits and Systems for Video Technology, vol. 24, no. 6, pp. 915-928, June 2014.