Scale-Compensated Nonlocal Mean Super Resolution

Qiaochu Li, Qikun Guo, Saboya Yang and Jiaying Liu*
Institute of Computer Science and Technology, Peking University, Beijing, P.R.China, 100871

***Abstract*—In this paper, we propose a novel algorithm for multi-frame super resolution (SR) with consideration of scale changing between frames. First, we detect the scale of each frame by *scale-detector*. Based on the scale gap between adjacent frames, we extract patches and modify them from different scales into the same scale to obtain more redundant information. Finally, a reconstruction approach based on patch matching is applied to generate a high resolution (HR) frame. Compared to original Nonlocal Means SR (NLM SR), the proposed *Scale-Compensated NLM* finds more potential similar patches in different scales which are easily neglected in NLM SR. Experimental results demonstrate better performance of the proposed algorithm in both objective measurement and subjective perception.**

I. Introduction

Multi-frame super resolution method intends to reconstruct a HR frame from a series of low resolution (LR) frames. It is based on an assumption that a large amount of redundant information exists in LR frames and they complement each other. Therefore, the key to multi-frame SR is to obtain precise and high quantity redundant information from the LR images.

Many researchers focus on improving motion estimation to get more precise locations of redundant information, such as Tanaka *et al*. [1]. Correct direct motion estimation assures the precision, but there are unavoidable motion estimation errors because scenes of videos are various and complex. Several wrong estimations severely degrade HR frames compared to slight imprecise estimations in a large domain, which is the bottleneck of the direct motion estimation methods. Potter *et al*. [2] proposed a method, NLM SR, free of explicit motion estimation enlightened by NLM denoising algorithm. NLM SR estimates the similarity of patches in the neighborhood, which reflects theirs possible motion estimation, and the final motion estimation is a weighted average from many possible motions, so it avoids severe motion estimation errors.

Recently, some new methods inspired by NLM SR have been proposed. Zeng *et al*. [3] separated an image into various regions and processed them with adaptive methods. Viewing an image as a signal and processing it in the frequency domain, Zheng *et al*. [4] combined wavelet theory with NLM and proposed a new method, wavelet-based nonlocal means (WNLM). Considering adaptive parameters in NLM, Cheng *et al*. [5] improved NLM SR by mobilizing its search window and adjusting block size adaptively. Our previous work [6] took rotation-invariance and search window into account simultaneously and proposed an SR reconstruction approach, Adaptive Rotation Invariance and Search Window Relocation (ARI-SWR) algorithm.

All methods mentioned above only have considered translation and/or rotation between adjacent frames. They neglect the zooming which changes scales of objects. Zooming caused by camera motion and objects motion is considered ubiquitous in videos. Regardless of scale changing, patches extracted from a same scale can not match flawlessly. Fig. 1 shows the scale changing effect in adjacent frames. Glanser *et al*. [7] took scale changing into account in single-frame SR. But their approach is blind to the explicit changing scales and fails to handle frames in arbitrary scales because LR frames are decimated by fixed scale factors. Furthermore, only the higher scale patch in a mapping relationship of the patches is taken into the reconstruction in their approach. Instead, an algorithm should make full use of all the redundant information from LR frames in multi-frame SR.

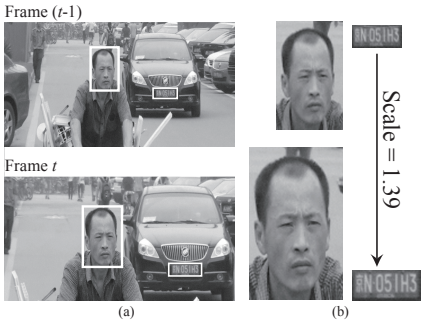

Fig. 1. Scale changing effects in adjacent frames. (a) Two adjacent frames, (b) some critical areas of the frames.

Taking account of the above issues, a scale-compensated measurement is used to estimate an accurate scale between adjacent frames. After that, patches are extracted from different scales and are modified into one scale so that patch-matching is more precise. Finally, the HR frame is reconstructed by the proposed algorithm, Scale-Compensated NLM.

The rest of this paper is organized as follows. In Sec. II, an improved NLM SR, ARI-SWR algorithm, is reviewed. Scale-Detector based on scale-invariant feature transform (SIFT) and verification is presented in Sec. III. Sec. IV focuses on proposed algorithm. Experimental results are shown in Sec. V. A brief conclusion is given in Sec. VI.

* Corresponding Author
This work was supported by National Natural Science Foundation of China under contract No.61101078, National Key Technology R&D Program of China under Grant 2012BAH18B03 and Doctoral Fund of Ministry of Education of China under contract No.20110001120117.

II. Overview of the Improved NLM SR: ARI-SWR Algorithm

The NLM SR [2] works effectively based on the assumption that image contents is likely to repeat itself within the neighborhood. Although NLM SR is a useful way to reconstruct higher resolution frames, it overlooks some affine transformations between the frames, including translation, rotation and zooming. ARI-SWR algorithm [6] is presented to compensate for part of the above omissions. Applying structure descriptor and local intensity kernel to NLM SR and relocating the search window by motion estimating, ARI-SWR algorithm performs better in videos with rotation and translation. In this specific scenario, we use a reference frame and several candidate frames which are adjacent to the reference frame to generate a HR frame.

The value of each pixel $Res(k,l)$ is calculated by

$$Res(k,l) = \frac{\Sigma_{t\in[1,\cdots,T]}\Sigma_{(i,j)\in N(k,l)} w(k,l,i,j,t) y_t(i,j)}{\Sigma_{t\in[1,\cdots,T]}\Sigma_{(i,j)\in N(k,l)} w(k,l,i,j,t)}, \tag{1}$$

where T is the number of reference frames, the neighborhood of the pixel (k,l) is represented by $N(k,l)$, and y_t stands for the t-th reference frame. $w_t(k,l,i,j)$ is given by

$$w(k,l,i,j,t) = \frac{1}{C(k,l)} \cdot exp\left\{-\frac{\|S(k,l)Y_r - S(i,j)Y_t\|_2^2}{2\sigma_1^2}\right\} \cdot exp\left\{-\frac{\|I(k,l)Y_r - I(i,j)Y_t\|_2^2}{2\sigma_2^2}\right\}, \tag{2}$$

where S and I denote the structure descriptor and the local intensity kernel, $C(k,l)$ is the normalization constant, Y_r and Y_t stand for the HR reference frame and the t-th HR candidate frame generated by bilinear interpolation. $w(k,l,i,j,t)$ is the weight that describes the similarity between the reference patch and the candidate patch. A higher weight implies more similarity between them. σ_1 and σ_2 control the proportion of S and I. σ_1 is defined as a piecewise function to be adjusted adaptively.

Although ARI-SWR algorithm considers rotation and translation in videos, it neglects zooming. Thus, we proposed the scale-detector which calculates the scale gap between the frames to consider zooming in our reconstruction approach.

III. Scale-Detector and the Verification

A. Scale-Detector Based on SIFT

SIFT [8] proves a reliable and effective method for extracting the distinctive features (keypoints) from frames regardless of affine transformation. There are two main stages of generating the set of keypoints, keypoints detection and keypoints description. In addition, when SIFT is served to compute the relationship of two frames, keypoints matching are proposed. Our algorithm uses SIFT to compute the keypoints in each frame and matches these keypoints between adjacent frames. Then the scale between adjacent frames is defined as follow:

$$s_t = \frac{1}{|M|} \times \left(\Sigma_{i\in M} \frac{s'_r(i)}{s'_c(i)}\right), \tag{3}$$

where s_t means the scale between the t-th candidate frame and the reference frame. $s'_r(i)$ and $s'_c(i)$ stand for the scale of the i-th matched keypoint descriptor in the reference frame and the i-th candidate frame, respectively. M is the set of matched keypoints and $|M|$ is the number of elements in M.

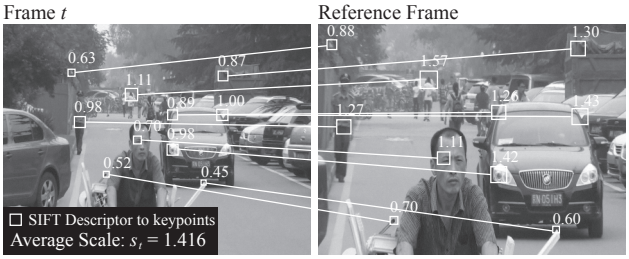

Fig. 3. Partial matched keypoints and the corresponding scale values.

Fig. 3 is a brief illustration of the scale-detector. Ratios of all the matched keypoints are summed and averaged so that influences from several keypoints computation errors in SIFT are weakened and a stable and accurate estimation of the scale between adjacent frames can be obtained.

B. The Verification of Scale-Detector

To demonstrate the precision and stability of the scale-detector, some verification experiments are conducted. We apply scale-detector to various changing scales and resolutions of the sequences. We use three CIF sequences (*Bus*, *City* and *Crew*) with a fixed resolution (352×288) on changing scales in Figs. 2(a) and (b). And we use three 1080p video sequences

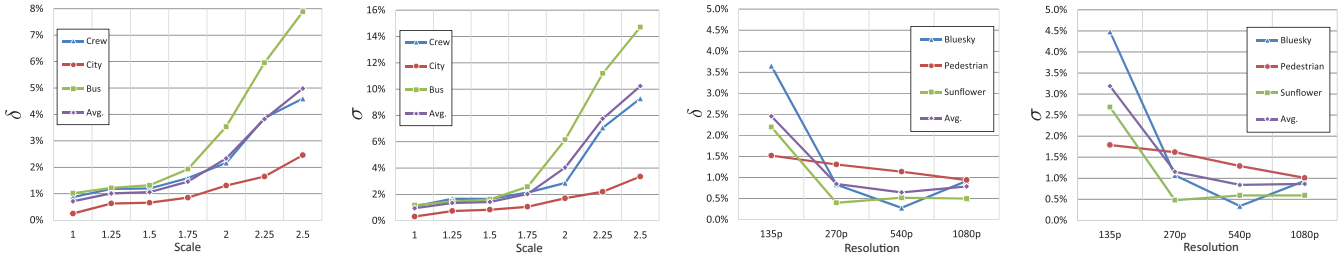

Fig. 2. The performances of scale-detector in different standard scales and different resolutions. 1080p, 540p, 270p and 135p represent resolution of 1920×1080, 960×540, 480×270 and 240×135 respectively.

(*Bluesky*, *Pedestrian* and *Sunflower*) with a fixed scale gap between testing frames ($2\times$) on changing resolutions in Figs. 2(c) and (d). In Figs. 2(a) and (c), the average relative error (axis-y, δ) reflects the precision of the scale-detector. And in Figs. 2(b) and (d), the standard deviation (axis-y, σ) reflects the stability of the scale-detector.

When the scale is under 2.5 and the resolution is higher than 270p, Figs. 2(a) and (c) show the low level of δ which indicates the high level of the precision of scale-detector. In Fig. 2(b), when the scale is under 1.5, over 99% of the results fluctuate in a very small range ($\pm 5\%$). Even when the scale reaches 2.5, over 60% of the fluctuation of results are still under 10%. In Fig. 2(d), over 99% of the results have very little fluctuation ($\pm 5\%$) when the resolution is higher than 270p. Therefore, when the scale is under 2.5 and the resolution is higher than 270p, scale-detector generates stable consequences.

Limited by the zooming speed of the camera, the scale changing of adjacent frames is rarely over 2.5. And the resolution of a video sequence is unlikely to be lower than 270p with the development of camera devices. Thus, scale-detector can be consider efficient and reliable on most video sequences.

IV. Scale-Compensated Nonlocal Means

After calculating the accurate scales between adjacent frames by scale-detector, we extract and modify patches from the candidate frames into the scale of the reference frames. Then, we reconstruct the HR frames based on NLM SR. Sec. IV-A focuses on the patch extraction, modification and matching. The whole algorithm is presented in Sec. IV-B.

A. Patches from Multi-Scales

In both NLM SR and ARI-SWR algorithm, the size of a patch has been decided before SR reconstruction. Therefore, any similar patch with zooming in different frames has lower weights, and is even more easily to be neglected when patch matching. After obtaining accurate scales from the scale-detector, we match the patches from different scales more precisely and obtain extra information.

First, patches in different scales are extracted on the basis of the scale between the reference frame and the t-th candidate frame (s_t in Sec. III). Then patches are modified by interpolation into the scale of the reference frame. To sum up, each modified patch can be described as follow:

$$MP(i,j,t) = I(s_t{\times}f)R(s_t,i,j)y_t, \tag{4}$$

where $MP(i,j,t)$ represents the modified patch, $I(s_t{\times}f)$ and $R(s_t,i,j)$ are interpolation operator and patch extraction operator in the scale of $s_t{\times}f$, y_t is the t-th candidate frame. In addition, we use bilinear interpolation as $I(s_t{\times}f)$, $s_t{\times}f$ is the interpolation scale factor.

With the set of modified patches, we match the patches and get the final value of each pixel as follow:

$$Res(k,l) = \frac{\Sigma_{t\in[1,\cdots,T]}\Sigma_{(i,j)\in N(k,l)}\hat{w}(k,l,i,j,t)y_t(i,j)}{\Sigma_{t\in[1,\cdots,T]}\Sigma_{(i,j)\in N(k,l)}\hat{w}(k,l,i,j,t)}, \tag{5}$$

which is the same as NLM SR, and specially, $\hat{w}_t(k,l,i,j)$ can be described as follow:

$$\hat{w}(k,l,i,j,t) = exp\left\{-\frac{\|R(1,k,l)Y_r - I(s_t{\times}f)R(s_t,i,j)y_t\|_2^2}{2\hat{\sigma}^2}\right\}, \tag{6}$$

where Y_r is the HR reference frame interpolated by bilinear, y_t is the t-th LR candidate frame, $\hat{\sigma}$ is a constant. As we have got the set of modified patches, the equation can be simplified as follow:

$$\hat{w}(k,l,i,j,t) = exp\left\{-\frac{\|R(1,k,l)Y_r - MP(i,j,t)\|_2^2}{2\hat{\sigma}^2}\right\}, \tag{7}$$

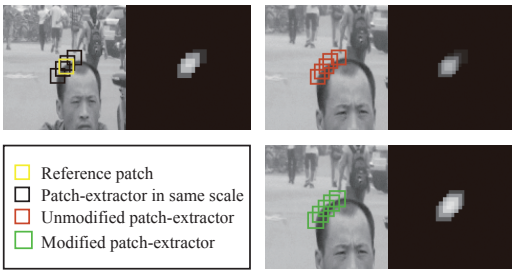

Fig. 4. Comparison of unmodified and modified patch-extractor in patch matching.

Algorithm 1 Scale-Compensated Nonlocal Means SR

Input:
- $y_t(t = 1,\cdots,T)$: input LR frames
- $Y_t(t = 1,\cdots,T)$: input HR frames generated from input LR

frames by interpolation
- f: the scale factor of SR
- $r(1 \le r \le T)$: the number of the reference frame in the sequence of LR frames

Initialization:

1. $V, W \leftarrow 0$

Scale detection: For each $t \in [1,T]$

2. SIFT processing
3. $s_t \leftarrow$ Eq. (3)

Patch Modification: For each $t \in [1,T]$, each $(i,j) \in y_t$

4. $p \leftarrow R(s_t,i,j)y_t$
5. $p \leftarrow I(s_t{\times}f)p$
6. $MP(i,j,t) \leftarrow p$

Reconstruction: For each $(k,l) \in Y_r$, each (i,j,t) that $(f{\times}i, f{\times}j, t) \in N(k,l)$

7. $\hat{w}(k,l,i,j,t) \leftarrow$ Eq. (7).
8. $V(k,l) += \Sigma_{t\in[1,\cdots,T]}\Sigma_{(i,j)\in N(k,l)}\hat{w}(k,l,i,j,t)y_t(i,j)$
9. $W(k,l) += \Sigma_{t\in[1,\cdots,T]}\Sigma_{(i,j)\in N(k,l)}\hat{w}(k,l,i,j,t)$

Result:For each $(k,l) \in Y_r$

10. $Res(k,l) \leftarrow V(k,l)/W(k,l)$

Output:
- Res: the HR reference frame

Since accurate scales are acquired and patches are modified from different scales into the same, we match the patches

more precisely and exploit more redundant information. In Fig. 4, when patch-extractor is modified, weights between truly similar patches are higher (lighter blocks in the figure), which means that the proposed patch-matching provides more useful information.

B. Scale-Compensated Nonlocal Means Algorithm

In this subsection, we integrate all the processes in the above and list a pseudocode of Scale-Compensated NLM in Algorithm 1 to give an overview of the proposed algorithm.

V. Experimental results

In experiments, we set patch size at 13×13, search window size at 37×37 and the scale factor of SR is 1:3 in each axis ($f = 3$). Our experiments have two main parts: testing on synthetic and real sequences. The improvements of the proposed algorithm can be easily observed. In measurements, Scale-Compensated NLM algorithm achieves the highest PSNR and SSIM among all the comparison experiments.

First, we test our algorithm on synthetic video sequences, such as *Foreman*. We down-sample several frames in *Foreman* to make a sequence with artificial zooming and reconstruct the 15-th of *Foreman* frames. To amplify the zooming effect in sequences, we use *Tempete* which is extracted with an interval of 20 frames, the 20^{th}, 40^{th}, 60^{th}, 80^{th} and 100^{th} frames, to reconstruct the 60^{th} frame.

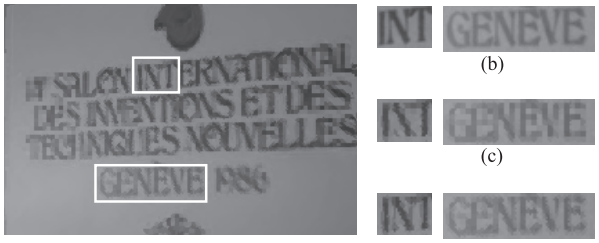

Fig. 5. Subjective comparison of two algorithms on *Text*. (a) An result of Scale-Compensated NLM on *Text*, (b) partial areas of the original frame, (c) NLM SR, (d) Scale-Compensated NLM.

TABLE I
Objective measurement of SR results in sequences (PSNR)

Sequence	NLM SR	ARI-SWR	Proposed
Foreman	31.15	30.96	**31.27**
Tempete	22.85	22.74	**23.00**
Text	29.23	30.06	**30.11**
Man	27.14	27.02	**27.29**

Considering the lack of natural zooming motion in synthetic video sequences, we shoot some real sequences with zooming. Our sequences (*Text* and *Man*) have lots of complex object motions during the camera motion and we choose some frames to conduct the experiments. Scale-Compensated NLM algorithm improves the original NLM SR by 0.1dB in objective measurement (PSNR) at Table I. Table II shows that Scale-Compensated NLM algorithm performs well in subjective measurement (SSIM). Scale-Compensated NLM algorithm preserves more details and produces less block effect in Figs. 5 and 6, compared to NLM SR. The above results indicate that Scale-Compensated NLM obtains more useful information from a same frame for the scale variation is considered. Thus, a positive conclusion to the proposed algorithm is made.

Fig. 6. Subjective comparison of two algorithms on *Man*. (a) An result of Scale-Compensated NLM on *Man*, (b) partial areas of original frame, (c) NLM SR, (d) Scale-Compensated NLM.

TABLE II
Subjective perception of SR results in sequences (SSIM)

Sequence	NLM SR	ARI-SWR	Proposed
Foreman	0.8109	0.8001	**0.8151**
Tempete	0.6927	0.6737	**0.7013**
Text	0.8592	0.8512	**0.8633**
Man	0.7780	0.7617	**0.7831**

VI. Conclusions

NLM SR is popular in SR reconstruction. However, most of these methods fail to concern the scale changes between frames. In this paper, we present two main contributions to solve this problem. One is the scale-detector which proves effective and reliable in detecting scale of video sequences. The other is the proposed Scale-Compensated NLM SR which improves NLM SR. To be extended, a more accurate scale-detector should be considered and a combination of rotation-invariant and translation-invariant algorithm is worth attempting.

References

[1] M. Tanaka, Y. Yaguchi and M. Okutomi. Robust and Accurate Estimation of Multiple Motions for Whole-Image Super-Resolution, *International Conference on Image Processing*, pp. 649-652, Oct. 2008.

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

[3] W. L. Zeng, X. B. Lu. Region-based Nonlocal Means Algorithm for Noise Removal, *Electronics Letters*, vol. 47, Issue. 20, pp. 1125-1127, 2011.

[4] H. Zheng, S. L. Phung. Wavelet based Nonlocal-Means Super-Resolution for Video Sequences, *IEEE International Conference on Image Processing*, pp. 2817-2820, Sept. 2010.

[5] M. H. Cheng, H. Y. Chen, J. J. Leou. Video Super-Resolution Reconstruction Using a Mobile Search Strategy and Adaptive Patch Size, *Signal Processing*, vol. 91, pp. 1284-1297, 2011.

[6] Y. Zhuo, J. Liu, J. Ren and Z. Guo. Nonlocal Based Super Resolution with Rotation Invariance and Search Window Relocation, *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Mar. 2012.

[7] D. Glasner, S. Bagon and M. Irani. Super Resolution from a Single Image, *International Conference on Computer Vision*, pp. 349-356, 2009.

[8] D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints, *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.