BILATERAL VIDEO SUPER-RESOLUTION USING NON-LOCAL MEANS WITH ADAPTIVE PARAMETERS

Yawei Li, Xiaofeng Li, Zhizhong Fu, Xiuxia Yin, Yufei Zhao

School of Communication and Information Engineering,
University of Electronic Science and Technology of China, Chengdu, Sichuan 611731 China

ABSTRACT

Super-resolution (SR) algorithms for video sequences with high resolution (HR) guide frames can provide outstanding performances. Non-local means (NLM) algorithm compares the similarity between a pixel and its neighbors. NLM replaces every pixel with a weighted average of its neighbors. The NLM based SR algorithm can super-resolve low resolution (LR) frames using the HR guide frames in the video sequence. However, the fixed decaying factor of NLM cannot satisfy regions of distinct characteristics in a LR frame. The fixed searching window fails to balance the requirements of low computational complexity and good SR images. Thus, we propose novel criteria for selecting the decaying factor and searching window adaptively. Bilateral adjacent HR frames are used to handle the occlusion problem. The experimental results verify the validity of the proposed method.

***Index Terms*—** Super-resolution, non-local means, adaptive parameters, video scaling.

1. INTRODUCTION

The bandwidth of the communication channel is the key factor that constrains the efficient transmission of videos. Various video compression methods have been explored to transmit video sequences with limited bandwidth [1]. To further compress videos, one can down-sample some frames of a video and obtain a low resolution (LR) video sequence with periodic high resolution (HR) frames (Fig. 1). On the other hand, the resultant multiple resolution video can also be captured by hybrid cameras which address the spatial-temporal resolution tradeoff of traditional cameras [2, 3]. Super-resolution (SR) could be applied to recover HR videos.

In order to recover the missing high frequency (HF) details of LR images and videos, SR algorithms use the spatial-temporal redundancies between related images and video frames [4–8]. There exist two categories of SR algorithms, that is, reconstruction-based SR and example-based SR. Reconstruction-based SR uses the sub-pixel motion among a series of LR images to recover a HR image [9]. However, when the magnification factor is large, reconstruction-based SR usually fails to provide satisfying detail information.

Example-based algorithms use known HR images to build a database which consists of pairs of low frequency (LF) information and HF information in a training phase [10]. Then the established database guides the learning phase to search a matching HR block for every block in the LR image. Inspired by Freeman *et al.* [10], Brandi *et al.* refined the LR frames using the periodic HR frames as references [11]. They added the HF information from the matching HR block to a target LR block, thus finishing the recovery of a LR frame.

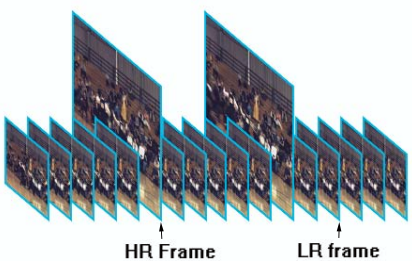


Fig. 1. Illustration of a video sequence with periodic HR frames. The LR frames are obtained by down-sampling the original full resolution frames.

Recently, SR based on non-local means (NLM) has been introduced. NLM is originally an image denoising algorithm assuming that the pattern of a patch may repeat within an image [12]. Protter *et al.* first generalized NLM to SR from the viewpoint of error energy minimization [13]. Basavaraja *et al.* combined the work in [11] and [13] to compute the HF part of a pixel using NLM [14]. Lengyel *et al.* incorporated illuminance and gradient information into the similarity comparison and reduced the averaging pixels by thresholding [15]. However, these NLM based SR use fixed parameters.

In this paper, we propose two adaptive parameters to improve the performance of NLM. The first parameter is an adaptive decaying factor that is used to accommodate different regions such as flat regions and texture regions. The second one is an adaptive searching window proposed to balance the computation complexity and the quality of the re-

This work was supported by the Natural Science Foundation of China (61075013).

fined frames. Moreover, bilateral adjacent HR frames are used to ease the occlusion problem between different video frames.

The rest of the paper is organized as follows. Section 2 explains the basic models of the algorithm. Section 3 describes the selection of the adaptive parameters. Section 4 contains the experimental results. Section 5 concludes the paper.

2. BASIC MODELS OF THE ALGORITHM

2.1. Preprocessing of the video sequence

The processed video sequence comprises of periodic HR frames and LR frames which are denoted by

$$\{F_k | k = Tz, z \in \mathbb{N}\} \tag{1a}$$

$$\{f_n | n = Tz + r, z \in \mathbb{N}, r = 1, 2, ..., R-1\} \tag{1b}$$

where F_k is a HR frame, f_n a LR frame, and T the period of HR frames. To generate the LF part of the HR frames, the HR frames are blurred, down-sampled, and interpolated, namely,

$$F_k^L = UDBF_k \tag{2}$$

where B, D, and U are the blurring operation, down-sampling operation, and interpolating operation, F_k^L is the LF part of F_k. Then F_k^L is subtracted from F_k, resulting in the HF detail of F_k, namely,

$$F_k^H = F_k - F_k^L. \tag{3}$$

The LR frames are also scaled to the same resolution as the HR frames, namely,

$$\tilde{f}_n^L = Uf_n \tag{4}$$

F_k^L and $\tilde{f}_n^L$ contain the basic structure information and are used to compute the weights in the NLM algorithm. F_k^H is used to recover the missing details of $\tilde{f}_n^L$.

2.2. NLM algorithm

NLM exploits the spatial redundancy of an image to recover a pixel. Each pixel is replaced by the weighted average of pixels in its neighborhood, namely,

$$\tilde{f}_n^H(x,y) = \frac{\sum\limits_{(i,j)\in\Omega_{xy}} \omega_{x,y}(i,j) \cdot F_k^H(i,j)}{\sum\limits_{(i,j)\in\Omega_{xy}} \omega_{x,y}(i,j)} \tag{5}$$

where Ω_{xy} is the neighborhood (searching window of size $\mathcal{N} \times \mathcal{N}$) of the pixel (x,y), (i,j) is a pixel in Ω_{xy}. The weight assigned to pixel (i,j) reflects the similarity between (i,j) and (x,y) and is computed as

$$\omega_{x,y}(i,j) = \exp\left(\frac{-\left\|(R_{x,y}^S \tilde{f}_n^L - R_{i,j}^S F_k^L) G_{\sigma_s}\right\|_2^2}{2\sigma^2}\right) \tag{6}$$

where σ is a decaying factor, $R_{i,j}^S$ is an operator that extracts a patch of size $S \times S$ centered at (i,j), and G_{σ_s} is a Gaussian kernel with 0 mean and variance σ_s^2. The patch difference is

$$E_{i,j}^2 = \left\|(R_{x,y}^S \tilde{f}_n^L - R_{i,j}^S F_k^L) G_{\sigma_s}\right\|_2^2. \tag{7}$$

At last, the HF detail $\tilde{f}_n^H(x,y)$ and LF structure $\tilde{f}_n^L(x,y)$ are added to form the recovered pixel value $\tilde{f}_n(x,y)$, namely,

$$\tilde{f}_n(x,y) = \tilde{f}_n^H(x,y) + \tilde{f}_n^L(x,y). \tag{8}$$

3. THE PROPOSED ALGORITHM

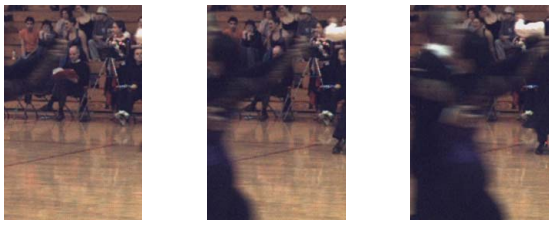
3.1. Bilateral video super-resolution

NLM searches in the neighborhood of a center pixel to find similar pixels to the center pixel. NLM can be considered as a coarse and implicit motion estimator. There is object or scene motion between consecutive frames. Thus, an object around the boundary may move in or out the current scene (see Fig. 2). If only one HR frame is used to refine a LR frame, NLM may fail to find similar pixels. Fortunately, we can use bilateral adjacent frames to ease this problem. For every LR frame, its bilateral adjacent HR frames are

$$\{F_b | b \in \Phi\} \tag{9}$$

where $\Phi = \{\lfloor n/T \rfloor \times T, \lceil n/T \rceil \times T\}$, $\lfloor \cdot \rfloor$ is the round-down operator, and $\lceil \cdot \rceil$ the round-up operator.

Bilateral SR is somewhat similar to motion-compensated frame rate up-conversion (MC-FRC) [16–18] in that both techniques use information in the adjacent frames to recover a middle frame. However, they differ mainly in two aspects. Firstly, bilateral SR refines a LR frame with HF information while MC-FRC aims at generating a non-existent frame. Secondly, compared with MC-FRC, bilateral SR doesn't rely on explicit motion estimation.



(a) (b) (c)

Fig. 2. Parts of (a) Frame 6, (b) Frame 9, and (c) Frame 12 of *Ballroom_0*. The pixels of the moving dancers in Frame 9 have no correspondences in Frame 6. Thus, it's impossible to refine these pixels using only Frame 6. This problem can be solved by including both Frame 6 and Frame 12 into the SR algorithm.

3.2. Adaptive searching window

The size of the searching window $\mathcal{N}$ of NLM is a fixed parameter. A frame of a video sequence can be divided into background and foreground objects. The background is usually stable or moves smoothly. So a small searching window is enough for NLM to find similar pixels to a center pixel. However, the foreground objects may move fast and out of the range of the searching window. In this case, one has to enlarge the size of the searching window in order to recover the details of the moving pixels. This treatment certainly increases the computational complexity. Thus, a fixed window size cannot meet both of the demands.

Motion estimation can be used to determine the size of the searching window of each pixel. However, motion estimation involves complex computation. Thus, we propose to use varying searching window whose computation is relative simple at the cost of accuracy. First of all, the absolute difference between two LF images F_b^L and $\tilde{f}_n^L$ is calculated, i.e.,

$$\Delta_b = |F_b^L - \tilde{f}_n^L|. \tag{10}$$

Then a map is established by comparing Δ_b with its mean

$$\mathcal{M}_b(x,y) = \begin{cases} 1, \Delta_b(x,y) > m_b \\ 0, otherwise \end{cases} \tag{11a}$$

$$m_b = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\Delta_b(p,q) \tag{11b}$$

where M and N are the height and width of the images. One can distinguish the background and the foreground using (10) and (11). In order to discriminate pixels with different motion, an indicator is calculated for every pixel, namely,

$$\mathcal{I}_b(x,y) = \sum_{(p,q)\in \mathrm{O}_{xy}} \mathcal{M}_b(p,q) \tag{12}$$

where O_{xy} is a neighborhood of (x,y) of size $K \times K$, (p,q) is a pixel in O_{xy}. Then, a window size is assigned to every pixel with respect to the indicator, namely,

$$\mathcal{S}_b(x,y) = s_l, n_l \le \mathcal{I}_b(x,y) < n_{l+1} \tag{13}$$

where s_l's are the adaptive window sizes and n_l's are the thresholds between two consecutive levels with $n_0 = 0$ and $n_L = \max\{\mathcal{I}(x,y)\}$. Finally, the maximum window size is calculated for every pixel

$$\mathcal{S}(x,y) = \max_{b\in\Phi}\{\mathcal{S}_b(x,y)\}. \tag{14}$$

The purpose of the above computation is the same as that of motion estimation, i.e., determining the size of the searching window for every pixel adaptively. But there are differences between them. The above computation avoids expensive motion estimation while its accuracy is worse than motion estimation. Thus, it's a tradeoff between computational complexity and accuracy.

3.3. Adaptive decaying factor

The decaying factor σ of NLM is a fixed parameter that needs to be carefully selected. If the decaying factor is large, then the weights of pixels will be close to each other. As a result, the output image tends to be over-smoothed. On the other hand, if the decaying factor is small, then the weights will be close to 0. In the worst case, all the weights will decays to 0. Thus, we propose an adaptive decaying factor.

The criterion here is to force the minimum value of the patch difference $E_{i,j}^2$, after smoothed by $2\sigma^2$, to decay to a predefined value α, i.e.,

$$\min_{(i,j)\in\Omega_{xy}^{\mathcal{S}}}\left\{E_{i,j}^2\right\} \Big/ 2\sigma_A^2 = \lambda/2\sigma_A^2 = \alpha \tag{15}$$

where $\lambda = \min_{(i,j)\in\Omega_{xy}^{\mathcal{S}}}\left\{E_{i,j}^2\right\}$ and $\Omega_{xy}^{\mathcal{S}}$ is the adaptive searching window of size $\mathcal{S}(x,y)$. Solving the above equation results in

$$\sigma_A = \sqrt{\lambda/2\alpha}. \tag{16}$$

Note that the minimum patch difference may be zero. This means that an exactly identical pixel is found in the neighborhood of a pixel. In this case, one can directly set the weight of the identical pixel to 1 and all the other weights to 0.

4. EXPERIMENTS AND RESULTS

This section shows the experimental results. We compare six SR methods including bilinear interpolation (BI), classical NLM (NLM) [12], detail warping based NLM (DW) [14]), NLM with adaptive decaying factor (ADF), NLM with adaptive searching window (ASW), and the proposed method (Pro.). These methods are tested on five common test sequences with different characteristics including *Ballroom_0*, *Foreman*, *Mobile*, *News*, and *Flower*. The parameter setup is listed in Table 1. The thresholds n_l's in (13) are $0, 10, 20, \cdots, 100$, respectively. The corresponding window sizes s_l's of each level are $5, 10, 15, \cdots, 50$, respectively.

The video sequences with periodic HR frames were generated as follows. The LR frames were acquired from the original full resolution frames after blurring and down-sampling. The blurring and down-sampling operators were the same as those in (2). The down-sampling factor was 2. The full resolution frames corresponding to the HR frames remained unchanged. Lanczos filter acted as a point spread function (PSF) to simulate the blurring operation B during image acquisition process [19]. Peak Signal-to-Noise Ratio

Table 1. Parameter setup.

Parameter	T	$\mathcal{N}$	S	K	L	α	σ	σ_s
Value	6	9	5	10	10	2	0.2	1

Table 2. PSNR (dB) results of 5 test sequences.

Sequence	BI	NLM	DW	ADF	ASW	Pro.
Ballroom_0	30.71	35.15	36.52	37.04	37.85	**38.93**
Foreman	31.76	35.71	36.74	37.62	37.83	**38.68**
Mobile	20.05	22.34	22.52	24.10	22.76	**24.42**
News	24.74	36.02	36.77	38.10	37.11	**38.68**
Flower	21.23	24.62	24.93	26.35	27.16	**28.38**

Table 3. SSIM results of 5 test sequences.

Sequence	BI	NLM	DW	ADF	ASW	Pro.
Ballroom_0	0.958	0.980	0.988	0.990	0.990	**0.991**
Foreman	0.938	0.955	0.962	0.967	0.967	**0.970**
Mobile	0.757	0.892	0.895	0.915	0.898	**0.917**
News	0.905	0.982	0.985	0.989	0.989	**0.990**
Flower	0.823	0.944	0.947	0.956	0.964	**0.969**

(PSNR) and Structural SIMilarity (SSIM) [20] were used to compare and evaluate different methods.

Due to the limitation on the paper length, we only list the PSNR and SSIM results for the ninth frame of the five test sequences in Table 2 and Table 3, respectively. All of the methods except bilinear interpolation use the sixth and twelfth frames as the HR references. In fact, better SR results are derived for LR frames that are not exactly in the middle between the HR references because they have a closer HR reference. It's obvious that each of the adaptive parameters can improve the SR results. The PSNR and SSIM results increase further if the two adaptive parameters are combined. The PSNR gain of the proposed method with respect to detail warping based NLM varies from sequence to sequence. The largest PSNR gain is achieved for *Flower*, reaching 3.45dB.

The details of the recovered images of *Foreman* and *Ballroom_0* are shown in Fig. 3 and Fig. 4, respectively. The differences between these figures can be better perceived if they are displayed on a computer screen. The images obtained by bilinear interpolation are blurred versions of the original images. The images obtained by traditional NLM and detail warping based NLM contains distortions around the edges. This phenomenon is especially obvious for the tongue of *Foreman* where the edges are blurred by their surrounding pixels. The left eye of *Foreman* in Fig. 3(c), Fig. 3(d), and Fig. 3(e) is also distorted. However, the proposed method provides relatively good results. For the *Ballroom_0* sequence, the edges of the left arm of the dancer contain some inaccurate pixels in Fig. 4(c), Fig. 4(d), and Fig. 4(e). Some other edges (i.e., the edges of the nameplate) also contain inaccurate pixels. These artifacts are not unexpected because the object moves fast. Thus, NLM without adaptive searching window cannot find good matching pixels for a center pixel. These artifacts are overcome in Fig. 4(f).

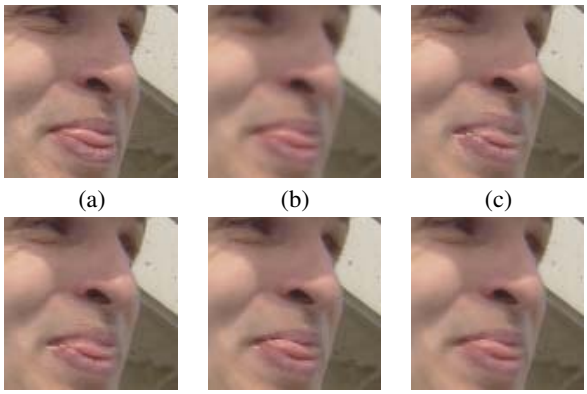


Fig. 3. Comparison of *Foreman* with different SR methods including (a) original HR image, (b) bilinear interpolation, (c) NLM, (d) detail warping based NLM, (e) NLM with adaptive searching window, and (f) the proposed method.

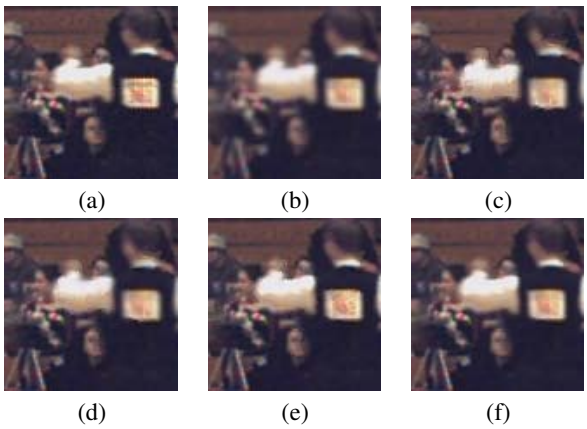


Fig. 4. Comparison of *Ballroom_0* with different SR methods including (a) original HR image, (b) bilinear interpolation, (c) NLM, (d) detail warping based NLM, (e) NLM with adaptive searching window, and (f) the proposed method.

5. CONCLUSION

In this paper, we propose novel methods to determine the decaying factor and searching window for every pixel. The adaptive decaying factor guarantees that the neighbors with high similarity to a pixel are assigned with large weights in the averaging process. The adaptive searching window can find right matching neighbors for pixels undergoing large motion. Bilateral adjacent HR frames are used in this method. The experimental results show the effectiveness of the adaptive searching window and decaying factor.

6. REFERENCES

[1] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Prentice Hall International, 2008.

[2] M. Ben-Ezra and S. K. Nayar, "Motion deblurring using hybrid imaging," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2003, pp. 657 – 664.

[3] Z. Hu, L. Xu, and M. Yang, "Joint depth estimation and camera shake removal from single blurry image," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2893 – 2900.

[4] R. Y. Tsai and T. S. Huang, "Multi-frame image restoration and registration," in *Proc. Advances in Computer Vision and Image Processing*, 1984, vol. 1, pp. 321 – 324.

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

[6] K. Zhang, X. Gao, D. Tao, and X. Li, "Single image super-resolution with multiscale similarity learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 10, pp. 1648 – 1659, 2013.

[7] T. Richter, J. Seiler, W. Schnurrer, and A. Kaup, "Robust super-resolution for mixed-resolution multiview image plus depth data," *IEEE Trans. Circuits Syst. Video Technol.*, 2015.

[8] Z. Jin, T. Tillo, C. Yao, and J. Xiao, "Virtual view assisted video super-resolution and enhancement," *IEEE Trans. Circuits Syst. Video Technol.*, vol. PP, no. 99, pp. 1, 2015.

[9] D. Capel and A. Zisserman, "Computer vision applied to super-resolution," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 75 – 86, 2003.

[10] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Comput. Graph. Appl.*, vol. 22, no. 2, pp. 56 – 65, 2001.

[11] F. Brandi, R. de Queiroz, and D. Mukherjee, "Super-resolution of video using key frames and motion estimation," in *Proc. IEEE International Conference on Image Processing*, 2008, pp. 321 – 324.

[12] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, 2005, vol. 2, pp. 60 – 65.

[13] M. Protter, M. Elad, T. Takeda, and P. Milanfar, "Generalizing the nonlocal-means to super-resolution reconstruction," *IEEE Trans. Image Process.*, vol. 18, no. 1, pp. 36 – 51, 2009.

[14] S. V. Basavaraja, A. S. Bopardikar, and S. Velusamy, "Detail warping based video super-resolution using image guides," in *Proc. IEEE International Conference on Image Processing*, 2010, pp. 2009 – 2012.

[15] R. Lengyel, S. M. Reza Soroushmehr, and S. Shirani, "Multi-view video super-resolution for hybrid cameras using modified NLM and adaptive thresholding," in *Proc. IEEE International Conference on Image Processing*, 2014, pp. 5437 – 5441.

[16] Sung Hee Lee, Ohjae Kwon, and Rae Hong Park, "Weighted-adaptive motion-compensated frame rate up-conversion," *IEEE Trans. Consum. Electron.*, vol. 49, no. 3, pp. 485 – 492, 2003.

[17] D. Wang, L. Zhang, and A. Vincent, "Motion-compensated frame rate up-conversion-łpart i: fast multi-frame motion estimation," *IEEE Trans. Broadcast.*, vol. 56, no. 2, pp. 133 – 141, 2010.

[18] D. Wang, A. Vincent, P. Blanchfield, and R. Klepko, "Motion-compensated frame rate up-conversion–part ii: new algorithms for frame interpolation," *IEEE Trans. Broadcast.*, vol. 56, no. 2, pp. 142 – 149, 2010.

[19] C. E. Duchon, "Lanczos filtering in one and two dimensions," *J. Appl. Meteorol.*, vol. 18, pp. 1016 – 1022, 1979.

[20] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600 – 612, 2004.