NONLOCAL BASED SUPER RESOLUTION WITH ROTATION INVARIANCE AND SEARCH WINDOW RELOCATION

Yue Zhuo[1,2], *Jiaying Liu*[1,*], *Jie Ren*[1], *Zongming Guo*[1]
[1]Institute of Computer Science and Technology, Peking University, Beijing, China 100871
[2]College of Information Science and Technology, Beijing Normal University, Beijing, China 100875

ABSTRACT

In this paper, we present a novel method for Super Resolution (SR) reconstruction with rotation invariance and search window relocation. To combine complementary information in observed images to generate a higher resolution image, we first relocate search window to involve potential similar patches and then use rotation invariance similarity measure to find accurate similar patches. Comparing with Nonlocal Means SR, our algorithm can find more similar patches for weighted average. Experimental results demonstrate superior performance of the proposed method in terms of both objective measurements and subjective evaluation.

***Index Terms*—** Super Resolution, Rotation invariance, Search window relocation, Nonlocal means

1. INTRODUCTION

Multi-frame SR reconstruction aims to fuse a set of observed LR images into one HR image. Due to subpixel shifts, each observed LR image contains complementary information. With knowledge of the shifts, these LR images can be combined to remove the aliasing and generate a higher resolution image.

In conventional multi-frame algorithms, it is essential to know the subpixel displacements between LR images. Thus, accurate motion estimation plays a critical role in conventional multi-frame SR. But unavoidable motion estimation errors lead to disturbing artifacts. To avoid motion estimation, Potter *et al*. [1] generalized Nonlocal means (NLM) from denoising algorithm to a motion estimation free SR method which averages neighbors by measuring patch similarity to reconstruct the center pixel. Taking account of the NLM SR or NLM denoise, two categories of improvements are proposed: *Adaptive parameters selection* and *invariance-based similarity measure*. Adaptive parameters selection is to discuss the relationship among patch size, search window size and performance of NLM, then adaptively select these parameters. Cheng *et al*. [2] proposed a SR reconstruction approach using a mobile search strategy and adaptive patch size. However, the mobile search strategy is pixel-wise and is likely to be trapped into local minimum. Zontak *et al*. [3] used Parzen window estimation to evaluate the average number of good nearest neighbors and eventually get a function of search window size and mean gradient magnitude. But this function only fits single image thus it cannot be directly used for multi-frame SR. Invariance-based similarity measure is to change the measure of patch similarity to be insensitive to mirror, rotation, *etc*. Grewenig *et al*. [4] proposed two rotation invariance (RI) approaches on denoising: moment-based approach and rotationally invariant block matching approach. Gao *et al*. [5] proposed a Zernike moment based method for SR reconstruction. However, invariance-based similarity measure assumes similar patches are always inside fixed search window which may not be satisfied in practice.

In this work, to perform SR reconstruction, we aim to find similar patches of reference patch and then combine their information to reconstruct the center pixel of reference patch. To achieve our goal, we start from finding potential similar patches and then assign them suitable weights for average. We conclude that the reason of falling to find enough similar patches is complex motion or limitation of search window. Based on the analysis, we propose a novel method for RI similarity measure which involves local gradient and intensity information and block-based search window relocation (SWR).

The rest of the paper is organized as follow: Section 2 reviews NLM SR and discusses its drawbacks. In Section 3, we propose adaptive rotation invariance similarity measure with search window relocation (ARI-SWR) algorithm. Experimental results are described in Section 4. Finally, concluding remarks are given in Section 5.

2. REVIEW ON NONLOCAL MEANS SUPER RESOLUTION

NLM takes the advantage of the redundancy of similar patches existing inside image for denoising. Then, the idea is generalized to Multi-frame SR in [1] that is NLM SR.

For each pixel (k,j) in the reference frame, its estimated value $X(k,l)$ is given by:

$$X\left(k,l\right)=\frac{\sum_{t\in[1,\cdots,T]}\sum_{(i,j)\in N(k,l)} w_t\left(k,l,i,j\right) y_t\left(i,j\right)}{\sum_{t\in[1,\cdots,T]}\sum_{(i,j)\in N(k,l)} w_t\left(k,l,i,j\right)}, \qquad (1)$$

* Corresponding author. This work was supported by National Natural Science Foundation of China under contract No.61101078, Research Fund for Doctoral Program of Higher Education of China under contract No.20110001120117 and Beijing Nova program under contract No. 2010B001.

where T is the number of candidate frames, y_t represents for t-th LR image, $N(k,l)$ is the neighborhood centered at (k,l) which is predetermined by a fixed search window and similarity weight $w(k,l,i,j)$ is defined as:

$$w_t(k,l,i,j) = \exp\left\{-\frac{\left\|R_{k,l}Y_{t0} - R_{i,j}Y_t\right\|_2^2}{2\sigma^2}\right\}, \tag{2}$$

where $R_{i,j}$ represents an operator extracts a patch of a fixed and predetermined size $(q \times q)$ from an image and get a vector of length q^2, σ acts as a smoothing parameter controlling the effect of the grey-level difference between these two image patches, Y_{t0} and Y_t are HR images of t_0-th and t-th respectively, obtained by interpolation.

Notice that deblurring is separated from the above reconstruction process which is performed at last. In this work, we do not focus on deblurring so that all of our experimental results are free of deblurring.

As we can see, NLM SR only takes translational motion into consideration. But natural videos usually contain complex motions, and even inside one frame, textured patches rotate in some places, which decrease the number of similar patches that NLM SR can find. Thus NLM SR does not fully exploit redundancy and complementary information in observed images. To solve this problem, we propose ARI-SWR algorithm to make use of more available information.

3. ARI-SWR SR ALGORITHM

3.1. Search window relocation

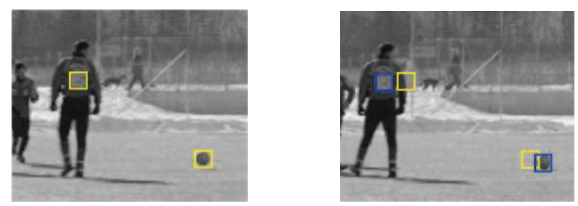
In this section, we focus on finding potential similar patches. Due to objects motion or camera motion, fixed search window fails to locate potential similar patches when objects are out of search window as the region highlighted with yellow rectangle shown in Fig.1. One solution is to enlarge the window size, but it increases computational complexity. In this work, we propose an alternative solution: SWR approach, which uses a fixed search window size but centered at different location in different frames. It not only reduces computational complexity but also acquires better results than relatively global search as [3] indicates that textured patches benefit more from local search.

The proposed SWR approach is block-based which aims to find potential similar patches without determining which patches are similar to reference patch. To avoid block-matching search trapped into local minimum, we use predicted MV assuming that motion is continuous spatially and temporally. For each macroblock (MB) (x,y) in reference frame t_0, the way to calculate its motion vector (MV) (v_x,v_y) in subsequent frame t $(t>t_0)$ is described in Table 1. We define $M(d_i,d_j,i,j,k)$ as a function which uses adaptive rood pattern search (ARPS) [6] to search for corresponding MB in $(k+1)$-th frame of MB (i,j) in k-th frame with predicted MV (d_i,d_j) and returns its MV (v_i,v_j).

As a result, all the pixels within one MB share the same MVs. Fig.1 shows the location of search window of our algorithm and NLM SR in 15-th frame and 17-th frame. In 17-th frame, relocated search windows are highlighted with blue rectangle. Comparing with fixed search window of NLM SR, SWR increases the probability to find similar patches.

Table 1. SWR algorithm

if x = 1
  $(d_i,d_j) \leftarrow (0,0)$
else
  $(d_i,d_j) \leftarrow (v_{x-1},v_y)$
end if
$i \leftarrow x$
$j \leftarrow y$
$(v_x,v_y) \leftarrow (0,0)$
for $k = t_0$ **to** $t-1$ **do**
  $(v_i,v_j) \leftarrow M(d_i,d_j,i,j,k)$
  $(d_i,d_j) \leftarrow (v_i,v_j)$
  $i \leftarrow i + d_i$
  $j \leftarrow j + d_j$
  $(v_x,v_y) \leftarrow (v_x,v_y) + (v_i,v_j)$
end for



(a) (b)

Fig.1. Location of search windows in two frames: (a) Two search windows in 15-th frame (reference frame); (b) Corresponding search windows in 17-th frame.

3.2. Rotation invariance similarity measure

3.2.1. Structure term and intensity term

After finding potential similar patches, in this section, we focus on assigning suitable weights to these candidate patches to let accurate similar patches stand out. It is reasonable that patches which rotate around their center should be considered similar to ones that are not rotated. However, in NLM SR they are considered extremely different. In this work, we extract local structure and intensity information to obtain a RI descriptor. Then, we measure similarities between patches based on their RI descriptors.

To describe local structure, we simplify Scale-invariant feature transform (SIFT) [7] to obtain local RI structure descriptor. The main steps show as follow:

1) Calculate the gradient magnitude and orientation of each pixel in each image. Fig.2 uses arrows to present gradient magnitude and orientation at each pixel.

2) To each pixel, assign one dominant direction based on the gradient orientation histogram within radius r.

3) To get a RI descriptor at each pixel, rotate the coordinates and the gradient orientations within radius r to its dominant direction, and then a vector of length 128 is built.

Local intensity descriptor involves the neighborhood intensity at each pixel. In particular, all the pixels within radius r that have the same Manhattan distance from the center pixel are grouped into one cluster. Then, we compute the mean of intensity of each cluster and get a vector of $(r+1)$ elements at each pixel. The t-th element of the vector at pixel (k,l) is calculated as follow:

$$I_t(k,l,r) = \underset{(i,j)\, s.t.\, |i-k|+|j-l|=t}{mean} Y(i,j), \tag{3}$$

where $Y(i,j)$ represents intensity at pixel (i,j). Fig.2 shows four clusters around center pixel within r = 3, where each grayscale stands for one cluster of the intensity descriptor.

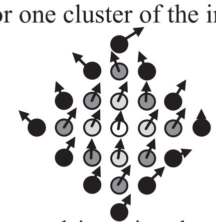


Fig.2. Local gradients and intensity clusters around center pixel within r = 3.

Similarity measured by Gaussian function of structure distance and intensity distance is defined as follow:

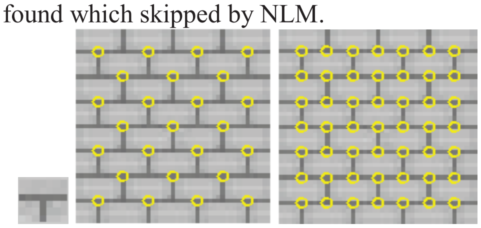
$$w(k,l,i,j) = \frac{1}{C(k,l)} \exp\left\{-\frac{\|P(i,j,r)-P(k,l,r)\|_2^2}{\sigma_1^2}\right\} \cdot \exp\left\{-\frac{\|I(i,j,r)-I(k,l,r)\|_2^2}{\sigma_2^2}\right\}, \tag{4}$$

where $P(i,j,r)$ represents vector built by local gradient, $I(i,j,r)$ represents vector built by local intensity and $C(k,l)$ is the normalization constant defined as

$$C(k,l) = \sum_{i,j\in N(k,l)} \exp\left\{-\frac{\|P(i,j,r)-P(k,l,r)\|_2^2}{\sigma_1^2}\right\} \cdot \exp\left\{-\frac{\|I(i,j,r)-I(k,l,r)\|_2^2}{\sigma_2^2}\right\}, \tag{5}$$

σ_1 and σ_2 control the effect of structure distance and intensity distance respectively.

Fig.3 shows the most similar patches of center patch our algorithm and NLM can find in one image. It presents that by using the proposed similarity measure, rotated patches can be found which skipped by NLM.



(a) (b) (c)

Fig.3. Similar patches found by NLM and RI: (a) Center patch. (b) Similar patches that NLM can find. (c) Similar patches that can be found by using the proposed similarity measure.

3.2.2. Adaptive parameter selection

Since we explicitly separate local structure and intensity, σ_1 and σ_2 should be selected carefully to balance structure term and intensity term. In this work, we fix σ_2 and think about an approach to choose σ_1 adaptively. We have observed that when the highest value of structure term is very small before normalization, the second highest value of structure term is often several orders of magnitude lower than the highest while the values of intensity item between candidate pixels are not so different that the intensity item does not work. In addition, due to the normalization, the weight of the patch with most similar structure to the reference patch almost equals to one while the weights of other candidate patches almost equal to zero. It leads to mismatches and degrades visual and objective quality. In that case, we say structure term is not so important thus σ_1 should be increased.

To select σ_1 adaptively, we first evaluate the importance of structure item by calculating the minimum distance of structure from reference patch. When structure item is not so important, which in this work means the minimum distance of structure from reference pixel is large, we set σ_1 with a relative large value in order to weaken the effect of structure item and enhance the effect of intensity item. The calculating formula of σ_1 is defined as a piecewise function:

$$\sigma_1 = \sigma_0 + step \cdot \left\lfloor \frac{\min\limits_{i,j\in N(k,l)} \left\{\|P(i,j,r)-P(k,l,r)\|_2^2\right\}}{L} \right\rfloor, \tag{6}$$

where σ_0 is the initial value of σ_1, L is the piecewise length and *step* controls the increase rate of σ_1 when the minimum distance of structure from reference pixel increases. By using adaptive σ_1 selection, most mismatches can be eliminated.

3.3. Adaptive rotation invariance similarity measure with search window relocation

Combining SWR, RI similarity measure and approach to balance structure item and intensity item, our complete ARI-SWR algorithm is summarized in Table 2:

Table 2. ARI-SWR algorithm

Input:
- $\{y_t\}_{t=1}^T$ - LR images.
- $\{Y_t\}_{t=1}^T$ - interpolated HR images.
- s - the scaling factor.
- r - the radius of descriptor.

Preprocessing: For each $t\in[1,T]$ and each $(i,j)\in t$-th HR
- Calculate descriptor including structure descriptor $P_t(i,j,r)$ and intensity descriptor $I_t(i,j,r)$ as Section 3.2 shows.

Fusion: For each $t_0\in[1,T]$, each $(k,l)\in t_0$-th HR image and for each candidate image $t\in[1,T]$
1) Compute MV (v_k,v_l) of the MB in t_0-th image which (k,l) belongs to in t-th image as Section 3.1 shows.
2) For each $(i,j)\in t$-th LR image and $(si,sj)\in N(k+v_k,l+v_l)$,
   - Compute its structure distance from (k,l).
   - Select the minimum distance to determine σ_1 as Eq.(6) shows.
3) For each $(i,j)\in t$-th LR image and $(si,sj)\in N(k+v_k,l+v_l)$
   - Compute w(i,j,k,l) using Eq.(4).
   - Add $y_t(i,j)$ to $X_{t0}(k,l)$ weighted by w(i,j,k,l).

Output:
- X is the result of estimated HR images.

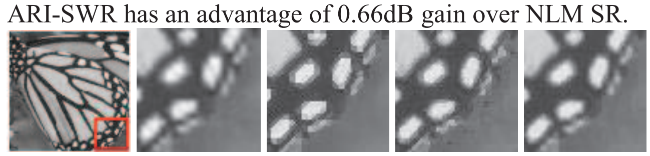
4. EXPERIMENTAL RESULTS

In the experiments, all the tests are blurred using a 3×3 uniform mask, decimated by a factor of 1:3 (in each axis), and then contaminated by an additive noise with standard deviation 2.

We first test on Monarch of 255×255 to perform single image SR without search window relocation to prove our RI similarity measure can find more similar patches so that eliminates block artifacts. Patch size is 7×7 for NLM SR, radius r is set to 3 for our algorithm and search window size is 21×21 for all the tests. As Fig. 4 shows, when textured patches rotate somewhere, NLM SR can not find their similar patches, thus produces block artifacts, while the results of our algorithm are free of block artifacts. Furthermore, adaptive parameter selection can eliminate mismatches (see the flat part in Fig.4).

Finally, we evaluate our algorithm on two real video sequences: Soccer and Ice. To accelerate computation, we only fuse 11 frames instead of whole frames to estimate one frame. Patch size is 13×13 for NLM SR, radius r is set to 6 for our algorithm and search window size is 27×27 for all the tests. Figs. 5 and 6 show that NLM SR generates disturbing block artifact (see legs part in Figs.5 and 6) and some details in flat part are regarded as noises so that they are denoised by NLM SR (see lawn part in Fig.5). Comparing with NLM SR, our algorithm not only eliminates block artifacts but also preserves details in flat part due to the structure term in similarity measure. Complete experimental results appear in our projects webpage [8].

The PSNR results for these two video sequences are summarized in Table 3, in which NLM-SWR SR represents NLM with search window relocation SR, and RI-SWR represents rotation invariance similarity measure with search window relocation. It shows that RI and SWR have 0.44dB and 0.25dB on average, respectively, higher than NLM SR, adaptive parameter selection further improve RI by 0.09dB and their combinations produce better results, among which ARI-SWR has an advantage of 0.66dB gain over NLM SR.

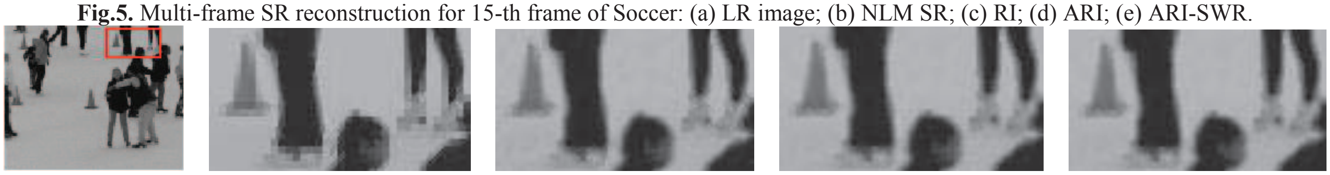


(a) (b) (c) (d) (e)

Fig.4. Partial results of single image SR and their global PSNR: (a) LR image; (b) Bicubic (BI), 22.64dB (c) NLM SR, 22.30dB; (d) RI, 22.58dB; (e) Adaptive rotation invariance (ARI), 23.00dB.

(a) (b) (c) (d) (e)

Fig.5. Multi-frame SR reconstruction for 15-th frame of Soccer: (a) LR image; (b) NLM SR; (c) RI; (d) ARI; (e) ARI-SWR.



(a) (b) (c) (d) (e)

Fig.6. Multi-frame SR reconstruction for 15-th frame of Ice: (a) LR image; (b) NLM SR; (c) RI; (d) ARI; (e) ARI-SWR.

Table 3. PSNR (dB) performance of two video sequences: Soccer and Ice.

Sequence	BI	NLM SR	RI	NLM-SWR SR	RI-SWR	ARI	ARI-SWR
Soccer	28.4499	27.8068	28.3302	28.2411	28.5366	28.3844	**28.6145**
Ice	28.3791	28.3717	28.7354	28.4355	28.7396	28.8680	**28.8852**

5. CONCLUSIONS

In this study, based on NLM SR framework, we focus on how to find more similar patches. By analyzing the reason of missing some similar patches in NLM SR, we propose a novel method for RI similarity measure and adaptive parameter selection. Another contribution of this work is we incorporate search window relocation. Experimental results are free of artifacts and of high quality.

6. REFERENCES

[1] M. Protter, M. Elad, H. Takeda, P. Milanfar, "Generalizing the Nonlocal-Means to Super-Resolution Reconstruction", *IEEE Trans. on Image Processing*, vol.8, no.1, pp. 36-51, 2009.

[2] 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.

[3] M. Zontak, M. Irani, "Internal Statistics of a Single Natural Image", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 977-984, 2011.

[4] S. Grewenig, S. Zimmer b, J. Weickert, "Rotationally Invariant Similarity Measures for Nonlocal Image Denoising", *Journal of Visual Com. and Image Rep.*, vol. 22, no. 2, pp. 117-130, 2011.

[5] X. Gao, Q. Wang, X. Li, D. Tao, K. Zhang, "Zernike Moment-based Image Super Resolution," *IEEE Transactions on Image Processing*, vol. 20, no. 10, pp. 2738-2747, 2011.

[6] Y. Nie, K.-K. Ma, "Adaptive Rood Pattern Search for Fast Block-Matching Motion Estimation", *IEEE Transactions on Image Processing*, vol. 11, no. 12, pp. 1442-1449, 2002.

[7] D. G. Lowe, "Distinctive image features from scale-invariant keypoints", *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.

[8] http://www.icst.pku.edu.cn/course/icb/ARI_SWR.html