Segmentation-Based Scale-Invariant Nonlocal Means Super Resolution

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

***Abstract*— Zooming in/out appears frequently in video shooting, which makes scale vary between frames. And object motion in videos may cause scale change of the object. It leads to the difficulty in finding similar patches and causes the invalidation of nonlocal means super resolution (NLM SR). In this paper, we propose a novel scale-compensated NLM SR algorithm. First, by considering the parameter model, the image is segmented in order to detect regions with different scales. Then, scale variations in different regions are computed based on SIFT descriptor. And patches extracted from different regions are compensated into the same scale to eliminate the effect of scale change. It is shown by experimental results that our proposed algorithm achieves the average PSNR by up to 0.678dB comparing with the state-of-the-art methods. Subjective results demonstrate the proposed method reduces artifacts and preserves more details.**

I. Introduction

Multi-frame super resolution (SR) methods reconstruct a high resolution (HR) frame from multiple low resolution (LR) frames. They are based on the assumption that LR frames can complement each other by a large amount of redundant information. Motion estimation techniques are employed in SR to obtain redundant information. However, due to the complexity of motion, unavoidable motion estimation error leads to annoying artifacts in super resolved HR frame. To avoid this problem, Potter *et al.* [1] proposed a motion-estimation-free algorithm based on NLM. NLM SR takes the advantage of the redundancy of patches existing in images. It obtains a better HR image with no explicit motion estimation by replacing every pixel with a weighted average of its neighborhood.

In the past decade, researchers made progress in NLM SR by improving patch matching and exploiting more information. Some suggest adaptively choosing parameters. The adaptive choice of parameters for NLM SR relates to the size of the patch and search window. The proper size of the patch and search window can help us find more accurate patches to improve the performance of NLM. Cheng *et al.* [2] proposed using mobilized search window and adaptive patch size. But the method of mobilizing search window is based on pixel and might fall to a local optimal solution easily. Some develop a new way of calculating similarities to use more available information. Patches which better represent the image characteristics are captured and thus the efficiency of matching those patches is improved. Grewening *et al.* [3] proposed two rotation-invariant denoising methods, which are based on moment and rotation-invariant patch searching. Meanwhile, patch matching relates to the measurement of similarity, which assumes that similar patches can always be found in a fixed search window. But the assumption may not work in practice.

Our previous work [4] [5] took rotation and illumination into consideration. However, in practical captured videos, global and local scale change frequently appear. While global scale change is caused by camera motion, object motion leads to local scale change. Taking Fig.1 as an example, camera motion and object motion bring different scale changes in different regions. As a result of scale change, it is difficult to find similar patches for NLM, which affects the performance of NLM SR. Moreover, all of these aforementioned improved NLM methods do not consider the problem of scale change.

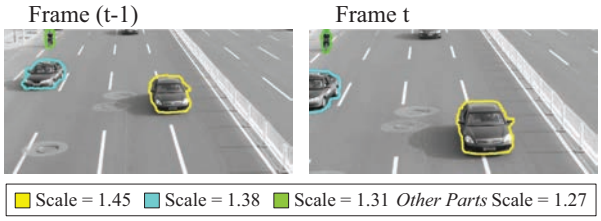

Fig. 1. Scale changing effects in adjacent frames.

In order to solve the above problem and consider the different scales in different regions, we propose a new method of patch matching to find similar patches for NLM. We segment the frames into several regions based on parameter model. The scale operator based on Scale-Invariant Feature Transform (SIFT) [6] is used to help us to obtain the scale differences and modify all matched patches into the same scale to compute the weights for NLM SR.

The rest of the paper is organized as follows. In Sec.II, improved NLM SR algorithms are reviewed. Sec.III focuses on the segmentation-based scale-invariant nonlocal means super resolution algorithm. The experimental results can be seen in Sec.IV and a brief conclusion is shown in Sec.V.

II. Overview of Improved Nonlocal Means SR

In the process of NLM SR [1] reconstruction, we usually build the cost function as a minimization problem.

$$X = \arg\min_{X}[\sum_{(k,l)\in\Omega}\sum_{t\in[1,\cdots,T]}\sum_{(i,j)\in N^L(k,l)} w(k,l,i,j,t) \|E^H_{k,l}D_{k,l}HX - E^L_{i,j}y_t\|_2^2],$$

$$w(k,l,i,j,t) = exp\left\{-\frac{\|P_{k,l}y_{t_0} - P_{i,j}y_t\|^2}{2\hat{\sigma}^2}\right\} f(k,l,i,j),$$

*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.

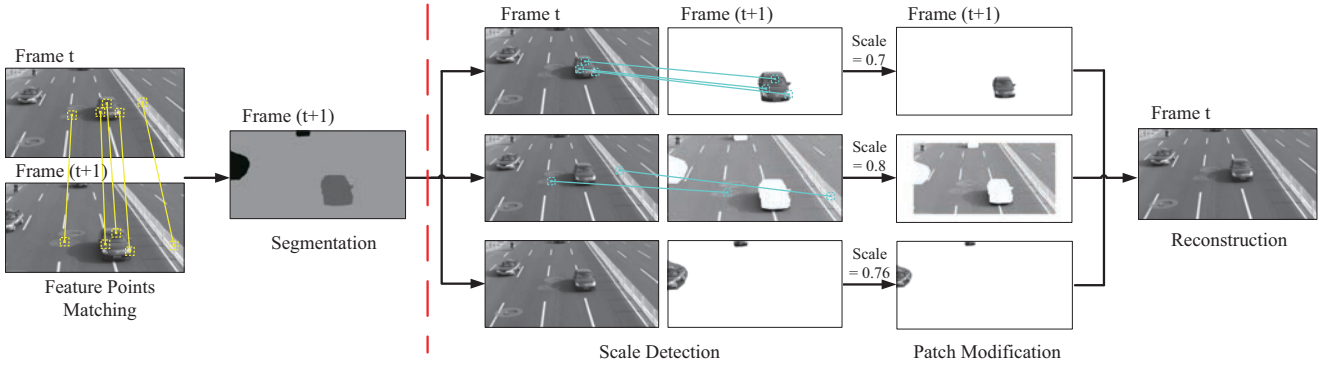

Fig. 2. The Framework of the Segmentation-Based Scale-Invariant Nonlocal Means Super Resolution Algorithm.

where X represents the desired output image to be created. T is the number of candidate frames and t is the frame index. $N^L(k,l)$ is the equivalent low-resolution neighborhood of pixel (k,l). The operator $E^H_{k,l}$ extracts a patch of size $p \times p$ pixels centered at (k,l) from HR frames while $E^L_{i,j}$ extracts a patch of size $q \times q$ pixels centered at (i,j) from LR frames, where $p = s(q-1)+1$ and s is the value of the scaling parameter. y_t defines the t-th LR candidate frame and y_{t_0} represents the reference frame. $w(k,l,i,j,t)$ defines the weight relating to pixel (k,l) in the reference frame and pixel (i,j) in the t-th candidate frame. $f(k,l,i,j)$ measures the spatial distance between (k,l) and (i,j). $P_{i,j}y_t$ and $P_{k,l}y_{t_0}$ represent a patch of size $q \times q$ with the center (i,j) extracted from the t-th candidate image y_t and with the center (k,l) extracted from the reference image y_{t_0} respectively. $\hat{\sigma}$ is a smoothing parameter which controls the effect of the gray-level difference between different patches. And H defines the blurring operator while $D_{k,l}$ refers to a patch decimation operator that ensures that the center pixel (k,l) of the patch is on the decimation grid.

However, NLM SR only considers translational motion. Complex situations such as rotation motions and illumination changes, which obstruct finding similar patches, exist in natural videos. To solve these two problems, adaptive rotation invariance similarity measure with search window relocation algorithm (ARI-SWR) [4] and illumination-invariant NLM based SR algorithm [5] were proposed in our previous work. The ARI-SWR algorithm relocates search window to involve potential similar patches and then uses rotation invariant similarity measurement to find accurate similar patches. The performance of NLM is improved by using more available information. Moreover, the illumination-invariant NLM SR algorithm improves NLM SR by eliminating the effects of illumination change. It adjusts the contrast between different search windows and selects proper candidate patches. The algorithm produces robust results to illumination changes by incorporating structure information and combining structure with intensity information.

Both of these methods did not consider scale as an important factor although the scale affects patch matching. Thus these existing NLM based methods do not fully exploit redundancy and complementary information in observed images. So we propose a scale-compensated NLM SR to address this issue.

III. Segmentation-Based Scale-Invariant Nonlocal Means

Taking camera motion and object motion into consideration, we propose a scale-invariant NLM SR algorithm based on segmentation. The framework of the proposed algorithm is illustrated in Fig.2.

In the segmentation stage as shown in the left side of Fig.2, we detect the feature point correspondences and compute the local parameter models for each pair using the weighted Lucas-Kanade algorithm [7]. Based on edge detection, support regions are extracted for each correspondence to generate the segmentation result. More details can be viewed in Sec.III-A. In Sec.III-B, the image is reconstructed by improved NLM SR. As shown in the right side of Fig.2, the similarity is measured between different patches with various scales. Scale differences between corresponding region pairs are calculated by weighted average algorithm. And the corresponding patches are adjusted into the same scale to acquire the similarities of each patch pair. The super resolved pixel is obtained by computing the weighted average of center pixels in every patch.

A. Parameter Model Based Segmentation

To obtain the segmentation results and detect the regions with different scales, we develop the algorithm in [8] and further improve it with clustering. The framework of this algorithm is shown in Fig.3. In the algorithm, segmentation is based on the parameter models of feature points, so that it is important to assure the accuracy of the original estimation of the parameters.

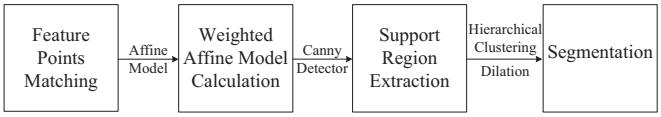

Fig. 3. The Framework of Parameter Model Based Segmentation

First, we use SIFT [6] to extract and match feature points from the LR candidate and reference frames. And the affine model T_i, which is used to describe local optical flow in i-th candidate frame, is initialized based on the feature information SIFT extracted. It has six unknown variables $p_j, j \in [1,6]$ and two of them relate to the scale in X and Y axis:

$$T_i = \begin{bmatrix} 1+p_1 & p_3 & p_5 \\ p_2 & 1+p_4 & p_6 \\ 0 & 0 & 1 \end{bmatrix}. \quad (1)$$

To detect boundaries, we reduce the influence of texture by decomposing the frame and we compute NLM mask in structure layer. The following weighted affine model in Eq.(2) is calculated by the weighted Lucas-Kanade algorithm.

$$\sum_m W(m)[L^r(T_i m) - L^c(m)], m \in B(f_i^c), \quad (2)$$

where L^c and L^r define the input candidate and reference frames. f_i^c is a matched feature point in candidate frame, and $B(f_i^c)$ is the neighborhood of the feature point f_i^c. $W(m)$ is a weight computed by the NLM scheme.

Confidence map C_i is used to extract support regions for each corresponding feature pair (f_i^r, f_i^c). And all pixels in one support region share the same affine parameters so that they share the same scale parameters.

$$C_i = \begin{cases} 1, |Q_z(L^r(T_i m) - L^c(m))| < \eta_c \\ 0, |Q_z(L^r(T_i m) - L^c(m))| \geq \eta_c \end{cases}, \quad (3)$$

where η_c is a predefined threshold. Q_z is the blurring matrix assumed known in this work.

In order to improve the robustness of confidence map and preserve boundaries, Canny detector is used to detect edge points which provide useful information for support region searching. And instead of a pixel, we use a ball to search the field to avoid those gaps between points, proposed as the trapped ball method [9].

Besides, the obtained segmentation is fragmentized. The fragmentation problem makes it difficult to find feature points in many small regions and the scale of these regions cannot be obtained accurately. The performance of our method is affected by inaccurate patch matching. To solve this problem, the method of hierarchical clustering is utilized to investigate grouping regions. So we get the average scale parameters in X and Y axis in affine model for one region and group two clusters (regions) with the shortest scale distance together every time. In addition, a threshold Z is set to make sure that regions with a big difference of scale cannot be clustered together.

$$\min_{i,j} d_{r_i,r_j} < Z, d_{r_i,r_j} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}, \quad (4)$$

where d_{r_i,r_j} represents the scale distance between region r_i and r_j. x_i and y_i refers to the average scale parameters in X and Y axis in region r_i.

When clusters at one level reach the threshold, the process of clustering ends. Morphological dilation also applies to the image. Those tiny holes in the segmentation are then filled to optimize the segmentation image without ruining the edge. The final segmentation result is obtained for the next step.

B. Scale-Invariant Patch Translation

Based on the segmentation and feature information extracted by SIFT in Sec.III-A, the local scale in every region is calculated and patches are modified into the same scale. Then we follow the traditional NLM to reconstruct the frame.

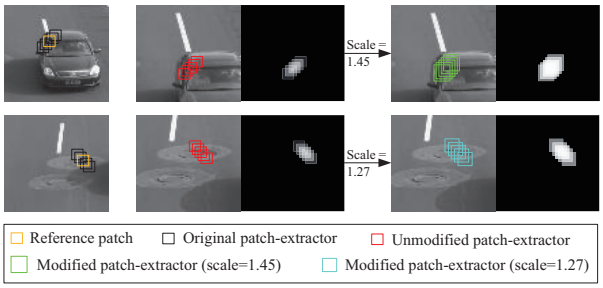

Fig. 4. Comparison of unmodified and modified patch in patch matching.

First, feature information extracted by SIFT in Sec.III-A is used to do weighted summing for every region in each frame. $s_{t,p}$ is obtained as the average scale difference of all matched points in corresponding regions of two frames.

$$s_{t,p} = \frac{1}{|R_{t,p}|} \times \left(\Sigma_{i \in R_{t,p}} \frac{s'_{r,p}(i)}{s'_{t,p}(i)} \right), \quad (5)$$

where $s_{t,p}$ means the scale variation in the p-th region between the t-th candidate frame and the reference frame. $R_{t,p}$ is the set of matched point pairs in the p-th region of the t-th candidate frame while $|R_{t,p}|$ is the number of point pairs in $R_{t,p}$. $s'_{r,p}(i)$ and $s'_{t,p}(i)$ respectively stand for the scale of the i-th matched feature point pair in the p-th region in the reference frame and the t-th candidate frame. When a corresponding region cannot be related because of the mismatching of feature points, $s_{t,p}$ can be computed as follows

$$s_{t,p} = \frac{1}{|\Sigma_p R_{t,p}|} \times \left(\Sigma_{i \in \Sigma_p R_{t,p}} \frac{s'_{r,p}(i)}{s'_{t,p}(i)} \right), \quad (6)$$

which equals to the global scale between two frames.

Given those scale differences, corresponding patches in different scales are modified into the same scale by interpolation. For every pixel (i, j) in the p-th region of the t-th candidate frames, the patch centered with this pixel (i, j) is translated to compensate the scale.

$$R_{t,p} \cdot D(i, j, t) = K(s_{t,p}) \cdot G(s_{t,p}, i, j) \cdot y_t, \quad (7)$$

where $R_{t,p} \cdot D(i, j, t)$ represents the modified patch, and y_t defines the t-th candidate frame. $G(s_{t,p}, i, j)$ is the patch extraction operator with the center pixel (i, j) and the scale $s_{t,p}$. $K(s_{t,p})$ is defined as a interpolation operator to modify the patch into the same scale with the region in the reference frame by bicubic. The similarity of every corresponding patch pair, which acts as the weight, is calculated as follows

$$\hat{w}(k, l, i, j, t) = exp\left\{ -\frac{\|G(1, k, l)Y_r - R_{t,p}D(i, j, t)\|_2^2}{2\hat{\sigma}^2} \right\},$$

where Y_r is the interpolated high resolution reference frame by bicubic. In Fig.4, when patch is modified, weights between similar patches are higher (lighter blocks in the figure), which means the proposed method provides more useful information.

To reconstruct the reference frame, we calculate the weighted average of all pixels (i, j) in the equivalent low-resolution neighborhood of pixel (k, l) in the reference frame, $I(k, l)$ in t-th candidate frames. For every pixel (k, l),

$$A(k, l) = \frac{\Sigma_{t \in [1, \cdots, T]} \Sigma_{(i,j) \in I(k,l)} \hat{w}(k, l, i, j, t) y_t(i, j)}{\Sigma_{t \in [1, \cdots, T]} \Sigma_{(i,j) \in I(k,l)} \hat{w}(k, l, i, j, t)}, \quad (8)$$

where $A(k, l)$ is the reconstruction result of the pixel (k, l).

C. Segmentation-Based Scale-Invariant NLM Algorithm

In this subsection, we integrate all the processes above and describe the proposed Segmentation-Based Scale-Invariant NLM SR algorithm as follows.

Step 1: **Feature Points Matching**
Detect feature point correspondences and compute weighted affine models shown in Eq.(2).

Step 2: **Segmentation**
Segment the image based on parameter model and find out which region each pixel belongs to.

Step 3: **Scale Detection**
Compute the scales $s_{t,p}$ in different regions $R_{t,p}$ as Eq.(5) or Eq.(6).

Step 4: **Patch Modification**
Modify similar patches into same scale as Eq.(7).

Step 5: **Reconstruction**
Calculate the weight for NLM and the reconstruction result $A(k,l)$ is shown in Eq.(8).

IV. Experimental Results

To evaluate the effectiveness of the proposed method, we conduct experiments of $2\times$ super resolution on several test sets. The LR input images are generated from the original HR images by downsampling with bicubic method. We find and shoot some sequences with zooming to test the algorithm. There are five video sequences, *Walk*, *Girl*, *Man*, *Owl* and *Road*. These sequences have been released on our website*.

In the experiments, we compare the proposed algorithm with traditional NLM SR. And to demonstrate the necessity of segmentation, we specially propose an algorithm to reconstruct the frame only considering the global scale of the frame without segmentation, called Scale-Compensated NLM SR (SC NLM). The algorithm is evaluated in both objective and subjective ways. The objective results are measured by Peak Signal to Noise Ratio (PSNR) in Table I. And the subjective results are shown in Fig.5.

TABLE I
PSNR of SR results in sequences

Sequence	NLM SR	SC NLM	Proposed
Walk	23.51	24.00	**24.98**
Girl	22.84	23.21	**23.45**
Man	24.94	25.62	**25.97**
Owl	28.06	28.17	**28.18**
Road	26.77	26.90	**26.93**

It is shown in Table I that our proposed algorithm performs better than NLM SR and SC NLM SR. The average gain of our proposed algorithm is 0.678 dB compared with NLM. And on test set *Walk*, due to the obvious zooming of camera and object motion, scale changes significantly between frames. Our algorithm has a large gain of 1.47 dB. This proves that the segmentation-based scale-invariant NLM SR algorithm we propose extracts more effective information from the sequences. And it is necessary to segment the image so that we take local scale changes into account instead of only considering global scale change.

*http://www.icst.pku.edu.cn/course/icb/Projects/SSI-NLM.html

Fig. 5. Subjective comparison of different algorithms on *Man*. (a) Original frame, (b) An result of NLM SR , (c) SC NLM SR, (d) Segmentation-Based Scale-Invariant NLM.

Subjective results on test set *Man* in Fig.5 show zoomed comparison of the face and car window part in the original image by different methods. Compared with NLM, Fig.5(d) reduces blocking effects and artifacts along the edges.

V. Conclusions

In this work, according to NLM SR framework, we focus on how to make the most of similar information in adjacent frames. Parameter model is used to segment the frame. And we compensate the patch to make it scale-invariant and accurate for NLM. Experimental results show the proposed method outperforms other methods in both objective and subjective quality.

References

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

[4] Yue Zhuo, Jiaying Liu, Jie Ren and Zongming Guo, Nonlocal Based Super Resolution with Rotation Invariance and Search Window Relocation, *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Kyoto, Japan, Mar. 2012.

[5] Mengyan Wang, Jiaying Liu, Wei Bai and Zongming Guo, Illumination-Invariance and Nonlocal Mean Based Super Resolution, *IEEE International Symposium on Circuits and Systems*, Beijing, China, May 2013.

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

[7] B. D. Lucas and T. Kanade, An iterative image registration technique with an application to stereo vision, *Proc. Imaging Understanding Workshop*, pp. 121-130, 1982.

[8] Y. Zhuo, J. Liu, M. Li and Z. Guo. Super Resolution with Edge-Constrained Motion Estimation, *Proc. Asia Pacific Signal and Information Processing Association*, Dec. 2012.

[9] S. Zhang, T. Chen, Y. Zhang, S. Hu and R. Martin. Vectorizing Cartoon Animations, *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 4, pp. 618-629, July. 2009.