Multi-frame Super Resolution Using Refined Exploration of Extensive Self-examples*

Wei Bai, Jiaying Liu, Mading Li, and Zongming Guo**

Institute of Computer Science and Technology, Peking University,
Beijing, P.R. China 100871

Abstract. The multi-frame super resolution (SR) problem is to generate high resolution (HR) images by referring to a sequence of low resolution (LR) images. However, traditional multi-frame SR methods fail to take full advantage of the redundancy in LR images. In this paper, we present a novel algorithm using a refined example-based SR framework to cope with this problem. The refined framework includes two innovative points. First, based upon a thorough study of multi-frame and single frame statistics, we extend the single frame example-based scheme to multi-frame. Instead of training an external dictionary, we search for examples in the image pyramids of the LR inputs, *i.e.*, a set of multi-resolution images derived from the input LRs. Second, we propose a new metric to find similar image patches, which not only considers the intensity and structure features of a patch but also adaptively balances between these two parts. With the refined framework, we are able to make the utmost of the redundancy in LR images to facilitate the SR process. As can be seen from the experiments, it is efficient in preserving structural features. Experimental results also show that our algorithm outperforms state-of-the-art methods on test sequences, achieving the average PSNR gain by up to 1.2dB.

1 Introduction

The super resolution process of a LR image is inverse to the LR imaging process, during which partial high-frequency information is lost. And for the SR problem, such loss leads to non-unique solutions, making the problem ill-posed. Multi-frame SR methods make use of their redundancy in LR information to constrain the solution space. The problem is to reconstruct a HR image from a sequence of low-resolution images. In the literature, these LR images can either be images acquired from the same scene but slightly differ in viewing angles, or a sequence of consecutive video frames. A key point in the former situation is that the LR images should be sub-pixel aligned. Integer pixel alignment is meaningless because that supplies the same amount of information as the single image. In consecutive video frames, information not included in one frame may

* This work was supported by National Natural Science Foundation of China under contract No.61071082, National Basic Research Program (973 Program) of China under contract No.2009CB320907 and Doctoral Fund of Ministry of Education of China under contract No.20110001120117.

** Corresponding author.

S. Li et al. (Eds.): MMM 2013, Part I, LNCS 7732, pp. 403–413, 2013.

appear in adjacent frames. In a word, multi-frame SR approaches exploit redundancy in input LRs for extra information and exchange temporal resolution for spatial resolution.

As current SR methods become more and more sophisticated, people demand more details restored from the LR input. Details are high frequency information. To achieve this, contemporary SR methods either apply pre-defined priori knowledge or refer to a learned dictionary for external information. For example, Tsai and Huang [1] proposed the frequency domain method by transforming the LR image data into the discrete Fourier transform (DFT) domain. Here the relationship between DFT coefficients of LR and HR images can be considered as priori knowledge. Farsiu *et al.* [2] also proposed an l_1-norm regularization method based on a bilateral total variation. However, the performance of these methods degrades rapidly when applied with a large magnification factor, which constrains their application. In recent years, example-based SR methods draw tremendous attention around the world because of their simplicity and potential to break the upper bound of magnification factor compared with the aforementioned algorithms. They usually depend on extra information provided by an image database. Freeman *et al.* [3] estimated high frequency details from a large training set of HR images that encode the relationship between HR and LR images. Glasner *et al.* verify the internal similarity in natural images statistically in [4], meaning that it is feasible to perform the example-based SR from a single image. However, these methods fail to take into consideration the structural features while exploiting examples for extra information.

The foundation of example-based SR is the recurrence of small image patches in different resolutions. These HR/LR pairs indicate how the input LR patches be super resolved. A typical example-based SR algorithm [4] is presented in the following steps:

- Let L be the LR image while H is the HR image. B is the blur kernel. Thus the imaging model is formulated as:

$$L_j(p) = H * B_j = \sum_{q_i \in S_{B_j}} H(q_i) \cdot B_j(q_i - q), \tag{1}$$

  where p is a pixel in the LR image, corresponding to q in the HR image. And the HR pixels q_i belong to the support of kernel B, which centers at q.
- Down-sample the input image L to a cascade of scales, comprising an image pyramid with multiple resolutions.
- For each patch in L, search for similar patches in the above image pyramid. If the similar patches are in lower scales than the input scale, their parent patches are of higher resolution, thus providing examples for the upsampling of the LR patch. Similar patches of the same scale as the input scale also count. By fitting these parent patches or similar patches in the imaging model, the solution space of (1) is constrained.

In this scheme, apparently, it is very important to find plausible examples with enough information, especially when we are handling the multi-frame SR problem. Similar patches recur in multiple frames, more frequent than in a single frame according to the statistics, making example-based method a proper choice. Grounding on this, we

propose a new metric to search for reliable examples whereas lower the search cost at the same time. Our method, REESE (refined exploration of extensive self-examples), takes into account of the structure feature and intensity feature of a local image patch and search its nearest neighbors in a multi-frame and multi-resolution image set, which is derived totally from the input images. With all the similar patches constraining the final HR result, we utilize numerical method to solve the least square problem. Experiments show the superiority of our method, which not only magnifies images efficiently but also preserves the structure details better.

The rest of this paper is organized as follows: Section 2 describes each part of the proposed algorithm in detail. Experimental results are shown in Section 3. Finally, concluding remarks are given in Section 4.

2 Proposed Multi-frame SR Algorithm

2.1 Motivation and Algorithm Framework Overview

According to the prior work that patches recur in a image pyramid, we can easily deduce that multiple frames can provide more similar patches. Fig.1 demonstrates this assumption by comparing the possibilities to find a certain number of similar patches in a single frame and multi-frame at different scales. The horizontal axis shows the number of similar patches found in a frame or frames while the vertical axis indicates the corresponding percentage of input patches. Thus we can come to the aforementioned conclusion that it's reasonable to introduce the self-similarity property to solve the multi-frame SR problem.

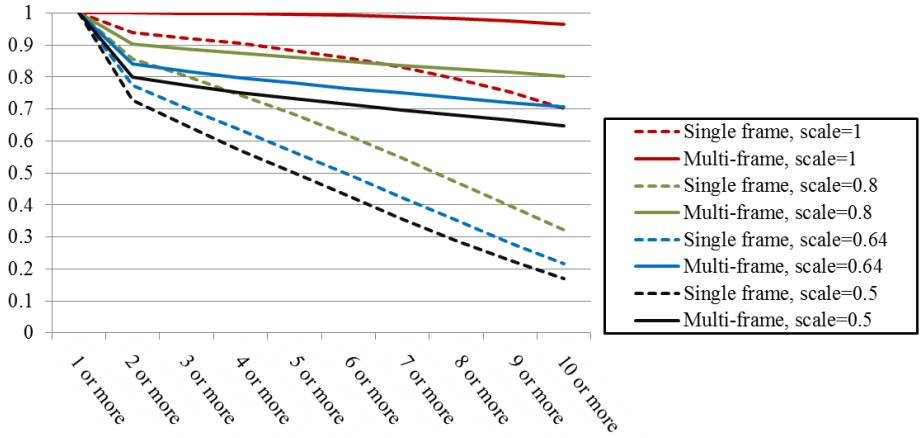


Fig. 1. Comparison of the amount of similar patches found in single frame and multi-frame, searched at different scales

Motivated by the existence of extensive examples in multiple frames, we propose an algorithm which follows the process illustrated in Fig.2. For the convenience of expression, the input LR images are represented as $\{..., L_0^{n-1}, L_0^n, L_0^{n+1}, ...\}$, and L_0^n is the

to-be-magnified image. The down-sampled images are denoted as $\{..., L_{-k}^{n-1}, L_{-k}^{n}, L_{-k}^{n+1}, ...\}$, $k = 1, 2, 3, ...$, where the down-sampling scale is α^{-k}, $\alpha > 1$. In order to get the corresponding HR image, L_k^n, we need to solve a least square problem. Generally, it can be presented as:

$$min(B * L_k^n \downarrow -L_0^n)^2, \tag{2}$$

where $\downarrow$ denotes the down-sampling process.

At essence, the fundamental idea is for the patch at every pixel of L_0^n, we search for its similar patches in both $\{..., L_0^{n-1}, L_0^n, L_0^{n+1}, ...\}$ and $\{..., L_{-k}^{n-1}, L_{-k}^n, L_{-k}^{n+1}, ...\}$. Theoretically, a global search for similar patches needs to be done at each pixel in the to-be-super-resolved image, L_0^n. However, it is a huge amount of calculation to do a global search, which is very time-consuming. So we can take a pre-processing step before the search part. This method is elaborately described in [5], using a motion estimation process to relocate the search window, thus reducing search cost. Then the similar patches in $\{..., L_0^{n-1}, L_0^n, L_0^{n+1}, ...\}$ and the parent patches of those in $\{..., L_{-k}^{n-1}, L_{-k}^n, L_{-k}^{n+1}, ...\}$, as the blue line show in the diagram, will render constraints on the final HR image in (2), resulting in an optimal solution.

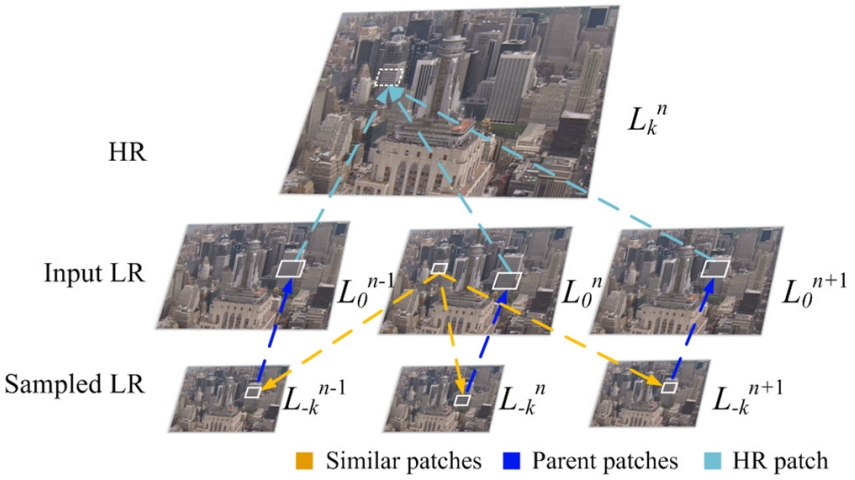


Fig. 2. The framework of the proposed multi-frame SR algorithm

2.2 Structure-Preserving Similarity Measure

In previous work, patch similarity is measured by GW-SSD, which is short for gaussian-weighted SSD (sum of squared difference). It assigns greater weight to pixels that are closer to the center of the patch. However, this is not enough to find the most structurally similar patches. GW-SSD just discriminates centering pixels from peripheral pixels, ignoring different structural features of each patch. Thus, we introduce a new similarity measure to define the distance between two patches. It takes both intensity

similarity and structure similarity into consideration, corresponding to the two terms in the following formulation, respectively.

$$dist(P_{x_1,y_1}, P_{x_2,y_2}) = \lambda \cdot d_i(P_{x_1,y_1}, P_{x_2,y_2}) + (1-\lambda) \cdot d_s(P_{x_1,y_1}, P_{x_2,y_2}), \quad (3)$$

where P_{x_1,y_1} and P_{x_2,y_2} are two patches centered at the pixel (x_1, y_1) and (x_2, y_2). $dist(P_{x_1,y_1}, P_{x_2,y_2})$ represents the distance between the two patches. d_i is the intensity term, while d_s is the structure term. λ is a parameter to weight the two terms which depends on the complexity of the image to be super resolved. The conventional GW-SSD is utilized to calculate the intensity distance:

$$d_i(P_{x_1,y_1}, P_{x_2,y_2}) = \sum_{x,y} G_\sigma \cdot (P_{x_1,y_1}(x,y) - P_{x_2,y_2}(x,y))^2, \quad (4)$$

where G_σ stands for the gaussian kernel, which is the same size as the patches.

As to the structure similarity measure, we utilize the covariance matrix to extract the local structure feature of the patches. A covariance matrix can reflect to what extent neighboring pixels rely on each other. In other words, it implicitly indicates the structure of a local area. The method of calculating the covariance matrix is stated below.

1. Let $P_{x,y}$ denote a patch centered at the point (x, y) in L_0^n. For each pixel in $P_{x,y}$, we build a sample vector V_i. Take point (x, y) (the 3-tagged pixel) for example, as shown in Fig.3, the pixels tagged from number 1 to 5 compose the elements of the sample vector. Note that, V_i is a 1×5 row vector.

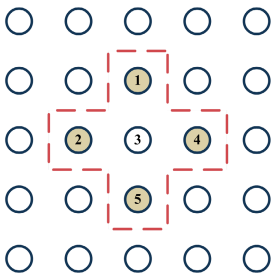


Fig. 3. Formation of sample vectors

2. Assume the patch size to be $N*N$, *i.e.* N^2 pixels. Their sample vectors are denoted as $V_i, i = 1, 2, 3, ..., N^2$, so the patch vector $V_{x,y}$ can be written as:

$$V_{x,y} = \begin{pmatrix} V_1 \\ V_2 \\ ... \\ V_{N^2} \end{pmatrix} \quad (5)$$

3. Then we calculate the self-covariance matrix of $P_{x,y}$'s sample vector $V_{x,y}$, represented as $C_{x,y}$. Finally the structure similarity measure is formulated as:

$$d_s(P_{x_1,y_1}, P_{x_2,y_2}) = \sum_{x,y} G_\sigma \cdot (C_{x_1,y_1}(x,y) - C_{x_2,y_2}(x,y))^2. \tag{6}$$

After we get the intensity and the structure term, the weight parameter λ is applied as mentioned before. In practice, the distance changes as the value of λ varies, making a fixed λ inappropriate. Thus, considering the impact of parameter selection, we adaptively choose λ based upon the smoothness of an image patch. If the patch is not so smooth, *i.e.*, there are plenty of textures, the weight of the structure term should be magnified. In [6], the gradient strength in two perpendicular directions is used to measure the smoothness of local image patches.

$$S_i = \frac{1}{n_i} \sum_{(x,y)\in b_i} S(x,y) = \frac{1}{n_i} \sum_{(x,y)\in b_i} \nabla L(x,y) \cdot \nabla L(x,y)^T, \tag{7}$$

In the above equation, n_i stands for the number of pixels in block b_i. Let $\lambda_1^{(i)}$ and $\lambda_2^{(i)}$ denote the eigenvalues of matrix S_i, which represent the gradient strength then the smoothness s_i of a block b_i is defined as:

$$s_i = |\lambda_1^{(i)}| + |\lambda_2^{(i)}|, \tag{8}$$

Thus, the weight parameter λ can be formulated as follows with μ a constant.

$$\lambda = exp\left(-\frac{s_i}{2\mu^2}\right), \tag{9}$$

2.3 Solving the Weighted Least Square Problem

We use the metric described in Sec.2.2 to exploit similar patches. These patches differ slightly in content, thus resulting in different contribution to the eventual HR patch. So we give them weights as follows, where P_{x_1,y_1} is the reference patch in L_0^n and σ is a parameter.

$$weight_{P_{x,y}} = G_\sigma * exp\left(-\frac{dist(P_{x_1,y_1}, P_{x,y})}{2\sigma^2}\right), \tag{10}$$

Considering the difference between similar patches, the aforementioned least square problem is further extended to a weighted least square problem, shown as below:

$$min\frac{1}{2}(B * L_k^n \downarrow - L_0^n)^T W (B * L_k^n \downarrow - L_0^n), \tag{11}$$

where W is a diagonal matrix, composed of each patch's weight calculated by eq.(10). Thus, we see that we have weighted constraints at the unknown pixels of the images in the higher resolution image L_k^n and we can use gradient descent or other numerical methods to solve the weighted least square problem.

3 Experimental Results

The experiments are performed using MATLAB R2010a on Intel Core CPU 2.4GHz Microsoft Windows platform. All the test image frames are blurred and decimated by a factor of 1:2 (in each axis), and then contaminated by an additive noise with standard deviation 2.

To demonstrate the validity of our proposed method, we first test on single images to perform single image SR. The size of the patch is set as 5*5, and the cascade of low-resolution images is simplified to a 1/2 size reduced one of the input image. In Table 1, we compare our result (S-REESE, single image REESE) with those obtained by Bicubic, [4], and [7], which are relatively state-of-the-art methods that can be found in the literature.

Specifically, since we utilize a similar self-similarity framework with the one in [4], we compare our results with Glasner's to verify that the proposed new metric for similarity does work. We zoom to see the details of HR images super resolved by Glasner's method and the proposed method, as Fig.5 shows. We see that by our algorithm the jaggy effect is reduced and the edges are clearer.

Fig.4 presents another group of results obtained by KR, Jurio's [8] and the proposed algorithm S-REESE. When we zoom in to see the details (best viewed on screen), for example, as presented in *Lena*, we can see that KR over-smoothes the image in contrary to Jurio's, which over-sharps the image even to render jaggies. Our method seeks to balance between smoothness and sharpness.

Then we implement the complete version of the proposed algorithm, *i.e.*, for the input video sequence, we enhance each frame's resolution with reference to its adjacent frames. The efficiency of the proposed multi-frame super resolution algorithm is evaluated both objectively and subjectively. In the objective part, we compare the PSNRs (average of 5 frames in the experiment) of different image sequences obtained by

Table 1. PSNR (dB) Comparison of Different Methods on Single Image Implementation

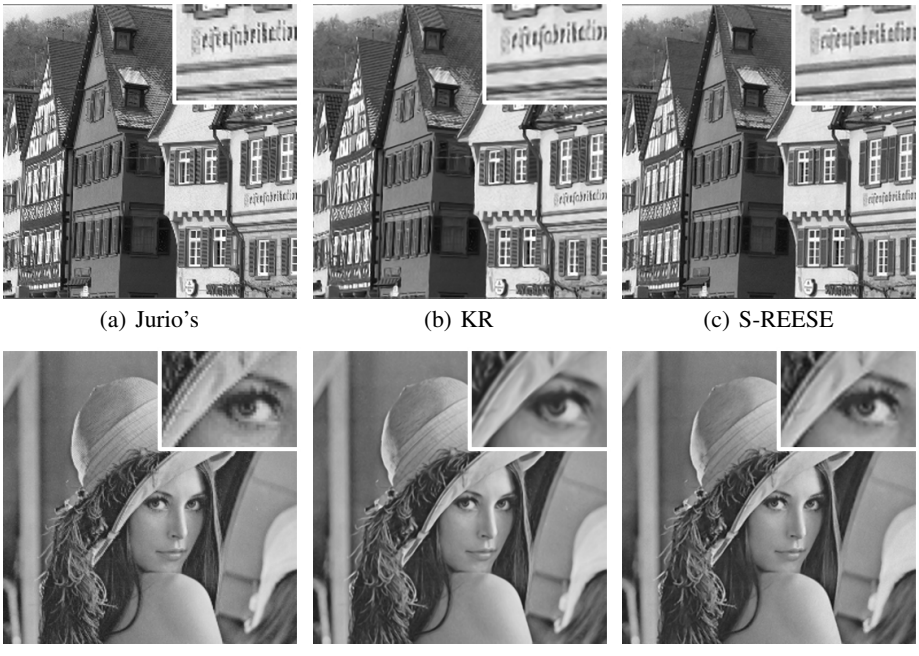
Images	Bicubic	KR [7]	Glasner's [4]	S-REESE (vs. Glasner's)
House	31.29	32.00	31.82	**32.85(+1.03)**
Cameraman	25.17	25.46	26.18	**26.36(+0.18)**
Elaine	32.20	31.89	31.92	**32.51(+0.59)**
Boat	28.80	29.16	29.16	**30.05(+0.90)**
Bridge	25.78	25.59	26.03	**26.69(+0.66)**
Car	29.58	30.01	29.99	**30.91(+0.92)**
Clock	28.81	29.36	29.78	**30.37(+0.59)**
Peppers	29.44	30.14	30.10	**31.17(+1.07)**
Ship	29.42	30.23	30.29	**31.04(+0.75)**
Window	21.74	21.69	22.62	**22.84(+0.22)**
Average	28.22	28.55	28.79	**29.48(+0.69)**

Table 2. Average PSNR (dB) Comparison of Test Sequences

Sequences	Bicubic	3DKR	S-REESE	REESE
Ice	30.58	31.12	33.24	**33.43**
Soccer	27.85	28.25	29.17	**29.26**
Harbour	23.35	23.86	24.63	**24.70**
City	26.82	27.43	27.87	**27.90**
Foreman	31.36	32.68	33.46	**33.62**
Crew	30.16	30.80	32.40	**32.45**
Average	28.35	29.02	30.13	**30.23**

different methods. To add, 3DKR [7] is the mutli-frame version of KR and the single frame REESE is also listed to demonstrate the superiority of multiple frames, as indicated below in Table 2.

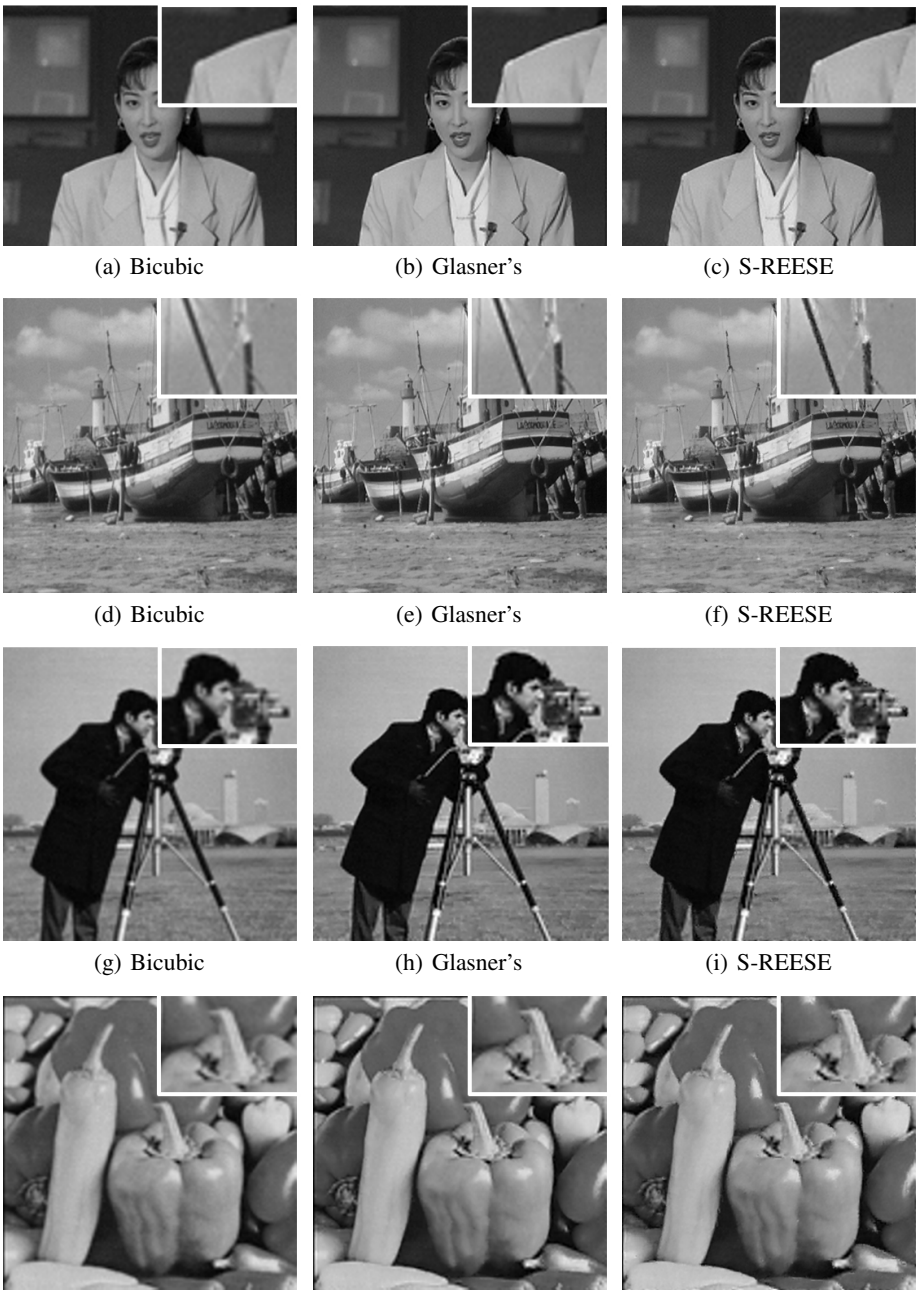
Fig.6 are interpolated results on the video by different methods. For the *Foreman* sequence, our algorithm obtains better result visually with a PSNR of 34.18 dB, achieving a 1.2 dB gain compared with the KR algorithm while for *City*, our algorithm also obtains better result visually with a PSNR of 27.94 dB, achieving a 1.07 dB gain.



(a) Jurio's (b) KR (c) S-REESE

(d) Jurio's (e) KR (f) S-REESE

Fig. 4. Results of our algorithm compared with other methods



(a) Bicubic (b) Glasner's (c) S-REESE

(d) Bicubic (e) Glasner's (f) S-REESE

(g) Bicubic (h) Glasner's (i) S-REESE

(j) Bicubic (k) Glasner's (l) S-REESE

Fig. 5. Comparison of results obtained by Glasner's and S-REESE on test images

(a) Original (b) Bicubic, PSNR=31.68dB

(c) 3DKR, PSNR=32.98dB (d) REESE, PSNR=34.18dB

(e) Original (f) Bicubic, PSNR=26.87dB

(g) 3DKR, PSNR=27.51dB (h) REESE, PSNR=27.94dB

Fig. 6. Comparison of different SR methods on sequence *Foreman*

4 Conclusions

In this work, based on example-based SR framework, we focus on how to find more reliable similar patches. Considering the inadequacy of contemporary example-based methods, we propose a novel method for similar patch exploration. Another contribution of this work is we incorporate multiple frames to enrich the details of the HR images while keeping computing cost as low as possible. Experimental results are free of jaggies and of high quality.

References

[1] Tsai, R., Huang, T.: Multiple frame image restoration and registration. In: Advances in Computer Vision and Image Processing, vol. 1, pp. 317–339 (1984)
[2] Farsiu, S., Robinson, D., Elad, M., Milanfar, P.: Fast and robust multi-frame super-resolution. IEEE Transactions on Image Processing 13, 1327–1344 (2003)
[3] Freeman, W., Jones, T., Pasztor, E.: Example-based super-resolution. IEEE Computer Graphics and Applications 22(2), 56–65 (2002)
[4] Glasner, D., Bagon, S., Irani, M.: Super-resolution from a single image. In: IEEE International Conference on Computer Vision, pp. 349–356 (2009)
[5] Zhuo, Y., Liu, J., Ren, J., Guo, Z.: Nonlocal based super resolution with rotation invariance and search window relocation. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 853–856 (2012)
[6] Su, H., Tang, L., Wu, Y., Tretter, D., Zhou, J.: Spatially adaptive block-based super-resolution. IEEE Trans. on Image Processing 21(3), 1031–1045 (2012)
[7] Takeda, H., Milanfar, P., Protter, M., Elad, M.: Super-resolution without explicit subpixel motion estimation. IEEE Transactions on Image Processing 18(9) (2009)
[8] Jurio, A., Pagola, M., Mesiar, R., Beliakov, G., Bustince, H.: Image magnification using interval information. IEEE Trans. on Image Processing 20(11), 3112–3123 (2011)