

文章编号:2095-6134(2016)05-0596-08

Super resolution reconstruction of single image with denoising and upscaling^{*}

LIU Xiao¹, GUO Tiande^{1,2}, HAN Congying^{1,2†}, LI Mingqiang¹

(1 School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China;

2 Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences, Beijing 100190, China)

(Received 25 December 2015; Revised 15 March 2016)

Liu X, Han C Y, Guo T D, et al. Super resolution reconstruction of single image with denoising and upscaling[J]. Journal of University of Chinese Academy of Sciences, 2016, 33(5):596-603.

Abstract In this study, a novel approach to single image resolution reconstruction is proposed based on nonlocal means, total variation-regularization, and sparse coding. Firstly, low-resolution (LR) image is denoised by nonlocal means which preserves geometric structure well. Then, high-resolution (HR) regularization-based component is obtained from up-scaling the LR image by using a reconstruction model. Image generation process is combined with total-variation regularization so that new image maintains part of the sharpness of the edges and some details. Meanwhile, the HR learning-based component is reconstructed by exploring the sparse coding which explores the co-occurrence relationship between LR training patches and their corresponding HR high-frequency patches. The regularization-based component and the learning-based component are combined to obtain an initial HR image. Finally, the global reconstruction constraint is applied to the initial image for making the final HR image natural. Experimental results show that our method is natural and robust.

Key words super resolution; nonlocal means; total variation; sparse coding

CLC number: O221.7; O235 **Document code:** A **doi:** 10.7523/j.issn.2095-6134.2016.05.004

具有去噪和尺度放大的单幅图像超分辨率重建

刘 晓¹, 郭田德^{1,2}, 韩丛英^{1,2}, 李明强¹

(1 中国科学院大学数学科学学院, 北京 100049; 2 中国科学院大数据挖掘与知识管理重点实验室, 北京 100190)

摘 要 采用非局部均值, 全变差正则化和稀疏编码, 提出一种新的单幅图像超分辨率重建方法. 首先, 用非局部均值方法对低分辨率图像进行去噪处理, 可保证几何结构的一致性. 然后, 将去噪后的低分辨率图像代入重建模型得到高分辨率图像的低频分量, 而高分辨率图像的高频分量则通过学习的方法在数据库中查找得到. 将得到的低频分量与高频分量相加得到初始

^{*} Supported by the National Nature Science Foundation of China(11331012, 11571014, 71271204)

[†]Corresponding author, E-mail: hancy@ucas.ac.cn

的高分辨率图像的恢复结果. 最后对初始的恢复图像进行一些全局优化即可得到最终恢复的高分辨率图像结果. 实验结果证明该方法得到的高分辨率图像结果具有鲁棒性且更自然.

关键词 超分辨率; 非局部均值; 全变差; 稀疏编码

In many practical applications, high-resolution (HR) images are often desired in many practical applications such as video surveillance, biometrics identification, medical imaging, and so on. However, due to the limitations of physical imaging devices, the images we observe are not always at a desired resolution level. The apparent aliasing effects often seen in digital images are caused by the limited number of charge-coupled device (CCD) pixels used in digital device. Using denser CCD arrays (with smaller pixel size) not only increases the cost but also adds more noises^[1]. Moreover, there are many low-resolution (LR) valuable images and available videos. The goal of super-resolution (SR) is to recover a high-resolution image from one or a few low-resolution images.

In the SR literature, SR approaches are classified into three groups: interpolation-based methods^[2-4], regularization-based methods^[5-6] and example-based methods^[7-10].

Interpolation-based methods upscale the size of an LR image by estimating the pixels in the HR grids via a base function^[2,4] or interpolation kernel^[3]. This kind of method always performs well in low-frequency areas (smooth areas) but it cannot recover high-frequency areas (edge) because it often blurs image along edges.

Regularization-based methods assume that the high-frequency information lost in an LR image is split across multiple LR images with sub-pixel misalignments. There are many different priors used in these methods, e. g. nonlocal means (NLM)^[6] and total-variation (TV)^[5]. However, in practical applications, it is hard to obtain these previous priors. Moreover, if HR images are computed by simulating the image formation process, these approaches are limited^[11].

In example learning-based SR methods, the local patches in the desired HR image and those in

the input LR image are assumed to admit a sparse approximation over the same dictionary learned from the HR and LR database. The missing HR pixels can be learnt from low-resolution and high-resolution pairs of examples in the database. Freeman et al.^[7] first proposed a relation model between LR image patches and corresponding HR image patches by using the Markov network. Yang et al.^[8] created a database which combines LR image patches with HR image patches and used sparse coding to reconstruct HR images. Based on an observation that natural images tend to contain repetitive visual content, Glasner et al.^[9] proposed a super resolution method from a single image. Li et al.^[10] proposed a single image super-resolution reconstruction method, which was based on global non-zero gradient penalty and non-local Laplacian sparse coding.

1 Our proposed approach

In this study, we propose a novel approach based on nonlocal means, total variation-regularization, and sparse coding. The overall framework of the proposed approach is shown in Fig. 1. As shown in Fig. 1, firstly, LR image is denoised by using nonlocal means (NLM)^[6] which preserves geometric structure well. This denoised LR image is used twice to produce two components in initial HR image. The regularization-based component is obtained from a reconstruction model with total variation-regularization. The regularization-based component maintains part of the sharpness of the edges and some details. Meanwhile, the HR learning-based component is reconstructed by exploring the sparse coding which explores the co-occurrence relationship between LR training patches and their corresponding HR high-frequency patches. Finally, the global reconstruction constraint is applied to the initial image for making the final HR image natural.

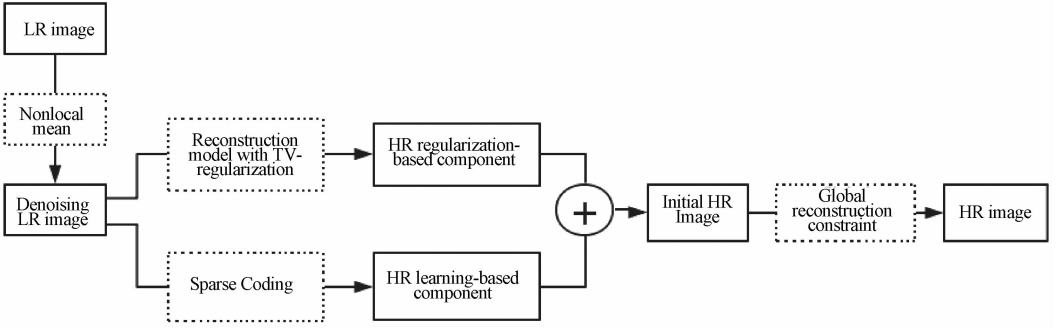


Fig. 1 SR framework

1.1 Regularization-based component reconstruction

In practical applications, the real images are always noisy. In this study, we use noisy LR image as the input to adapt to real-world conditions. During the regularization-based component reconstruction, LR image is denoised and upscaled.

1.1.1 Denoising by nonlocal means

We use NLM to denoise the noisy LR image. This approach assumes that image content tends to repeat itself within some similar neighborhoods among natural images. Therefore, the similar patches found in other locations provides more information of the target patch. By averaging all similar patches in the neighborhoods, the target patch is denoised^[6,12].

The target patch is calculated by

$$x(k, l) = \frac{\sum_{(i,j) \in N(k,l)} w[k, l, i, j] y(i, j)}{\sum_{(i,j) \in N(k,l)} w[k, l, i, j]}, \quad (1)$$

where $N(k, l)$ stands for the neighborhood of the pixel (k, l) in the input image and the term $w[k, l, i, j]$ is the weight for the (i, j) -th neighborhood pixel of pixel (k, l) . The input pixels are $y(i, j)$, and the pixel value of output image in (k, l) is $x(k, l)$.

The weights $w[k, l, i, j]$ for each pixel are computed based on both radiometric (gray-level) proximity and geometric proximity between the pixels, namely

$$w[i, j, k, l] = f(\sqrt{(k - i)^2 + (l - j)^2}) \cdot \exp$$

$$\left\{ - \frac{\sum_{m=-q}^q \sum_{n=-q}^q [y(k + m, l + n) - y(i + m, j + n)]^2}{2 \sigma_r^2} \right\}, \quad (2)$$

where the former function f takes the geometric distance into account and is monotonically non-increasing. It may take many forms such as a Gaussian, a box function, a constant, and others. The latter computes the Euclidean distance between two image patches centered around two involved pixels. The size of these patches is $(2q + 1) \times (2q + 1)$. The parameter σ_r controls the effect of the grey-level difference between the two pixels.

After NLM, the output LR image preserves many low-frequencies and some high-frequencies of input LR image. Besides, the output LR image is denoised. Figure 2 (c) shows the denoised LR image of “meadow” image.

1.1.2 Reconstruction model with total variation – regularization

The output image of the previous step should be upscaled to satisfy the demand.

For single image SR problem, the image generation process of an LR image from HR image is defined as

$$Y = DHX + \epsilon, \quad (3)$$

where X and Y are the HR and LR images, respectively, which are rearranged in lexicographic order. D is the decimation operation, H represents the blurring operation, and ϵ is the system noise which is always assumed as Gaussian noise^[13].

Solving (3) is an ill-posed problem^[14]. There exists an infinite number of solutions which satisfy

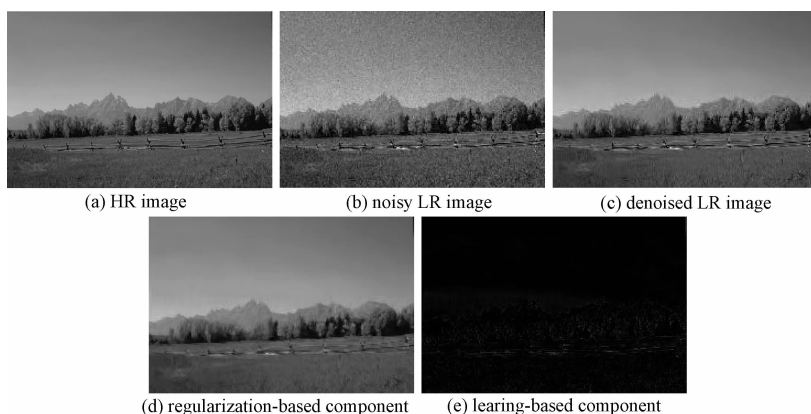


Fig. 2 HR image (“meadow”) decomposed by NLM + TV

(3). Therefore, regularization is a useful method to produce a stable solution. Besides, regularization improves the rate of convergence.

A regularization term is usually implemented as a penalty factor. One of the most successful regularization methods for denoising and deblurring is the total-variation (TV) method^[15]. The TV criterion penalizes the total amount of change in the image as measured by the L_1 norm of the magnitude of the gradient. Then the regularizing function becomes

$$\hat{X} = \arg \min_X \| \mathbf{D} \mathbf{H} \mathbf{X} - \mathbf{Y} \|_2^2 + \lambda \| \nabla \mathbf{X} \|_1, \quad (4)$$

where λ , a global parameter, balances similarity cost against regularization cost. ∇ is the gradient operator.

We use a fast iterative shrinkage/thresholding algorithm (FISTA) to solve this minimization problem^[16]. Figure 2 (d) shows the regularization-based component of “meadow” image.

1.2 Learning-based component reconstruction

Before the learning-based component reconstruction, we need to train several over-complete dictionaries.

1.2.1 design of training databases

Clustering of the training image patch pairs is linked to the effect of dictionary learning. The collected patch pair is HR patch together with LR patch.

Recently, extracting the local geometry structure for the low-resolution image patch in order to boost the prediction accuracy was suggested.

There are many filters to extract the edge information such as a high-pass filter^[17], a set of Gaussian derivative filters^[18], the first-order gradients and second-order gradients of the patches^[4]. Here, we use the LR feature patches, which are the first-order gradients and second-order gradients, to replace the original LR patches.

To obtain better HR high-frequency information, we use HR learning (high-frequency) component to replace the original HR patch. HR learning component is obtained in the following steps. Firstly, the HR image is downsampled by the desired factor s and added Gaussian noise as the noisy LR image. Secondly, the denoised LR image is constructed by NLM. Thirdly, the regularization-based component of HR image is constructed by reconstruction model with TV-regularization. Finally, we can obtain the HR learning component by calculating the residual between HR and HR regularization-based component. Figure 1 shows the decomposition process.

Once these training patch pairs are collected, we divide these pairs into different clusters by the dominant direction of LR patch. The patches with similar dominant direction will be collected into one cluster.

1.2.2 Couple dictionary learning

Sparse coding method is to find a sparse representation of one signal in an over-complete dictionary \mathbf{D} . The dictionary is usually learned from many training sets $\mathbf{X} = [x_1, \dots, x_n] \in R^{m \times n}$. Let \mathbf{D}

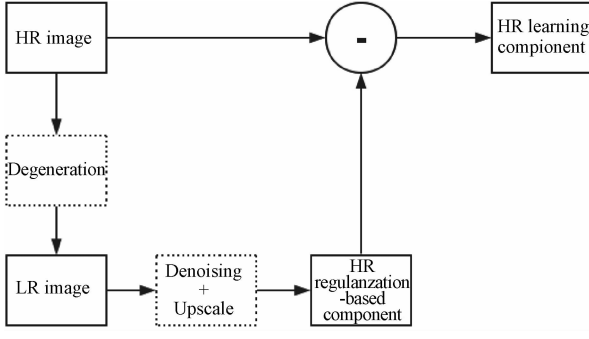


Fig. 3 The decomposition process

$= [d_1, \dots, d_k] \in R^{m \times k}$ be the dictionary matrix, and let $S = [s_1, \dots, s_n] \in R^{k \times n}$ be the coefficient matrix. Then, the objective function of sparse coding is

$$\arg \min_{D, S} \{ \|X - DS\|_F^2 + \alpha \|S\|_1 \} \quad \text{s. t. } \|d_i\|^2 \leq 1, i = 1, 2, \dots, k, \quad (5)$$

where minimizing $\|S\|_1$ is to enforce sparsity, $\|X - DS\|_F^2$ is the empirical loss function, and the parameter α balances fidelity of the approximation to X and sparsity of the solution.

Let the sampled training image patch pair in the j -th cluster be $P = \{X_h^j, Y_l^j\}$, where $X_h^j = \{x_1^j, x_2^j, \dots, x_n^j\}$ are the set of HR learning component patches and $Y_l^j = \{y_1^j, y_2^j, \dots, y_n^j\}$ are the corresponding LR feature patches. Joint dictionary training method is used to train two coupled dictionaries for each cluster. Similar to the treatment in Reb. [8], the single sparse coding problem in the HR learning component patch and the LR feature patch are

$$D_h^j = \arg \min_{D_h^j, S} \{ \|X_h^j - D_h^j S\|_F^2 + \alpha \|S\|_1 \}, \quad (6)$$

and

$$D_l^j = \arg \min_{D_l^j, S} \{ \|Y_l^j - D_l^j S\|_F^2 + \alpha \|S\|_1 \}, \quad (7)$$

where D_h^j and D_l^j represent j -th cluster's coupled dictionaries.

In order to force the HR and LR representations to share the same codes, we combine these objectives

$$\arg \min_{D_h^j, D_l^j, S} \left\{ \frac{1}{N_1} \|X_h^j - D_h^j S\|_F^2 + \frac{1}{N_2} \|Y_l^j - D_l^j S\|_F^2 + \alpha \left(\frac{1}{N_1} + \frac{1}{N_2} \right) \|S\|_1 \right\}, \quad (8)$$

where N_1 and N_2 are the dimensions of the HR learning component and LR feature patch in vector form, respectively. Here, $\frac{1}{N_1}$ and $\frac{1}{N_2}$ balance the two cost terms of (6) and (7). By (6) and (7), (8) can be rewritten as

$$\arg \min_{D_c^j, S} \{ \|B_c^j - D_c^j S\|_F^2 + \hat{\alpha} \|S\|_1 \}, \quad (9)$$

where

$$B_c^j = \left[\frac{1}{\sqrt{N_1}} X_h^j, \frac{1}{\sqrt{N_2}} Y_l^j \right]^T, \\ D_c^j = \left[\frac{1}{\sqrt{N_1}} D_h^j, \frac{1}{\sqrt{N_2}} D_l^j \right]^T, \\ \hat{\alpha} = \alpha \left(\frac{1}{\sqrt{N_1}} + \frac{1}{\sqrt{N_2}} \right).$$

For a given test LR image, which is set as the denoise LR image in our method, we sample patches with overlap and extract four features of these patches. Then, for each feature patch y_i , we adaptively select coupled dictionaries D_h^j and D_l^j from all coupled dictionaries, by computing the dominant orientation of y_i to find the most similar cluster. We compute the sparse coefficient s_i under the proper dictionary D_l^j , and we can obtain the HR patch estimation $X^{j*} = D_h^j s_i$. The whole HR learning component is reconstructed by averaging all the reconstructed patches such as X^{j*} . Figure. 2 (e) shows the learning-based component of “meadow” image. Then the initial HR image X_0 is obtained by adding the reconstructed component and the learning component.

1.3 The global reconstruction constraint

Unexpected artifacts are always found in the initial HR image X_0 because of the simple addition operation. Considering the image generation process of an LR image from HR image, we find the closest image to X_0 which satisfies the image generation process

$$\mathbf{X}^* = \arg \min_{\mathbf{X}} \{ \|\mathbf{D}\mathbf{H}\mathbf{X} - \mathbf{Y}\|_2^2 + c \|\mathbf{X} - \mathbf{X}_0\|_2^2 \}, \tag{10}$$

where c is regularization paramant, \mathbf{X}^* is the final HR image, and \mathbf{Y} is set as the denoised image.

2 Experiments

In our experiments, we perform eleven test images to measure our proposed method by magnifying them two times. All the images come from the Berkeley Segmentation Dataset^{[19] ①}.

2.1 Experimental configuration

The noisy LR image is produced by artificially adding Gaussian noise to original LR image. The standard deviation of Gaussian noise is set to be 10. We denote it as σ .

In (1) and (2), the research zone $N(k,l)$ is limited to a square neighborhood of 21 pixels \times 21 pixels. $q = 1$. $\sigma_r^2 = 0.4\sigma$. The function f is set to be constant, so that it gives no preference to adjacent patches over distant ones. This makes f more robust to various motion patterns.

In (3), the blurring operation \mathbf{H} is set to be an identity matrix.

In (4), global parameter λ is set to be 0.001 and \mathbf{Y} is chosen as the denoised LR image.

In (5), parameter α is set to be 0.15.

During the design of training databases, we have trained 12 clusters in total. For each cluster, the coupled dictionaries are trained from 50 000 patch pairs randomly sampled from 100 training images. The sizes of original LR training patch and the corresponding HR patch are set to be 25 (5 \times 5) and 100 (10 \times 10), respectively. So the LR feature patches which are the first – and second – order gradients of original LR training patch are set to be 100 (5 \times 5 \times 4). Considering computation cost and image quality, the dictionary size is always fixed at 512.

2.2 Results

For showing the effectivity of our proposed method, we use different methods to magnify the

‘Boat’ image two times in Fig. 2. Figure 4 (a) is the original HR image, Figs. 4 (b) to 4 (e) are the mages solved by using other methods, and Fig. 4 (f) is the final HR image by using our method. The input noisy LR image is produced by the generation process in (3). To be convenient to calculate PSNR, we firstly upscale LR image by bicubic interpolation to fix the size of HR image. (See Figs. 4 (b) and 4 (c)). However, Figs. 4 (d), 4 (e), and 4 (f) retain the same size as in Fig. 4 (a).

It is obvious that Fig. 4 (b) is very noisy. After denoising Fig. 4 (b) by NLN, a smooth image in Fig. 4 (c) is obtained and its PSNR value also increases. Figure 4 (d) is obtained by using (4) and \mathbf{Y} is set as in Fig. 4 (c). Figure 4 (e) is the initial HR image which has been described in Fig. 1. Figure 4 (f) is the final result by using (10) to optimize the initial image. It is easy to see that the final image in Fig. 4 (f) is most similar to the original HR image in Fig. 4 (a).

To certify the validity of our method, we deal with eleven images as same as (b) – (f) in Fig. 2. We show the PSNR results in Table 1.

Table 1 PSNR results at different steps					
images	(1)	(2)	(3)	(4)	(5)
1	25.70	27.14	27.80	27.41	27.75
2	24.38	25.10	25.55	25.73	25.91
3	22.79	22.91	22.67	22.68	23.04
4	25.18	26.31	26.47	26.26	26.65
5	24.62	25.65	26.20	25.82	26.11
6	26.82	28.54	28.82	28.50	28.69
7	18.74	18.74	19.09	19.49	19.68
8	24.77	25.16	25.17	24.95	25.36
9	26.49	27.81	28.15	27.25	28.12
10	27.47	29.99	29.82	29.58	30.02
11	24.19	24.62	24.98	24.93	25.28
average	24.65	25.63	25.88	25.69	26.06

Note: (1) original noisy LR image + bicubic, (2) denoised image + bicubic, (3) regularization-based component, (4) initial HR image, and (5) HR image treated by using our proposed method.

In Table 1, it is obvious that original noisy LR image always gives the worst performance because of

① <http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping>.

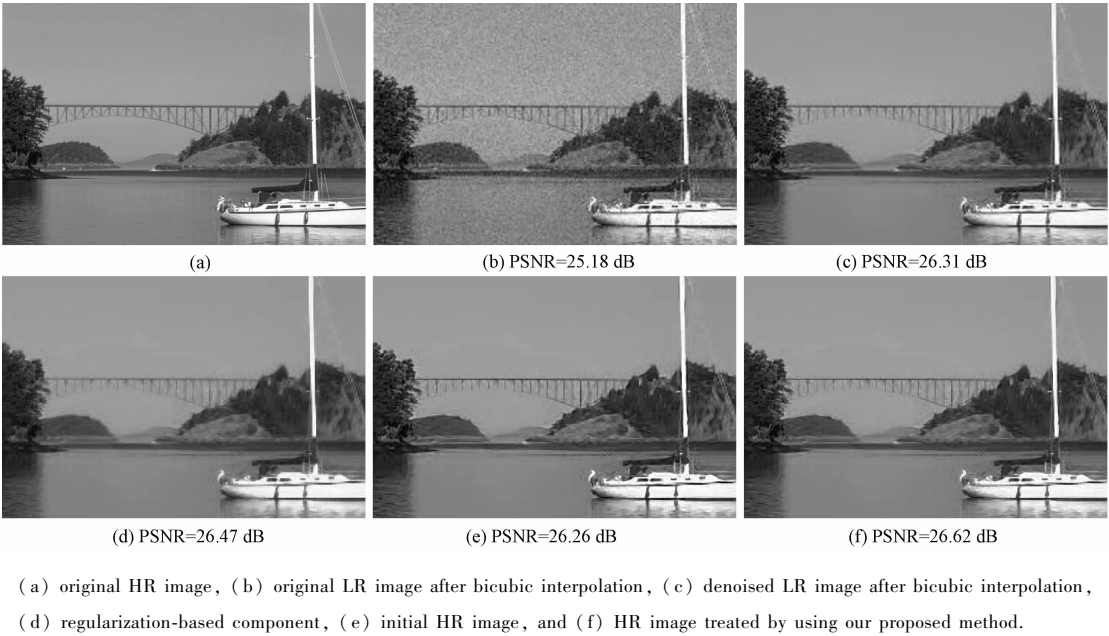
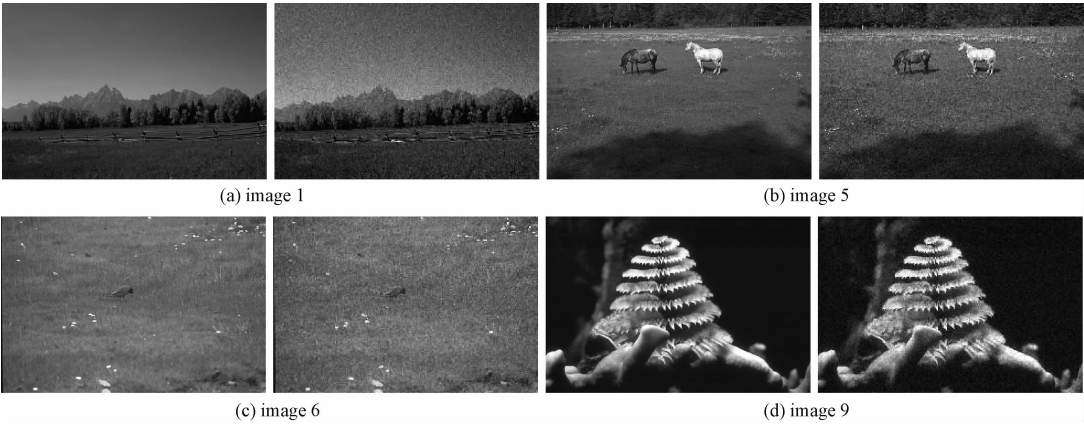


Fig. 4 Sequence “Boat”

the noteworthy noise. The denoised image achieves better results than original noisy LR image, which indicates that denoising is effective. Because of unexpected artifacts, the PSNR of initial HR image is often smaller than that of the regularization-based component. By using the global reconstruction constraint, these artifacts are eliminated and the PSNR always increases.

The regularization-based component sometimes

achieves the best performance. Figure 5 lists these images. Among these images, the important high-frequency information is destroyed by noise. When we try to recovery high-frequency information from the noisy LR image, it is hard to distinguish the true information from noise. The learning-base component has a low confidence so that the regularization-based component without the learning-base component achieves the best result.



Left: original HR images; Right: input noisy LR image.

Fig. 5 HR images whose regularization-based components achieve the best performance and the corresponding noisy LR image

3 Conclusion

In this study, we propose a new approach based

on regularization-based component and learning-based component. We apply the global reconstruction to make the images natural. As the

input image is noisy, we denoise image firstly by nonlocal means. Then the reconstruction-based component is obtained by reconstruction model with total variation-regularization. The learning-based component is produced by sparse coding and the initial HR image is obtained by adding the reconstructed component and the learning component. However, the initial HR image always looks unnatural. Hence, the global reconstruction constraint is used.

Experimental results demonstrate the effectiveness and robustness of the proposed approach.

References

- [1] Park S C, Park M K, Kang M G. Super-resolution image reconstruction: a technical overview [J]. Signal Processing Magazine, IEEE, 2003, 20(3): 21-36.
- [2] Li X, Orchard M T. New edge-directed interpolation [J]. Image Processing, IEEE Transactions on, 2001, 10(10): 1 521-1 527.
- [3] Takeda H, Farsiu S, Milanfar P. Kernel regression for image processing and reconstruction [J]. Image Processing, IEEE Transactions on, 2007, 16(2): 349-366.
- [4] Chang H, Yeung D Y, Xiong Y. Super-resolution through neighbor embedding [C] // Computer Vision and Pattern Recognition, 2004. Proceedings of the 2004 IEEE Computer Society Conference on. IEEE, 2004, 1: 275-282.
- [5] Farsiu S, Robinson M D, Elad M, et al. Fast and robust multiframe super resolution [J]. Image processing, IEEE Transactions on, 2004, 13(10): 1 327-1 344.
- [6] Buades A, Coll B, Morel J M. A non-local algorithm for image denoising [C] // Computer Vision and Pattern Recognition, 2005. IEEE Computer Society Conference on. IEEE, 2005, 2: 60-65.
- [7] Freeman W T, Jones T R, Pasztor E C. Example-based super-resolution [J]. Computer Graphics and Applications, IEEE, 2002, 22(2): 56-65.
- [8] Yang J, Wright J, Huang T, et al. Image super-resolution as sparse representation of raw image patches [C] // Computer Vision and Pattern Recognition, 2008. IEEE Conference on. IEEE, 2008: 1-8.
- [9] Glasner D, Bagon S, Irani M. Super-resolution from a single image [C] // Computer Vision, 2009 IEEE 12th International Conference on. IEEE, 2009: 349-356.
- [10] Li J, Gong W, Li W, et al. Single-image super-resolution reconstruction based on global non-zero gradient penalty and non-local laplacian sparse coding [J]. Digital Signal Processing, 2014, 26: 101-112.
- [11] Lin Z, Shum H Y. Fundamental limits of reconstruction-based superresolution algorithms under local translation [J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2004, 26(1): 83-97.
- [12] Protter M, Elad M, Takeda H, et al. Generalizing the nonlocal-means to super-resolution reconstruction [J]. Image Processing, IEEE Transactions on, 2009, 18(1): 36-51.
- [13] Yu J, Gao X, Tao D, et al. A unified learning framework for single image super-resolution [J]. Neural Networks and Learning Systems, IEEE Transactions on, 2014, 25(4): 780-792.
- [14] Nguyen N, Milanfar P, Golub G. A computationally efficient superresolution image reconstruction algorithm [J]. Image Processing, IEEE Transactions on, 2001, 10(4): 573-583.
- [15] Rudin L I, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms [J]. Physica D: Nonlinear Phenomena, 1992, 60(1): 259-268.
- [16] Beck A, Teboulle M. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems [J]. Image Processing, IEEE Transactions on, 2009, 18(11): 2 419-2 434.
- [17] Freeman W T, Pasztor E C, Carmichael O T. Learning low-level vision [J]. International Journal of Computer Vision, 2000, 40(1): 25-47.
- [18] Sun J, Zheng N N, Tao H, et al. Image hallucination with primal sketch priors [C] // Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on. IEEE, 2003, 2: 729-736.
- [19] Martin D, Fowlkes C, Tal D, et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics [C] // Computer Vision, 2001. Proceedings Eighth IEEE International Conference on. IEEE, 2001, 2: 416-423.