Int. J. Mach. Learn. & Cyber. (2011) 2:15–23
DOI 10.1007/s13042-011-0011-6

ORIGINAL ARTICLE

Single-image super-resolution via local learning

Yi Tang · Pingkun Yan · Yuan Yuan · Xuelong Li

Received: 23 November 2010 / Accepted: 16 January 2011 / Published online: 12 February 2011
© Springer-Verlag 2011

Abstract Nearest neighbor-based algorithms are popular in example-based super-resolution from a single image. The core idea behind such algorithms is that similar images are close in the sense of distance measurement. However, it is well known in the field of machine learning and statistical learning theory that the generalization of the nearest neighbor-based estimation is poor, when complex or high dimensional data are considered. To improve the power of the nearest neighbor-based algorithms in single-image based super-resolution, a local learning method is proposed in this paper. Similar to the nearest neighbor-based algorithms, a local training set is generated according to the similarity between the training samples and a given test sample. For super-resolving the given test sample, a local regression function is learned on the local training set. The generalization of nearest neighbor-based algorithms can be enhanced by the process of local regression. Based on such an idea, we propose a novel local-learning-based algorithm, where kernel ridge regression algorithm is used in local regression for its well generalization. Some experimental results verify the effectiveness and efficiency of the local learning algorithm in single-image based super-resolution.

Keywords Super-resolution · Local learning · Generalization · Reproducing kernel · Kernel ridge regression · Similarity

Y. Tang · P. Yan (✉) · Y. Yuan · X. Li
Center for OPTical IMagery Analysis and Learning (OPTIMAL),
State Key Laboratory of Transient Optics and Photonics,
Xi'an Institute of Optics and Precision Mechanics,
Chinese Academy of Sciences, Xi'an 710119,
Shaanxi, P. R. China
e-mail: pingkun@ieee.org

1 Introduction

1.1 Single-image super-resolution

Super-resolution is a technique for enhancing the resolution of given images. Traditionally, such methods can be categorized as multi-image super-resolution and single-image super-resolution based on the image input. Within the framework of multi-image super-resolution, a high-resolution image will be generated using multiple low images of same scene, e.g. [1,2]. Different low-resolution image contains different details related to the high-resolution image. The focus of multi-image super-resolution is locating and fusing the information of details connected with high-resolution image. Within the framework of single-image super-resolution, high-resolution image is generated with just a single low-resolution image. It is clear that the single-image super-resolution is a more challenging problem in the field of super-resolution for no enough information on high-resolution details in hand.

In order to make meaningful single-image super-resolution, prior information on high-resolution image is needed. The prior information can be presented directly or indirectly, which results in different strategies to deal with single-image super-resolution. First of all, the explicit prior information is the basis of most of single-image super-resolution. Based on the given prior information, the details missing in the low-resolution images can be estimated. The high-resolution image can be generated by combining the details with the given low-resolution image. For example, smoothing prior is extensively used in various interpolation or filtering algorithms for single-image super-resolution [3–6]. Though explicit priors are the representations of intrinsic rules behind images, it is hard to determine what and how much prior should be used in a special problem of

single-image super-resolution. Therefore, such algorithms are out of main stream of study in single-resolution super-resolution recently.

Implicit prior information contains more complex information on high-resolution images than explicit prior information [7, 8]. Generally, implicit prior information is represented by a set of image pairs, which consists of low-resolution images and their corresponding high-resolution images. Various explicit prior information on high-resolution can be found in such set if the size of the set is large enough.

For super-resolving a low-resolution with implicit prior information, the key is to learn the effective rule which matches the given problem of single-image super-resolution. The rule learned from the set of image pairs unveils the relationship between the low- and high-resolution images. Based on such rule, the high-resolution image corresponding to the given low-resolution image can be well estimated.

Single-image super-resolution with implicit prior information or a set of image pairs is tightly connected with machine learning. Many achievements in machine learning, such as local linear embedding (LLE) [9] and support vector regression (SVR) [10] are the tools in learning the relationship between low- and high-resolution images. In fact, the method of super-resolution with implicit prior information is mostly named as learning-based method or example-based method [11–15]. More and more effort is devoted to such method in the field of super-resolution along with the success of machine learning in various applications.

1.2 Related work

In this section, we review some nearest neighbor-based algorithms. For more details about example-based algorithms for single-image super-resolution, we refer the readers to [11, 12, 16].

Example-based method was firstly introduced into single-image super-resolution by Freeman et al. [13, 14] in 2000. Their algorithm is designed based on the nearest neighbor-based estimation i.e. [17]. With patching technique, all given images are represented as a set of patches. For super-resolving a test low-resolution patch, a small set of low-resolution patches which consists of the k-nearest neighbors to the test low-resolution patch is chosen from the set of training set. The high-resolution patches corresponding to the chosen low-resolution patches are considered as hidden states in a Markov network model. By maximizing posterior probability, the high-resolution estimation is chosen from those high-resolution patches. Following the idea, similar nearest neighbor-based algorithms are proposed based on different probability model, e.g. [18–20].

To increase the flexibility of the nearest sample based methods, Chang et al. [15] introduced a novel nearest neighbor embedding algorithm which is motivated by the idea of LLE. Similar to the algorithm proposed by Freeman et al. [14], the k-nearest neighbor low-resolution patches are chosen. However, the high-resolution estimation is not generated by choosing from the high-resolution patches corresponding to those k-nearest neighbor low-resolution patches, but generated by combining these high-resolution patches in a linear style. The coefficients of such linear combination are defined based on the idea of LLE. On the low-resolution manifold, the local optimal linear approximation of the test patch is calculated. With the assumption on the preserving of local linear structures on low- and high-resolution manifolds, the coefficients corresponding to the optimal linear approximation of low-resolution test patch is used to the high-resolution patches. Thus, more information of training samples is used in estimating the high-resolution patch [21].

Though the performance of the nearest neighbor embedding algorithm is better than most of nearest neighbor estimation algorithms, the assumption on the preserving of local linear structures on different manifolds is always doubted. Many efforts are devoted to refine such assumption to make it more convincing. Based on those refined assumptions, different nearest neighbor embedding algorithms are proposed, such as [22–24].

1.3 Our contribution

In this paper, we introduce a local learning framework to enhance the performance of nearest neighbor-based algorithms. Similar to the nearest neighbor-based algorithms, a subset of training samples, named as local training set, is chosen for super-resolving a test low-resolution patch. Based on such local training set, a local regression function can be estimated by machine learning technique. The high-resolution estimation of the test low-resolution patch is generated by the value of the learned regression function on the test low-resolution patch.

The main merits of the local learning framework are as follows.

No additional assumption Traditional nearest neighbor-based algorithms are connected with the assumptions on the relationship between the low- and high-resolution patches, for example, the probability prior in nearest neighbor estimation algorithms and manifold prior in nearest neighbor embedding algorithms. As for our local learning algorithm, such relationship is directly learned

from the local training samples, which make our algorithm independent on those additional priors.

Flexibility The local regression makes our algorithm more flexible in using the information of local training samples. Without strong prior on the relationship between low- and high-resolution patches, the more complex styles of using local training samples are available. In fact, the high-resolution estimation will be generated by some nonlinear combination of local training samples when kernel regression methods are used in local regression. It is completely different from the linear style used in traditional nearest neighbor-based algorithms.

The rest of the paper is organized as follows. In Sect. 2, the used notations are defined and the problem of single-image super-resolution is formulated strictly. In Sect. 3, a naive version of the local learning-based algorithm is presented. By comparing the performance of the local learning-based algorithm with the nearest neighbor-based algorithms, we show the merits of learning in single-image super-resolution. In Sect. 4, a refined version of the local learning-based algorithm is presented. Considering the connection between the generalization of regression algorithms and the number of training samples, the nearest neighbor training set is replaced by a larger local training set. Along with the expanding of the local range, the assumption of local linear structure is unacceptable. Therefore, the nonlinear estimation based on kernel method is adopted in local learning algorithm. In Sect. 5, various experimental results are presented to demonstrate the efficacy of the refined version of local learning-based algorithm. Finally, Sect. 6 concludes the paper.

2 Notations and problem formulation

Let $\mathcal{X}$ be a low-resolution image and $\mathcal{Y}$ its corresponding high-resolution image. Denote the magnification factor between $\mathcal{X}$ and $\mathcal{Y}$ as N.

The low-resolution image $\mathcal{X}$ will be represented as a set of overlapping image patches, that is

$$\mathcal{X} = \{x_i | i = 1, 2, \ldots, k_{\mathcal{X}}\}.$$

Here, x_i is a $l \times l$ image patch, $k_{\mathcal{X}}$ is the number of patches generated from the image $\mathcal{X}$. Similarly, the high-resolution image $\mathcal{Y}$ can be also represented as a set of small image patches $\{y_i | i = 1, 2, \ldots, k_{\mathcal{Y}}\}$. If the size of y_i is set to be $h \times h$ where $h = N \times l$, $k_{\mathcal{Y}} = k_{\mathcal{X}}$ there is. Following image patches are treated as vectors, for example, $x_i \in \mathcal{R}^{l^2}$ and $y_i \in \mathcal{R}^{h^2}$

Within the framework of example-based method for single-image super-resolution, a test low-resolution image $\mathcal{X}^{\text{test}}$ and a set of training image pairs

$$\{(\mathcal{X}^i, \mathcal{Y}^i) | i = 1, 2, \ldots, m\}$$

are available. For super-resolving the image $\mathcal{X}^{\text{test}}$, the patching technique is always used. The test image $\mathcal{X}^{\text{test}}$ is blocked as a test set of patches

$$T = \{x_i^{\text{test}} | i = 1, 2, \ldots, p\}.$$

Meanwhile the set of training image pairs is used to generate the training set S. For ith training image pair $(\mathcal{X}^i, \mathcal{Y}^i)$, a set of training patch pairs is generated, that is

$$S^i = \left\{ \left(x_j^i, y_j^i \right) | j = 1, 2, \ldots, k_i \right\}.$$

The training set S is the union of $S^i, i = 1, 2, \ldots, m$, that is

$$\begin{aligned} S &= \cup_{i=1}^m S^i \\ &= \cup_{i=1}^m \left\{ \left(x_j^i, y_j^i \right) | j = 1, 2, \ldots, k_i \right\}. \end{aligned}$$

Simply, we denote below the training set as

$$S = \{(x_i, y_i) | i = 1, 2, \ldots, n\}.$$

The focus of single-image super-resolution is to estimate the high-resolution image $\hat{\mathcal{Y}}$ corresponding to the test image $\mathcal{X}^{\text{test}}$ with the help of the training set S.

With the information in hand, there are two kinds of relationships can be estimated as showing in Fig. 1. One is low-low relation (LL-relation) which can be estimated by learning on

$$\{x_i^{\text{test}} | i = 1, 2, \ldots, p\} \cup \{x_i | i = 1, 2, \ldots, n\}.$$

LL-relation unveils the compatibility between the test samples and training samples. The other is low-high relation (LH-relation) which can be estimated by learning on training set S. LH-relation unveils the underlying rule of converting a low-resolution image to its corresponding high-resolution image. LL-relation helps us to collect meaningful information in super-resolution, and LH-relation helps us generalize such information to high-resolution patches and reconstructing a high-resolution image.

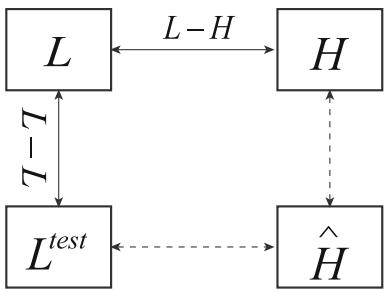

Fig. 1 Two key relationships in example-based super-resolution. *L-L* the relationship between the low-resolution patches, and *L-H* the relationship between the low-resolution patches and their corresponding high-resolution patches

It is clear that traditional nearest neighbor-based algorithms pay more attention on LL-relation. LH-relation is generally replaced by some prior. Though priors simplify the single-image super-resolution algorithms, they will also lower the performance of those algorithms. In the following sections, the machine learning method is introduced to generate sample-based LH-relation, which will enhance the performance of the nearest neighbor-based algorithms.

3 Nearest neighbor-based learning for super-resolution

In this section, we will show the power of learning in single-image super-resolution by comparing the naive local learning-based algorithm with some nearest neighbor-based algorithms.

Similar to most of the nearest neighbor-based algorithms, the local training set

$$S_L^t = \{(x_{t_1}, y_{t_1}), (x_{t_2}, y_{t_2}), \ldots, (x_{t_k}, y_{t_k})\} \subset S$$

corresponding to a test patch x_t^{test} is generated according to the distance between training and test low-resolution patches. When local training set S_L^t is available, a local regression function $f^t(\cdot)$ is estimated by the ridge regression. Then the high-resolution patch $\hat{y}_t$ corresponding to x_t^{test} is estimated by the local regression function, that is

$$\hat{y}_t = f^t\left(x_t^{\text{test}}\right).$$

Consider that any manifold structure can be well approximated by a linear structure in sufficiently small range. A linear function space is used as the hypothesis space. That is

$$\mathcal{H} = \{f(x) = Px | P \in \mathcal{R}^{h^2 \times l^2}, x \in \mathcal{R}^{l^2}\}. \tag{1}$$

The optimal local linear regression function is generated by minimizing the regularized empirical error, that is

$$\begin{aligned} \hat{f}^t &= \arg\min_{f \in \mathcal{H}} \sum_i \|y_{t_i} - f(x_{t_i})\|_2^2 + \lambda \|f\|_{\mathcal{H}}^2 \\ &= \arg\min_{P \in \mathcal{R}^{l^2 \times h^2}} \|Y - PX\|_2^2 + \lambda \|P\|_F^2, \end{aligned} \tag{2}$$

where $X = (x_{t_1}, x_{t_2}, \ldots, x_{t_k})$, $Y = (y_{t_1}, y_{t_2}, \ldots, y_{t_k})$, $\lambda > 0$ is a regularization parameter. Noting the necessary condition of extremum, there exits

$$\hat{P}^t = YX'(XX' + \lambda I)^{-1}, \tag{3}$$

where I is an identity matrix, which results in $\hat{f}^t(x) = \hat{P}^t x$.

The naive version of local learning-based algorithm is presented in Fig. 2. The difference between the nearest neighbor learning algorithm and the nearest neighbor-based algorithms is that the priors are replaced by a set of local regression functions which depends only on the training set. Some experimental results where five nearest neighbors are used show that the more details on high-resolution images are found by the nearest neighbor learning method. For more details, see Figs. 3 and 4.

Input
a set of training image pairs $\{(\mathcal{X}^i, \mathcal{Y}^i)_{i=1}^m\}$, a test image $\mathcal{X}^{\text{test}}$, the size of low-resolution patch $l \times l$, the magnification factor N and the number of nearest neighbors k
Step1 Generating the training set S and the test set T.
Step2 Nearest neighbor learning
For i=1 to $|T|$
a Generating i-th local training set corresponding to the i-th test patch
b Generating i-th local regression function f^i
c Generating i-th high-resolution estimation $\hat{y}_i^{\text{test}} = f^i(x_i^{\text{test}})$
End
Output
Merge all $\hat{y}_i^{\text{test}}$ to obtain high-resolution estimation $\hat{\mathcal{Y}}$

Fig. 2 A nearest neighbor learning-based algorithm for single-image super-resolution

4 Local learning for super-resolution

In the nearest neighbor learning algorithm, each local regression function is trained on a small local training set. According to the results in learning theory [25–27], however, more training samples can lead to better generalization of estimators. Generally, better generalization will be helpful in enhancing the quality of super-resolution within the framework of learning-based method. Therefore, it is helpful to enlarge the sizes of the local training sets. With the increasing of the number of the local training samples, the local linear model maybe inappropriate to simulate the local relationship between the low- and high-resolution patches.

To benefit from the more training samples, a nonlinear model is adopted here. By enlarging the size of local training set, the local linear structure around the test patch will be destroyed. Though the linear model is unacceptable, the model learned from a local training set will be available. The kernel ridge regression (KRR) [10] is adopted to estimate the local nonlinear model for its well generalization in the small sample problem.

The hypothesis space used to define nonlinear models is a reproducing kernel Hilbert space (RKHS)

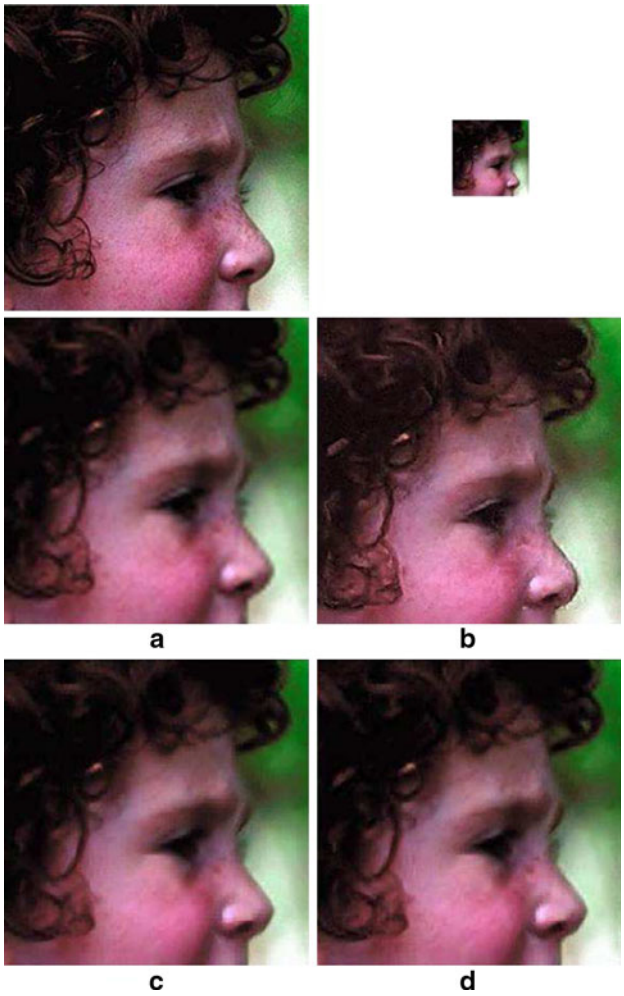

Fig. 3 High-resolution image and the low-resolution test images are shown on the *first line*. The super-resolved images are shown on the *second line*. **a** Bicubic interpolation; **b** nearest neighbor estimation method [14]; **c** nearest neighbor embedding method [15]; **d** nearest neighbor learning (Fig. 2)

$$\mathcal{H}_\kappa = \overline{\text{span}}\left\{ f = \sum_i \alpha_i \kappa(x_i, \cdot) | \alpha_i \in \mathcal{R},\, x_i \in \mathcal{R}^{l^2} \right\},$$

where κ is a reproducing kernel. For the definition of reproducing kernel and RKHS, see [27] or [28].

Similar to the formulation (2), the optimal local regression function can be estimated by minimizing the regularized empirical error, that is

$$\hat{f}^t = \arg\min_{f \in \mathcal{H}_\kappa} \sum_i \left\| y_{t_i} - f(x_{t_i})_2 \right\|^2 + \lambda \|f\|^2_{\mathcal{H}_\kappa} \tag{4}$$

Considering the representation theory [29] or [30], the optimal local regression function can be represented as

$$\hat{f}^t = \sum_i \alpha_i^t \kappa(x_{t_i}, \cdot), \tag{5}$$

where $x_{t_i} \in S_L^t$. Combining the formula (4) and (5), there exists

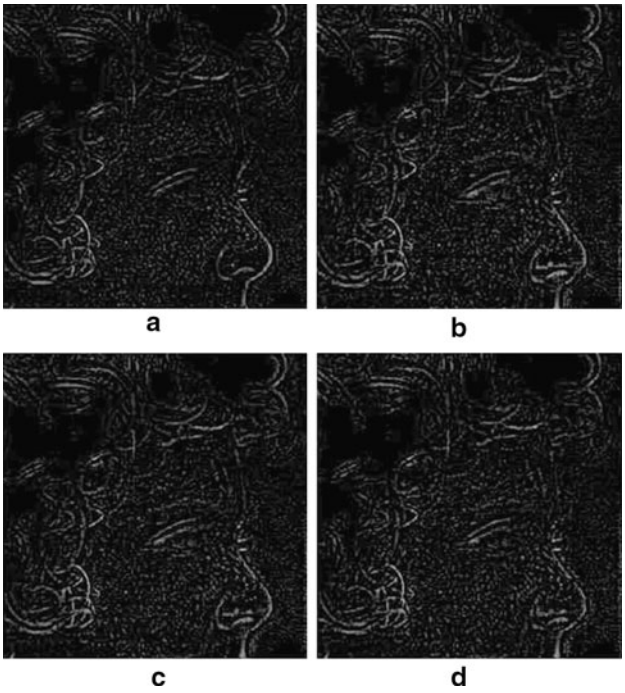

Fig. 4 Deviation between a super-resolved image and the original high-resolution image. **a** Bicubic interpolation;**b** nearest neighbor estimation method [14]; **c** nearest neighbor embedding method [15]; **d** nearest neighbor learning (Fig. 2)

$$\begin{aligned} \hat{A}^t &= \arg\min_{A \in \mathcal{R}^{h^2 \times t_k}} \|Y - AK\|_2^2 + \lambda \|A\|_2^2 \\ &= YK'(KK' + \lambda I)^{-1} \end{aligned} \tag{6}$$

where $\hat{A}^t = (\hat{\alpha}_1^t, \hat{\alpha}_2^t, \ldots, \hat{\alpha}_{t_k}^t)$, $K = \left(\kappa(x_{t_i}, x_{t_j})\right)$.

Combining the formulas (5) and (6), $\hat{f}^t$ can be defined as

$$\hat{f}^t(x) = \hat{A}^t (\kappa(x_{t_i}, x))_{i=1}^{t_k}. \tag{7}$$

It should be noted that the complexity of calculating the regression function $\hat{f}^t$ is related with the number of the local training samples. Too large size of local training set will make it trouble in calculating a large kernel matrix. Thus, it is time to consider the strategy of generating local training samples. Firstly, the k-nearest neighbors $\mathcal{N} = \{x_{t_1}^t, x_{t_2}^t, \ldots, x_{t_k}^t\}$ of ith test patch x_i^{test} are collected. Then the t_jth local training sample is replaced by a set of its k-nearest neighbors $\mathcal{N}_{t_j} = \{x_{n_s}^{t_j} | s = 1, 2, \ldots, k\}$. Finally, the local training set is generated by pooling all these k-nearest neighbor sets together, that is,

$$S_L^i = \left(\cup_{j=1}^k \mathcal{N}_{t_j} \right) \cup \mathcal{N}. \tag{8}$$

Based on such strategy, more information of training samples is used to estimated the high-resolution patches meanwhile the local geometrical structure is still not too complexity.

The local learning algorithm is summarized in Fig. 5.

5 Experiments

To validate the effectiveness of the local learning algorithm in single-image super-resolution, we compare such algorithm with other algorithms including bicubic interpolation algorithm, nearest neighbor-based estimation algorithm [14] and nearest neighbor embedding algorithm [15].

All training high-resolution images are shown on Fig. 6. The corresponding low-resolution images are generating by down-sampling and averaging. For all experiments, the magnification factor is set to be 4. The size of low-resolution patch is set to be 3×3 and the number of overlapping pixels is 2. The number of nearest neighbors is set to be various values including 5, 10, 50 and 100. A Gaussian kernel

$$\kappa(x, y) = \exp\{-\sigma\|x - y\|^2\} \tag{9}$$

is used in all experiments. The kernel parameter σ is set to be 10^{-1} in all experiments, which is chosen by cross-validation. The regularized parameter λ is empirically set to be 10^{-3}. In fact, no remarkable difference in performance there is when λ is around 10^{-3}. However, the performance will be decrease if λ is too large or too small, such as $\lambda > 10^{-2}$ or $\lambda < 10^{-6}$.

Before super-resolving a test image, some preprocessing is needed. Consider that humans are more sensitive to the changes in luminance than to the changes in color. The YIQ color model is used in experiments. Thus all images used in experiments should be converted into the YIQ model. The super-resolution is only made for luminance values from the Y channel. The features of images are generated by uniting the first-order and second-order gradients of the luminance.

Figure 9 shows the results of applying different super-resolution algorithms to a lightning image when different numbers of nearest neighbors are considered. Figure 7

Input
a set of training image pairs $\{(\mathcal{X}^i, \mathcal{Y}^i)_{i=1}^m\}$, a test image $\mathcal{X}^{\text{test}}$, the size of low-resolution patch $l \times l$, the magnification factor N, the number of nearest neighbors k and a kernel function κ
Step1 Generating the training set S and the test set T.
Step2 Nearest neighbor learning
For i=1 to $|T|$
a Generating i-th local training set corresponding to the i-th test patch by uniting all nearest neighbor sets
b Generating i-th local regression function f^i with kernel ridge regression
c Generating i-th high-resolution estimation $\hat{y}_i^{\text{test}} = f^i(x_i^{\text{test}})$
End
Output
Merge all $\hat{y}_i^{\text{test}}$ to obtain high-resolution estimation $\hat{\mathcal{Y}}$

Fig. 5 A local learning-based algorithm for single-image super-resolution

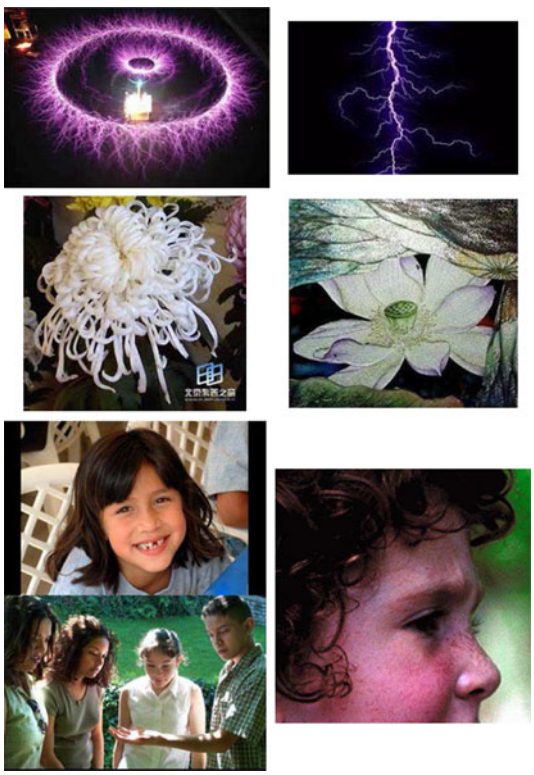

Fig. 6 All high-resolution images used in experiments. Both images on common row are used in an experiment. *Left images* are used to generate training samples, the *right* used to generate test samples

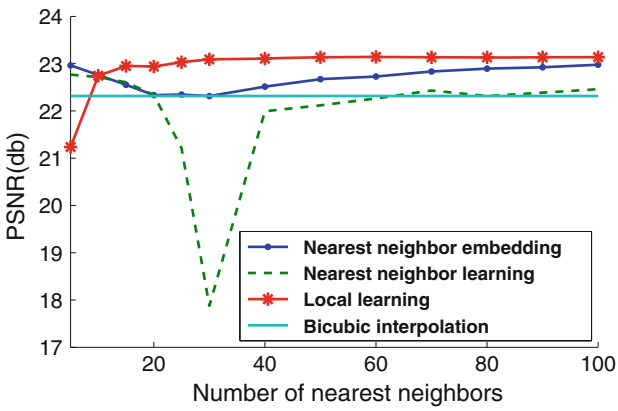

Fig. 7 Lightning: PSNR versus number of nearest neighbors

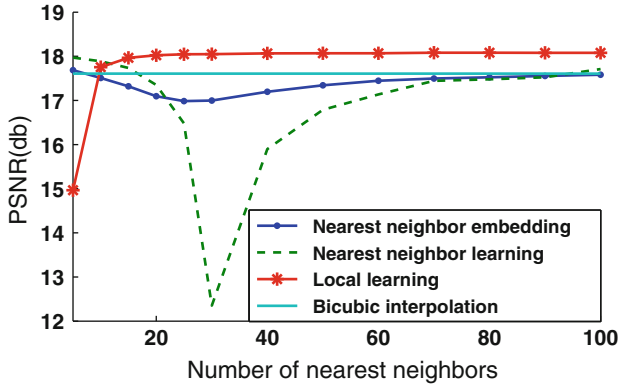

Fig. 8 Flower: PSNR versus number of nearest neighbors

shows the curve of values of peak signal-to-noise ratio (PSNR) corresponding different super-resolution algorithms when different numbers of nearest neighbors are adopted. In the experiment, two images of lightning in Fig. 6 are used as the original images to generate training and test samples. The left one is used to generate training samples and the other one to generate test samples. Similarly, the results about flowers are shown in Figs. 8 and 10. Those experimental results show that more details are well estimated by the local learning-based algorithms than those estimated by most of nearest neighbor-based algorithms. In Figs. 7 and 8, PSNR is assessed between luminance components.

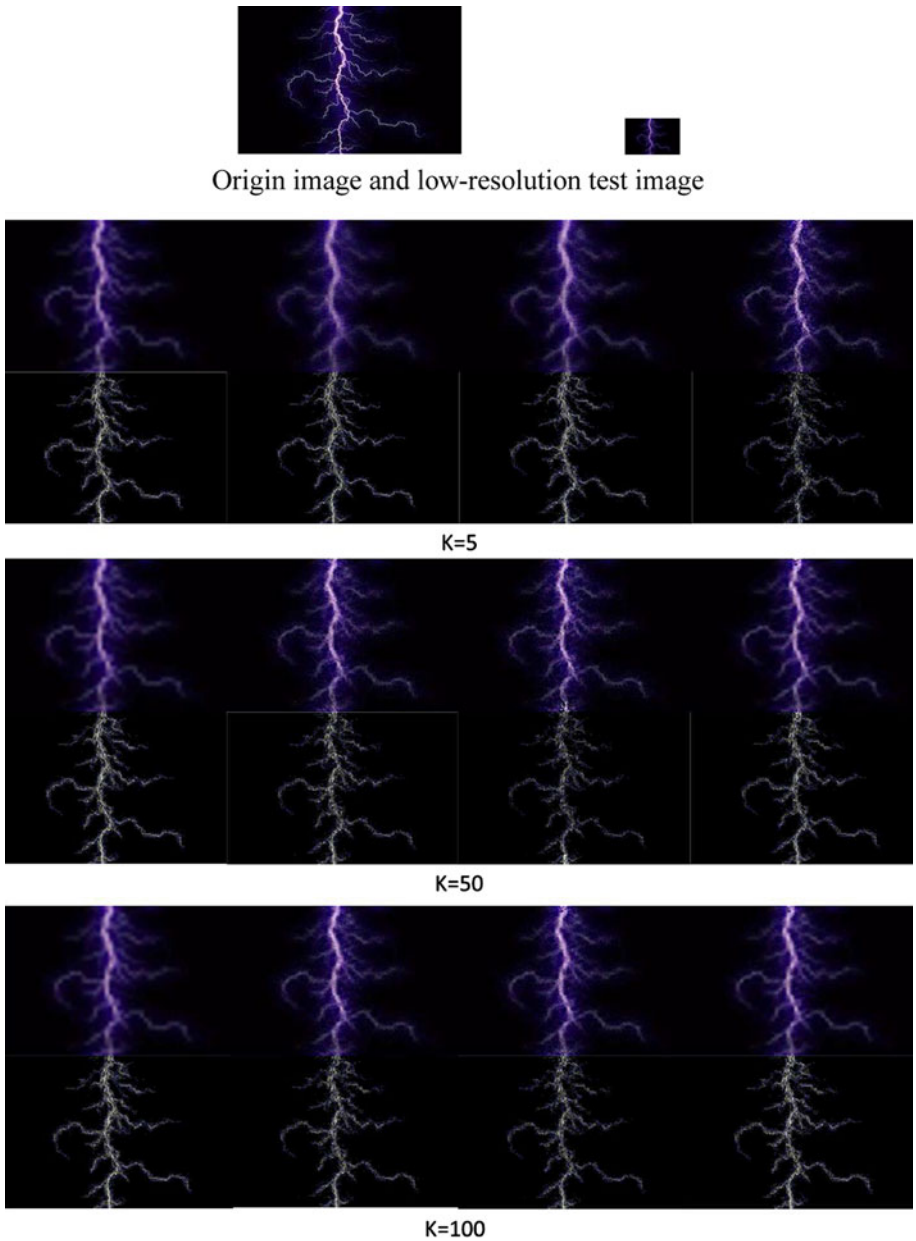

Fig. 9 Super-resolved images and deviation images for lightning, where K is the number of nearest neighbors. From *left to right*, the algorithms are bicubic interpolation, nearest neighbor embedding [15], nearest neighbor learning-based algorithm (Fig. 2) and local learning-based algorithm (Fig. 5)

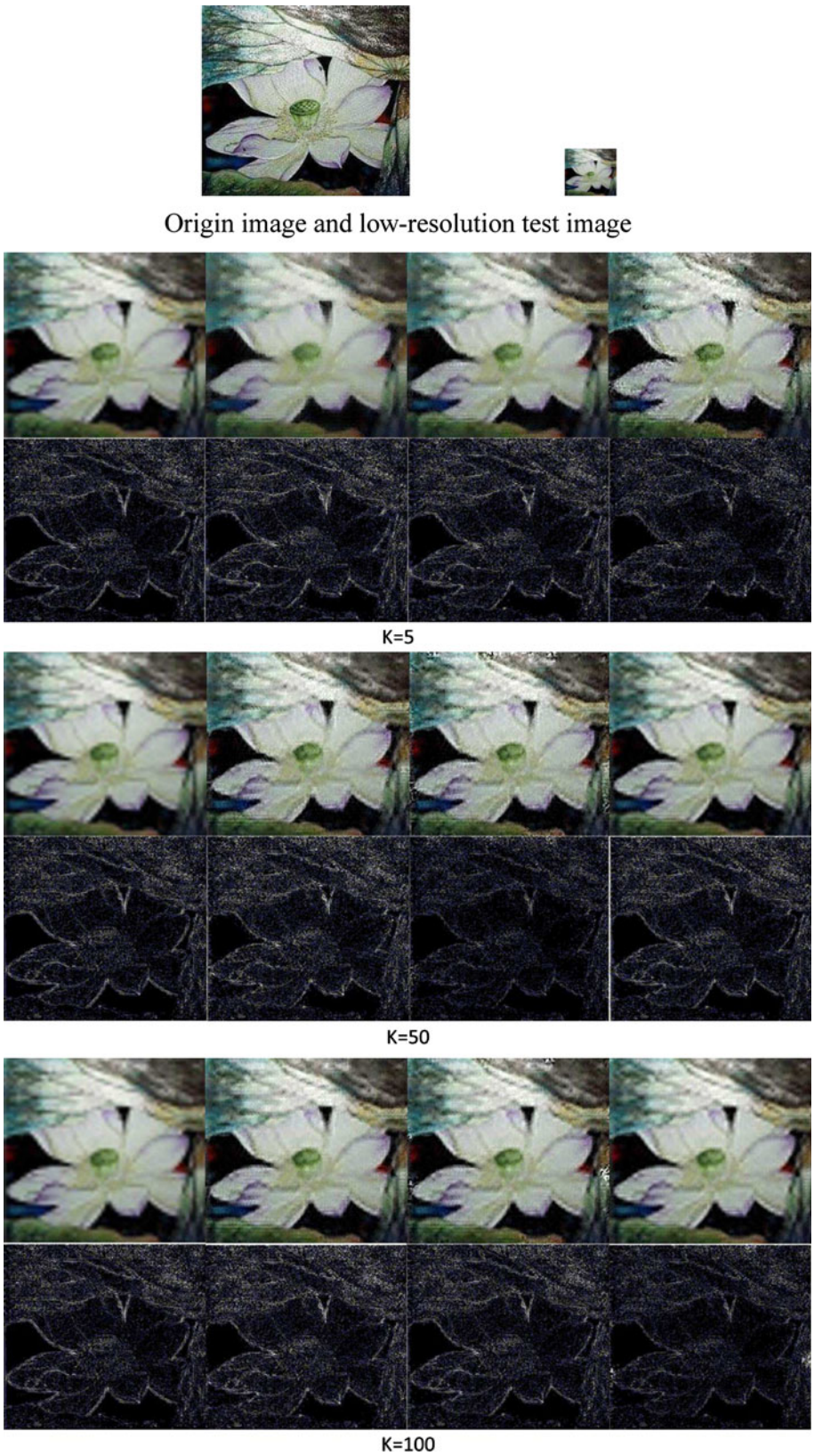

Fig. 10 Super-resolved images and deviation images for flower, where K is the number of nearest neighbors. From *left to right*, the algorithms are bicubic interpolation, nearest neighbor embedding [15], nearest neighbor learning-based algorithm (Fig. 2) and local learning-based algorithm (Fig. 5)

6 Conclusion

In this paper, a novel local learning-based method is introduced into the field of single-image super-resolution. The regression method is used on a local training set which is generated according to some similarity between low-resolution patches. The idea behind such method is making trade-off between generalization of estimators and the complexity of models, which is connected by the famous rule of structure empirical risk minimizing [27]. Experimental results show the effectiveness of the proposed algorithms, which also show the value of learning theory and machine learning methods in the single-image super-resolution.

Following the above idea, it is interesting to study the algorithms for super-resolution in the view of learning theory and machine learning. In the future, we will pay more attention on the theoretical analysis of example-based super-resolution methods with the help of learning theory.

Acknowledgments The work presented in this paper is supported by the National Basic Research Program of China (973 Program) (Grant No. 2011CB707000), the National Natural Science Foundation of China (Grant No. 61072093) and the Open Project Program of the State Key Lab of CAD & CG (Grant No. A1116), Zhejiang University.

References

1. Tsai R, Huang T (1984) Multi-frame image restoration and registration. Adv Comput Vis Image Process 1:317–339
2. Li X, Hu Y, Gao X, Tao D, Ning B (2010) A multi-frame image super-resolution method. Signal Process 90(2):405–414
3. Li X, Orchard MT (2001) New edge-directed interpolation. IEEE Trans Image Process 10(10):1521–1527
4. Unser M (1999) Splines: a perfect fit for signal and image processing. IEEE Signal Process Mag 16(6):22–38
5. Elad M (2002) On the origin of the bilateral filter and ways to improve it. IEEE Trans Image Process 11(10):1141–1151
6. Hardie R (2007) A fast image super-resolution algorithm using an adaptive wiener filter. IEEE Trans Image Process 16(12):2953–2964
7. Tschumperlé D, Deriche R (2005) Vector-valued image regularization with pdes: a common framework for different applications. IEEE Trans Pattern Anal Mach Intell 27(4):506–517
8. Dai S, Han M, Xu W, Gong Y (2007) Soft edge smoothness prior for alpha channel super resolution. In: Proc. IEEE ICIP'07, pp 1–8
9. Roweis S, Saul L (2000) Nonlinear dimensionality reduction by locally linear embedding. Sci Agric 290(5500):2323–2326
10. Scholkopf B, Smola A (2002) Learning with kernels. MIT Press, Cambridge
11. Kim K, Kwon Y (2010) Single-image super-resolution using sparse regression and natural image prior. IEEE Transa Pattern Anal Mach Intell 32(6):1127–1133
12. Yang J, Wright J, Huang T, Ma Y (2008) Image super-resolution as sparse representation of raw image patches. In: Proc. IEEE CVPR'08, June 2008, pp 1–8
13. Freeman W, Pasztor E, Carmichael O (1999) Learning low-level vision. In: The proceedings of the seventh IEEE international conference on computer vision, vol 2. pp 1182–1189
14. Freeman W, Jones T, Pasztor E (2002) Example-based super-resolution. IEEE Comput Graph Appl 22(2):56–65
15. Chang H, Yeung D, Xiong Y (2004) Super-resolution through neighbor embedding. In: Proc. IEEE CVPR'04, vol 1. pp 275–282
16. Ni K, Nquyen T (2007) Image superresolution using support vector regression. IEEE Trans Image Process 16(6):1596–1610
17. Pang Y, Yuan Y, Li X (2009) Iterative subspace analysis based on feature line distance. IEEE Trans Image Process 18(4):903–907
18. Bishop C, Blake A, Marthi B (2003) Super-resolution enhancement of video. In: Bishop CM, Frey B (eds) Proceedings of the ninth international workshop on artificial intelligence and statistics, Key West, FL, 3–6 January 2003
19. Pickup L, Robert S, Zisserman A (2003) A sampled texture prior for image super-resolution. In: Adv Neural Inf Process Syst 16, pp 1587–1594
20. Sun J, Tao H, Shum H (2003) Image hallucination with primal sketch priors. In: Proc IEEE CVPR'03, vol 2. pp II-729–II-736
21. Zhang T, Tao D, Li X, Yang J (2009) Patch alignment for dimensionality reduction. IEEE Trans Knowl Data Eng 21(9):1299–1313
22. Zhang K et al (2011) Partially supervised NE for example-based image super-resolution. IEEE J Sel Top Signal Process (to be published)
23. Li B, Chang H, Shan S, Chen X (2009) Locality preserving constraints for super-resolution with neighbor embedding. In: Proceedings of the IEEE ICIP'09, November 2009. pp 1189–1192
24. Zhou T, Tao D, Wu X (2011) Manifold elastic net: a unified framework for sparse dimension reduction. Data Mining Knowl Discov (to be published)
25. Smale S, Zhou D (2007) Learning theory estimates via integral operators and their approximation. Constr Approx 26:153–172
26. Shawe-Taylor J, Cristianini N. Kernel methods for pattern analysis. Cambridge University Press, Cambridge, July 2004
27. Vapnik V (1998) Statistical learning theory. Wiley, New York
28. Wu Q, Ying Y, Zhou D (2007) Multi-kernel regularized classifiers. J Complex 23(1):108–134
29. Cucker F, Zhou D (2007) Learning theory: an approximation theory viewpoint. Cambridge University Press, Cambridge
30. Smale S, Zhou D (2004) Shannon sampling and function reconstruction from point values. Bull Am Math Soc 41:279–305