Multi-stage Dictionary Learning for Image Super-Resolution Based on Sparse Representation

Dianbo Li, Wuzhen Shi, Wenfei Wang, Zhizong Wu and Lin Mei

Abstract Sparse representation has been proved successful in solving image super-resolution (SR) problems. It aims to compensate the high-frequency details from a pair of high–low (HL) resolution dictionary which is trained by the corresponding resolution of image patches. This paper presents a novel strategy to generate a super-resolution image via multi-stage HL dictionaries which are trained by a cascade training process. Extensive experiments on image super-resolution validate that the proposed solution can get much better results than some state-of-the-arts ones in terms of PSNR and FSIM.

Keywords Multi-stage dictionary learning · Image super-resolution · Sparse representation

1 Introduction

One single image SR problem has been a concerned issue in image processing for a long time. The goal is to recover the high resolution (HR) image from its low resolution (LR) form. However, it is an ill-posed inverse problem that some prior knowledge is in need to make the solution unique and stable. Lots of articles provide various methods to address this problem, which can be roughly divided into three categories, interpolating based, reconstructing based and learning based. Among them, the third one is more worth being researched than others in trend. For example, example learning based methods [1–4] employ a database consisting of co-occurrence examples from a training set of HR and LR image patches. Since they rely much more on the similarity between the training set and the test set,

D. Li (✉) · W. Wang · Z. Wu · L. Mei
The Third Research Institute of the Ministry of Public Security, Shanghai, China
e-mail: dianxinwu@126.com

W. Shi
School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China

N.Y. Yen and J.C. Hung (eds.), *Frontier Computing*, Lecture Notes in Electrical Engineering 422, DOI 10.1007/978-981-10-3187-8_10

they are not very practical in some situations. Another kind of efficient learning-based method [5–8] use the sparse-representation modeling to deal with this problem. Sparse-representation theory assumes that there is a linear relationship between high and low dimension, so that high dimension signal can be restored from their low dimension projection accurately [6]. Besides, [7] found that image patches can be well-represented as a sparse linear combination of elements from an appropriately chosen over-complete dictionary, so they made a compact representation for these patch pairs to capture the co-occurrence prior to improve the speed and the robustness significantly, achieving much better performance. Lately, [5] modified the approach above in various respects including computational complexity and algorithm architecture, which shows to be more efficient and much faster than [7]. Because of the limitation in recovering high-frequency details and the wide gap between the frequency spectrum of the corresponding HR image and that of the initial interpolation, [8] put forwards a dual-dictionary learning method via parse representation for image super-resolution, which consist of two steps to make up the wide gap. First, a main learned high-frequency dictionary was used to reduce the most gap of the frequency spectrum primarily. Then, a residual high-frequency dictionary was trained to recover the lack of residual high-frequency signal. According to [8], it obtained better results than [5] in PSNR.

However, the gap between the frequency spectrum of the corresponding HR image and that of the initial interpolation is so wide that two-layer progressive estimation of high frequency is not enough to recover the whole image high frequency details. In order to alleviate this problem, the multi-stage dictionary learning method is proposed. First, multiple stages of dictionary are trained offline, and each one also contains both high and low resolution parts. After that, high frequency details will be compensated by using these dictionaries via sparse representation stage by stage until the gap is smaller enough. This scheme can be treated as a cascade coarse-to-fine recovering progress, and the final results in the experimental section show that our method is better than expected.

This overall framework is as follows: some methods and research were introduced before in this section. In Sect. 2, the proposed SR scheme are described in detail including dictionary learning in Sect. 2.1 and image restoration in Sect. 2.2. Section 3 shows some experimental results in different views, and Sect. 4 makes some conclusions.

2 Method

When capturing image, it is easy to be affected by some factors such as deformation, blur, noise and down-scaling etc. Assuming that the original capturing image is an HR image, the actual obtained result is a LR image. This process can be described by formulation (1):

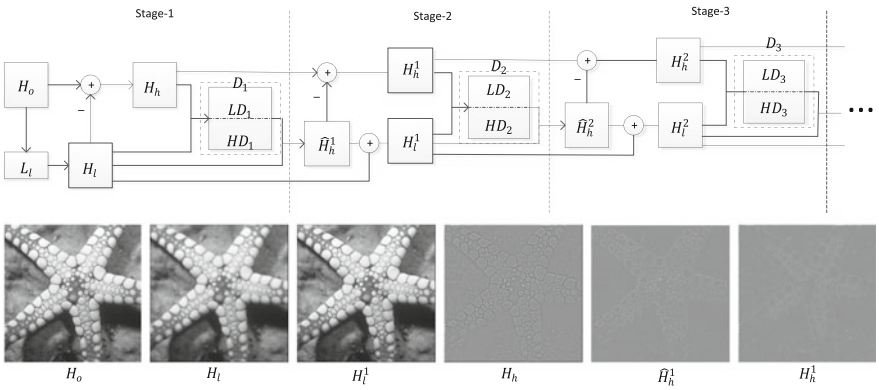

Fig. 1 The frame of dictionary learning stage

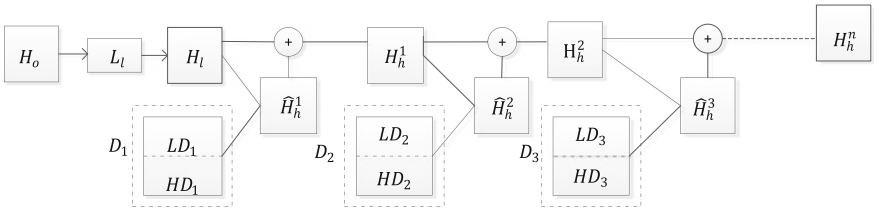

Fig. 2 Frame of image synthesis stage

$$y = GBDx + n \tag{1}$$

where x is the original HR image, y is the observed LR image. G denotes the geometric deformation operator, B denotes a blurring operator, D denotes a down-scaling operator and n is the additive Gaussian noise.

It can be seen that solving x is an ill-posed inverse problem. As a learning-based method, sparse representation method can get the coefficient between LR and HR image via a trained over-complete dictionary, which avoid to solving the equation directly. Both dictionary training and image generation are needed inescapability. We describe the training process as Fig. 1 and image generation progress in Fig. 2.

2.1 *Offline Dictionary Learning*

In this stage, multi-stage dictionaries are trained using sparse representation, i.e. D_1, D_2, D_3, …. Each dictionary like D_1 has two parts: low-frequency dictionary (LD_1) and high-frequency dictionary (HD_1), respectively. Our training scheme is similar in spirit to that of [7].

As shown in Fig. 1, H_l and H_h which represent HR low-frequency image and HR high-frequency image is the first pair input to train the first stage dictionary D_1, and some pre-progress to the defined original HR image H_o should have been done to get them before the true training stage. First, we down sampling H_o and get its blur image L_l. Then, applying bi-cubic interpolation method on L_l to construct the image H_l, which is of the same size as H_o. The final image H_h is generated by subtracting H_l from H_o.

Since we have said that each stage dictionary has two coupled sub-dictionaries (LD_1, HD_1), we need to extract the local patches from H_l and H_h to forming the training data $\left\{pa_l^n, pa_h^n\right\}$, where pa_h^n is the set of patches extracted from image H_h directly while pa_l^n is built in another way which has been explained in detail in [8].

In order to generate the dictionary LD_1 and HD_1, the following two Eqs. (2), (3) can be used to generate them. Formulation (2) is K-SVD dictionary learning [9] procedure and Formulation (3) is based on the theory of high-dimension image patches can be accurately recovered from their low-dimension projections.

$$\mathrm{LD}, \{q^n\} = \operatorname{argmin} \sum_n \left\| pa_l^n - LD \cdot q^n \right\|_2^2, \text{ s.t.} \|q^n\|_0 \le L, \forall n \tag{2}$$

where $\{q^n\}_n$ are sparse representation vectors, and $||\cdot||_0$ is the l_0 norm counting the nonzero entries of a vector.

$$\mathrm{HD} = \operatorname{argmin} \sum_n \left\| pa_h^n - HD \cdot q^n \right\|_2^2 = \operatorname{argmin} \sum_n \|P_h - HD \cdot Q\|_2^2 \tag{3}$$

where the matrices $P_h = \left\{pa_h^n\right\}_n$ and $Q = \{q^n\}_n$, respectively.

So far, the first stage dictionary D_1 has been trained, and we need set a stage number n to train more stage dictionaries. The next stages of dictionary can be built by using the same method of dictionary learning as D_1. As the input training image of the next stage, H_l^1 is generated by adding H_l and $\widehat{H}_h^1$, which contains more details $\left(\widehat{H}_h^1\right)$ than H_l. It is important to note that other stage of dictionary D_i is also consist of two coupled sub-dictionaries: low-frequency residual dictionary (LD_i) and high-frequency residual dictionary (HD_i).

Finally, all the rest stages of dictionary are trained as the same way described above. Theoretically, the back stage of dictionary contains less high-frequency signal than the previous stage and the dimension of the dictionary is higher, and at some point, the dictionary may has little use to compensate the high frequency signal.

2.2 Online Image Generation

After the offline training stage, there are multiple stages of dictionaries were generated. Each stage of dictionary can be used to compensate some high frequency

component for the low resolution image. More high frequency details can be got via a cascade compensating strategy in theory. However, too much compensation is not necessary, and even cause a distortion. Generally speaking, the problem of how many stages of dictionary should we use for generating the final HR image is hard to be determined, because we have not the strict evaluation standard to estimate the result. In this paper, we select the PSNR value as an indicator. When the PSNR value decline or stay the same, we stop the next image synthesis stage.

As shown in Fig. 2, H_h^n is the final synthetic image. $H_h^1, H_h^2, \ldots, H_h^i$ are the intermediate synthetic image after each stage of dictionary representation, which is also used to the next stage input. $\widehat{H}_h^1, \widehat{H}_h^2, \ldots, \widehat{H}_h^i$ are the lost high frequency of each input LR image.

Each stage of image synthesis has the same procedure to restore the loss. For the first stage example, suppose that an input LR image denoted by L_l has been done the same pre-progress in Sect. 2.1 to an HR image. Then, H_l is the first input target of the super resolution. With the use of dictionary D_1 and the method in [5], the first stage high-frequency image is generated $\widehat{H}_h^1$, which is just contain the lost high frequency signal, and add the input LR image H_l.

First, make sure that H_l is filtered with the same high-pass filters and PCA projection as the training stage, and then is decomposed into overlapped patches $\{pa_l^n\}_n$. After all, employ the traditional OMP method [8] to generate $\{pa_l^n\}_n$, and calculate the sparse representation vectors $\{q^n\}_n$ by allocating L atoms to their representation under LD_1. Next, the HR image patches can be reconstructed by the formulation: $\{\widehat{pa}_h^n\}_n = \{HD_1 \cdot q^n\}_n$. Finally, generate the first high frequency loss $\widehat{H}_h^1$ by solving the following minimization problem (4):

$$\widehat{H}_h^1 = \text{argmin} \sum_n \left\| R_n \widehat{H}_h^1 - \widehat{pa}_h^n \right\|_2^2 \tag{4}$$

More details of the solution can be referred to [8]. Then, the first HR temporary image H_{LF}^1 containing more details than H_{LF} is built by adding H_l to $\widehat{H}_h^1$.

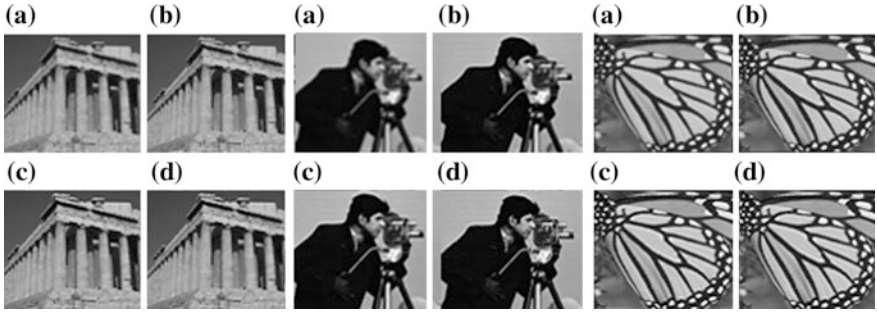

Fig. 3 Some vision comparison by different methods: **a** Bicubic interpolation; **b** J Zhang et al. [8]; **c** our method; **d** original images

In the same way, H_h^2 can be generated by using of H_l^1 and D_2, then, $H_l^3, \ldots, H_l^i$ and so on until reach the certain stopped condition. The last synthesized HR high-frequency image H_l^n contains much more details than the original HR high-frequency H_l. The stopped condition has been explained before in this section. Some synthesized image result is shown in Fig. 3.

3 Experiments

Extensive experiments on image super-resolution by using our method are demonstrated in this section. Bi-cubic interpolation method is a kind of complex interpolation method which is the best method of super resolution based on the interpolation method. It is comparable with sparse representation on the comprehensive performance, as a result, we employ it as a basic correlation method used in this paper. Besides, we take the comparison with the similar method in [8] to illustrate the advantages of our method.

First, we trained 9 stages dictionary as an offline library for the image synthesis step in Sect. 2.2. In order to test our performance with the methods bi-cubic interpolation method and dual-dictionary learning method [8], we take the same parameters as the method [8] including the Gaussian filter size and standard deviation of blurring operator which are set to 5×5 and 1 respectively, down sampling scale factor of decimation operator which is set to 2, and also the size of each level dictionary which is set to 500. Besides, the number of atoms for representing each image patch is fixed to 3, and the size of image patch is 9×9 with overlap of 1 pixel.

Some experimental results are shown in Fig. 4, which separately show the result of PSNR and FSIM with different stages of dictionary to be used in the image synthesis step. Each curve represents a test image, and each point in curve is an evaluation result corresponding to the stage in axis X. From the figures, we can see that in the front several stages, PSNR and FSIM increased significantly, and then remain the same or stay a little shock. In which, PSNR is the most widely used evaluation quality objective measurement and FSIM indicates the similarity of the original image and the interpolated high frequency image which is ranged from 0 to 1. Both of them are the bigger the better in their ranges.

To show the performance of the proposed method intuitively, we draw Table 1 as the compare result between different methods with the evaluation index PSNR. It can be seen that the proposed method can gain much better results of PSNR than the methods mentioned above, which increased 3.45 dB and 0.48 dB, respectively. The last column means how much the proposed method gain over Zhang's method [8], in which it claimed that his approach is better than the state-of-art method [5]. In conclusion, our method is effective in any way.

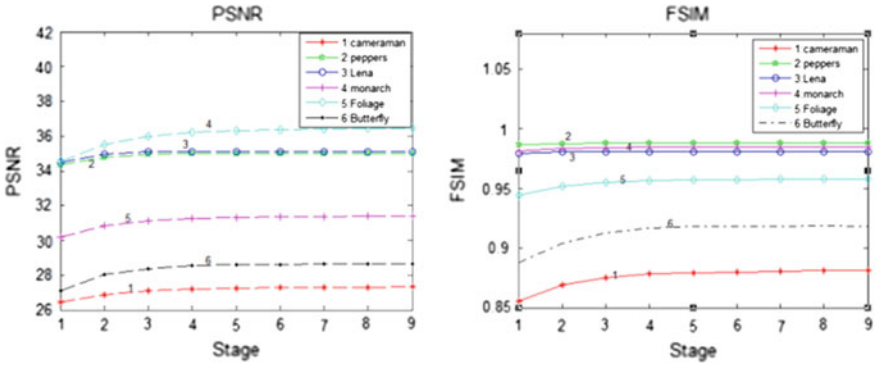

Fig. 4 PSNR and FSIM results on different test images

Table 1 PSNR comparisons with different algorithms (dB)

Images	Bicubic	J. Zhang [8]	Proposed	Gain
Cameraman	24.97	26.88	**27.31**	**0.43**
Foliage	31.65	35.50	**36.45**	**0.95**
Monarch	27.78	30.88	**31.39**	**0.51**
Peppers	32.32	34.78	**35.01**	**0.23**
Lena	32.19	34.96	**35.09**	**0.13**
Butterfly	24.23	28.01	**28.64**	**0.53**
Average	28.86	31.83	**32.31**	**0.48**

4 Conclusions

This paper presents a novel image super-resolution approach via multi-stages dictionaries learning based on sparse representation, which can restore a high-resolution image from a low-resolution one by a series of progressive high-frequency compensation utilizing multi-stages dictionaries. Experimental results show that the proposed method is able to narrow the gap between the frequency spectrum of the corresponding HR image and that of the initial interpolation, hence achieving better results in terms of both PSNR and FSIM. However, our method may spend some time off because of too much compensation in high frequency. Next, we will do some work to improve it.

Acknowledgements This work was supported in part by the Canada NSERC Business Intelligence Network and by the University of Waterloo, in part by the National Science and Technology Major Project under Grant 2013ZX01033002-003, in part by the National High Technology Research and Development Program of China (863 Program) under Grant 2013AA014601, in part by the National Science Foundation of China under Grants 61300028, in part by the Project of the Ministry of Public Security under Grant 2014JSYJB009.

References

1. J. Sun, N. N. Zheng, H. Tao, and H. Shum, "Image hallucination with primal sketch priors," IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 729–736, 2003.
2. Z. Xiong, X. Sun, and F. Wu, "Image hallucination with feature enhancement," IEEE Conference on Computer Vision and Pattern Classification, vol. 1, pp. 2074–2081, 2009.
3. W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, "Learning low-level vision," International Journal of Computer Vision, vol. 40, no. 1, pp. 25–47, 2000.
4. H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," IEEE Conference on Computer Vision and Pattern Classification, vol. 1, pp. 275–282, 2004.
5. R. Zeyde, M. Elad, and M. Protter, "On Single Image Scale-Up using Sparse-Representations," Curves & Surfaces, Avignon France, June, 24–30, 2010.
6. D. L. Donoho, "Compressed sensing," IEEE Transactions on Information Theory, vol. 52, no. 4, pp. 1289–1306, 2006.
7. J. Yang, J. Wright, T. Huang, and Y. Ma, "Image superresolution via sparse representation," IEEE Trans. on Image Processing, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.
8. J. Zhang, C. Zhao, S.W. Ma, D.B. Zhao."Image Super-Resolution via Dual-Dictionary Learning And Sparse Representation". ISC*AS*, *page 1688–1691*. *IEEE*, (*2012*).
9. M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing over complete dictionaries for sparse representation," IEEE Trans. on Signal Processing, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.