SUPER-RESOLUTION VIA K-MEANS SPARSE CODING

YI TANG[1], QI WANG[2]

[1]Faculty of Mathematics and Computer Science, Yunnan University of Nationalities, Kunming 650500, China
[2]School of Electronic and Control Engineering, Chang' an University, Xi'an 710064, China
E-MAIL: yitang4math@ynni.edu.cn, crabwq@chd.edu.cn

Abstract:

Dictionary learning and sparse representation are efficient methods for single-image super-resolution. We propose a new approach to learn a set of dictionaries and then choose the suitable one for a given test image patch of low resolution. Firstly, the training image patches are clustered into K groups with the information of the test image patches. Secondly, a best basis is learned to model each cluster using sparse prior. Finally, we employ this dictionary to estimate the high resolution patch for the given low resolution patch. This method reduces the complexity of dictionary learning greatly and also makes the representation of patches more compact compared to state-of-the-art methods, which learn a universal dictionary. Experimental results show the effectiveness of our method.

Keywords:

Super-resolution; K-means; Dictionary learning; Sparse representation

1. Introduction

The purpose of image super-resolution (SR) is to recover a high resolution (HR) image from one or more low resolution (LR) images. Image super-resolution is applied in many applications, such as video surveillance and remote sensing because it is expensive to directly obtain high-resolution images.

Interpolations may be the simplest and most convenient methods for super-resolving the low-resolution images. Generally, the bilinear interpolation and bicubic interpolation are effective in super-resolving a low-resolution nature image [1]. However, some artifacts such as "jaggy" and "ringing" will make the enlarged images blurred. To refine the basic interpolation methods, additional image priori is introduced in interpolation based methods. For example, local covariance coefficients regularization term was used to enhancing the interpolation process in [2]. Dai et al. [3] proposed soft-cuts algorithm to preserve the smoothness of edges in gradual variation region. Though the additional image priori is helpful in enhancing the performance of interpolation based methods, many nature images could not be well super-resolved because of the limits of the additional image priori. For example, interpolation based algorithms may be ineffective when LR images contain fine textures [3].

Noticing the shortage of interpolation based methods, Tsai and Huang [4] introduced multi-frame super-resolution where an HR image can be generated from a set of LR images. Because the information used in reconstructing an HR image is directly from the related LR images, the performance of multi-frame super-resolution is mostly better than the ones of the interpolation based methods. Unfortunately, an HR image will be connected to many different LR images, which means the optimization problem corresponding to the multi-frame super-resolution is an ill-posed problem [5]. Additionally, larger magnification is impossible because the inherent framework of multi-frame super-resolution [6].

Learning based super-resolution is thought as a promising method for refining the traditional super-resolution algorithms [6]. The first learning-based super-resolution algorithm was introduced by Freeman et al. in [7]. The relation between LR and HR images is modeled by Markov Random Field (MRF) in their algorithm. Experimental results show the effectiveness of their algorithm in super-resolving nature images with larger magnification. Motivated by the success of their algorithm, more learning based algorithms are proposed with different motivations. For example, Chang et al. [8] integrated manifold learning into the learning based super-resolution algorithms, and dictionary learning techniques are used in [9].

Yang et al. [9] proposed single image SR method based representation can be shown by the excellent results. It is shown in [9] that their SR method is robust to noise while other methods like neighbor embedding [8] fail to separate noise from real images and get noisy result. In [9], authors also shows that the visual effects of the reconstructed images are almost the same and outperform other methods, such as bicubic interpolation and neighbor embedding when the size of dictionary ranges from 256 to 2048. It means LR images can be well recovered by using small databases of HR and LR atom pairs rather than using large databases composed of

HR and LR raw image pairs. Thus the computation complexity and the storage requirement are greatly reduced.

The effectiveness of Yang's method largely relies on whether image patches can be sparsely represented by the atoms learned. However, the character of images always changes even in one single image. Then a question arises: Can a dictionary learned from a certain type of image patches represent the image structures more effectively than that learned from general image patches? If the answer is yes, we can learn a set of dictionaries for different types of images. For a given input LR image, we can pick up the most suitable dictionary and apply it in super-resolution progress. The above idea was employed by Dong et al. to learn a series of dictionaries in [11]. Firstly, the set of high resolution training image patches is partitioned into K clusters by using conventional K-means. Secondly, for each cluster, dictionary training is performed independently to learn K sub-dictionaries. If an input test image patch is classified as class i, the most suitable sub-dictionary is the one learned from that class.

Although the method illustrated in [11] can adaptively select the most suitable dictionary for each test image patch，the cluster progress in [11] doesn't take the information of test images into account. Therefore, the sub-dictionaries learned from the training set in [11] may be neither optimal nor effective for the test images.
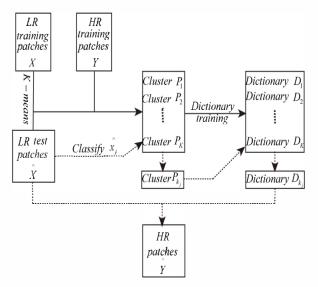

Figure 1.   An illustration of our approach

In this paper, we propose a novel learning-based approach in which a set of compact sub-dictionaries are adaptively learned from example image patches. By incorporating the test image into the training images, the dictionary is adaptively learned which is suitable for every

input image in this paper. In this case, we can achieves much better results than [9] in terms of both PSNR and visual perception.

As Fig. 1 illustrates, firstly, we divide the training images into K clusters or K sub-training sets with the information of the LR test image. Our cluster method is different from the conventional K-means which is adopted in [11]. Because the prior information of test image is incorporated in cluster progress, training images and test image of the same cluster are more similar than [11]. Hence, the representation of test images can be more compact. Secondly, a set of dictionaries are learned using sparse prior for each cluster. Finally, the input LR test image patch $\hat{x}_j$ in cluster P_{k_j} will be represented by LR dictionary $D_{k_j}^h$. The HR test image can be recovered in the same way of [9].

The rest of the paper is organized as follows. In section 2, we will discuss the details of our approach. Section 3 demonstrates the effectiveness of our adaptive dictionary selection by experimental results. Finally, section 4 concludes this paper.

2.   Super-resolution via K-means sparse coding

Yang et al. proposed an image SR via sparse coding [9]. The key idea of sparse coding is that natural signals can be sparsely represented as a linear combination of atom signals. SR problem using sparse prior   states that each pair of LR and HR image patches have the same sparse representations with respect to the two dictionaries D_l and D_h. Based on the sparse prior assumption, cluster method is considered. Because signals in the same cluster have similar feature information, each signal can be represented by fewer atom signals in the same cluster. Hence signals in the same cluster have a sparse representation. In this paper, we employ K-means method, which is a popular unsupervised learning algorithm that can effectively solve the clustering problem [12]. Among most of the super-resolution algorithms based on cluster method, only training image is used by cluster algorithms and the information of test images is usually ignored. Besides, K-means method is also significantly sensitive to the initial centroids. Hence, SR algorithms based on cluster method may have bad effect on the recovered image. It is interesting to make use of the prior information of test image which can improve the rationality of cluster. We will give a brief formulation of K-means-sparse coding for SR as follows.

Let $X = \{x_1, x_2, \ldots, x_m\}$ be a set of LR training patches, $Y = \{y_1, y_2, \ldots, y_m\}$ be the set of corresponding HR training

patches, $\hat{X} = \{\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_m\}$ be a set of LR test patches, and $\hat{Y} = \{\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_m\}$ be the set of corresponding HR patches to be estimated. K-means-sparse coding for super-resolution can be divided into five steps:

Firstly, we combine X and $\hat{X}$ together in one set $\tilde{X}$. K-means method is employed to divide the test patches $\hat{X}$ into k clusters, and the centroids of k test clusters are $c_1, c_2, \ldots, c_k$. We set the initial centroids of $\tilde{X}$ equal to the centroids of test cluster $c_1, c_2, \ldots, c_k$ at the initial step. Then the set $\tilde{X}$ is classified into k clusters $X_1, X_2, \ldots, X_k$ by employing K-means in the following step. The centroids of $X_1, X_2, \ldots, X_k$ are $\overline{c}_1, \overline{c}_2, \ldots, \overline{c}_k$. The LR training patches can be picked up from X_i and construct a new cluster $\overline{X}_i$. With regard to each cluster $\overline{X}_i$, $\overline{Y}_i$ of Y is the corresponding HR image patches cluster. So we get a series of training patch-pairs subsets $P_i = \{\overline{X}_i, \overline{Y}_i\}, i = 1, 2, \ldots, k$.

Secondly, corresponding to each subset P_i, joint dictionary training method [9] is employed to learn sub-dictionaries D_i^l and D_i^h for low and high resolution image patches respectively.

Thirdly, given the input LR image patch $\hat{x}_j$, we select the sub-dictionary best fitted to $\hat{x}_j$ by comparing $\hat{x}_j$ to the centroid $\overline{c}_k$. For example, we can select the sub-dictionary for $\hat{x}_j$ based on the minimum Euclidean distance between $\hat{x}_j$ and $\overline{c}_k$, i.e.,

$$k_j = \arg \min_k \|\hat{x}_j - \overline{c}_k\|_2 \tag{1}$$

Fourthly, each LR test patch $\hat{x}_j$ in $\hat{X}$ can be faithfully represented by a sub-dictionary $D_{k_j}^l$. The representation coefficient vector $\tilde{\alpha}$ corresponding to the sub-dictionary $D_{k_j}^l$ can be intuitively formulated by the following objective function:

$$\tilde{\alpha} = \arg \min_\alpha \|D_{k_j}^l \alpha - \hat{x}_j\|_2^2 + \lambda \|\alpha\|_1 \tag{2}$$

The HR patch can be recovered as $\hat{y}_j = D_{k_j}^h \tilde{\alpha}$.

Finally, all HR patches are merged together to obtain HR image Y_t.

The pseudo-codes are listed as follow:

Algorithm: Super-resolution via K-means-sparse coding
Task: Estimate HR image Y_t of a LR test image X_t.
Input:

- A LR test image X_t.
- Training HR images Y_s and its LR version X_s.
- Size of LR patches q×q.
- Zooming factor s.
- Dictionary size n.
- Cluster number k.

Output: HR image Y_t.
Procedure:

1) Cut X_s and X_t into patches of size q×q with one or two pixels overlap. Generate training and test patches set as $X = \{x_1, x_2, \ldots, x_m\}$ and $\hat{X} = \{\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_m\}$ respectively.

2) Cut Y_s into patches of size sq×sq with s or 2s pixels overlap accordingly, i.e., $Y = \{y_1, y_2, \ldots, y_m\}$. Generate low-high training patch-pairs $\{X, Y\}$.

3) Employ K-means to divide $\hat{X}$ into K clusters, the centroids of these clusters are $c_1, c_2, \ldots, c_k$.

4) Combine X and $\hat{X}$ together into one set $\tilde{X}$. K-means is employed once again to divide $\tilde{X}$ into K clusters $X_1, X_2, \ldots, X_k$ with $c_1, c_2, \ldots, c_k$ as the initial centroids. The centroids of these new clusters are $\overline{c}_1, \overline{c}_2, \ldots, \overline{c}_k$.

5) Select LR training patches from X_i, and use these patches to form a new set $\overline{X}_i$. $\overline{Y}_i$ is the corresponding HR patch cluster. Composing $\overline{X}_i$ and $\overline{Y}_i$ together, we get a training patch pairs subset $P_i = \{\overline{X}_i, \overline{Y}_i\}, i = 1, 2, \ldots, k$.

6) Sub-dictionary for each $P_i = \{\overline{X}_i, \overline{Y}_i\}$ is learned by joint dictionary training [9]. D_i^l and D_i^h are dictionaries for LR and HR patches respectively.

7) For each patch $\hat{x}_j$ in $\hat{X}$, select the best fitted dictionary for $\hat{x}_j$ based on the minimum distance between $\hat{x}_j$ and $\overline{c}_i$, i.e.,

$$k_j = \arg \min_k \|\hat{x}_j - \overline{c}_k\|_2$$

8) Solve the optimization problem

$$\tilde{\alpha} = \arg \min_\alpha \|D_{k_j}^l \alpha - \hat{x}_j\|_2^2 + \lambda \|\alpha\|_1$$

9) Generate the HR patch

$$\hat{y}_j = D_{k_j}^h \tilde{\alpha}$$

10) Merge all $\hat{y}_j$s together to obtain a high resolution image Y_t.

3 Experimental results

3.1. Image feature representation

During K-means-sparse coding, the training set contains feature vectors rather than original image patches. The feature vectors of LR and HR patches are the same with the methods of Chang e al. [8] and Yang et al. [9].

To verify the effectiveness of the proposed image super-resolution reconstruction algorithm, we compare our results with those of Bicubic interpolation and Yang's method in [9].

3.2. Training datasets

Although image contents can vary a lot from image to image, it can be represented by a small number of structural elements, e.g., edges, line segments and other elementary features. Therefore, we select images which are rich in edges and textures to train dictionaries. These dictionaries can represent the natural images well. Training images are shown in Fig. 2.

Figure 2. Training images.

3.3. Comparison

In the experiments of super-resolution, we zoom the input LR image patches up by a factor of 3. We mainly compare our method with Yang's method in [9]. Two algorithms use the same features to represent LR and HR patches. The size of LR patches is 3×3, with one pixel overlapped between adjacent patches, and the corresponding 9×9 HR image patches with three pixels overlapped. Similar to Yang's method, we also extract gradient features from the upsampled version (6×6) of the LR image. 40000 LR and HR patch pairs are randomly extracted as the training set for sparse dictionary. The number of clusters is 5 in K-means method. The size of sparse dictionary is fixed as 512, which depends on the tradeoff between the image quality and the robustness to noise. We set λ =0.1 in this experiment. Fig. 3 shows SR results of our method compared to Bicubic and Yang's method. Among these three super-resolution algorithms, our method shows the best visual effect and achieves the highest PSNR value. More experimental results

of PSNR value are shown in Table 1. All of these experimental results of PSNR value shows that our performance is always better than the other algorithms.
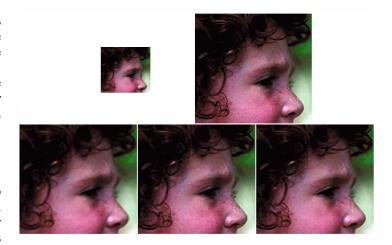

Figure 3. Results of the girl with zooming factor 3 the corresponding PSNRs. Upper Left: low-resolution image; Upper Right: high-resolution image; Lower Left: Bicubic interpolation image (PSNR: 28.4848dB); Lower Middle: Yang's super-resolution image (PSNR: 28.4792dB); Lower Right: Our super-resolution image (PSNR: 28.5673dB).

TABLE 1. PSNR VALUES OF SUPER-RESOLUTION IMAGES

Test Image	Super-Resolution Algorithms		
	Bicubic interpolation	Yang's algorithm	Proposed algorihtm
Test 1	31.7463	32.7666	32.9325
Test 2	25.3544	26.9930	27.0223
Test 3	27.4573	29.1355	29.5491
Test 4	27.3372	27.9439	28.2107
Test 5	26.3357	26.8426	26.9374
Test 6	29.3379	30.9573	31.0024

4. Conclusions

In this paper, we propose a novel super-resolution method based on *K*-means-sparse coding. In order to adaptively select the most suitable dictionary, *K*-means method is utilized. Our main contribution is combining test image information with conventional *K*-means method so as to make each cluster more reasonable. Experimental result shows our method is better than Yang's method both in visual perception and in PSNRs. However, there are some works for future improvement, such as how to select the number of cluster and reduce the empirical parameters.

285

Acknowledgements

References

[1] H. S. Hou and H. C. Andrews, "Cubic splines for image interpolation and digital filtering," IEEE Trans. Acoustics, Speech, Signal Process., vol. ASSP-26, no.6, pp. 508-517, Dec. 1978.

[2] X. Li and M. T. Orchard, "New edge-directed interpolation," IEEE Trans. Image Process., vol. 10, no. 10, pp. 1521-1527, Oct. 2001.

[3] S. Dai, M. Han, et al., "Soft-cuts: a soft edge smoothness prior for color image super-resolution," IEEE Trans. Image Process., vol. 18, no. 5, pp. 969-981, May. 2009.

[4] R. Tsai and T. Huang, "Multi-frame image restoration and registration," in Advances in Computer Vision and Image Processing, 1984, vol. 1, pp. 317–339.

[5] R. C. Hardie, K. J. Barnard, et al., "Joint MAP registration and HR image estimation using a sequence of undersampled images," IEEE Trans. Image Process., vol. 6, no. 12, pp. 1621-1633, Dec. 1997.

[6] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 9, pp. 1167-1183, Sep. 2002.

[7] W. T. Freeman, E. C. Pasztor, et al., "Learning low-level vision," International Journal of Computer Vision and, vol. 40, no. 1, pp. 25-47, 2000.

[8] H. Chang, D.-Y. Yeung, et al., "Super-resolution through neighbor embedding," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2004, vol. 01, pp. 275-282.

[9] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representation," IEEE Trans. Image Process., vol. 19, no. 11, pp. 2861-2873, Nov. 2010.

[10] I. Daubechies, M. Defrise, et al., "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," Commun. Pure Appl. Math., vol. 57, No. 11, pp. 1413-1457, Nov. 2004.

[11] W. Dong, L. Zhang, et al., "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," accepted in IEEE Trans. Image Processing.

[12] M. Aharon, M. Elad, and A.M. Bruckstein, "The K-SVD: An Algorithm for Designing of Overcomplete Dictionaries for Sparse Representation," IEEE Trans. On Signal Processing, Vol. 54, no. 11, pp. 4311-4322, Nov. 2006.