CONVEX DICTIONARY LEARNING FOR SINGLE IMAGE SUPER-RESOLUTION

Pak Lun Kevin Ding, Baoxin Li *

Arizona State University
School of Computing, Informatics,
Decision Systems Engineering
Tempe, AZ 85287, USA

Kan Chang †

Guangxi University
School of Computer and
Electronic Information
Nanning, Guangxi 530004, China

ABSTRACT

In recent years, dictionary learning approaches have been used in image super-resolution, achieving promising results. Such approaches train a dictionary from image patches and reconstruct a new patch by sparse combination of the atoms of the dictionary. Typical training methods do not constrain the dictionary atoms. In this paper, we propose a convex dictionary learning (CDL) algorithm by constraining the dictionary atoms to be formed by non-negative linear combination of the training data, which is a natural, desired property. We evaluate our approach by demonstrating its performance gain over typical approaches.

***Index Terms*—** Dictionary learning, sparse representation, super-resolution.

1. INTRODUCTION

High-resolution images are useful in many real world applications, such as medical imaging and satellite imaging. When only low-resolution images are available, super-resolution is a common technique deployed for resolution enhancement. The task of single-image super-resolution, which is the focus of this paper, is to reconstruct a high-resolution version from only a low-resolution image. Conventional techniques for solving this problem can be categorized into interpolation-based (such as bilinear interpolation)[1], reconstruction-based [2, 3], and example-based methods [4, 5]. Interpolation-based schemes tend to generate over-smoothed images, and thus complex details are often missing in the reconstructed high-resolution images. In the reconstruction-based methods, super-resolution is viewed as an inverse problem, which is typically ill-posed, and thus different regularization schemes have been introduced. Most existing efforts on this regard show that such approaches only work for small up-scaling factor. Example-based approaches use images in the training set as priors for reconstruction. Typically they require a huge dataset of high-resolution and low-resolution patch pairs as the examples.

In recent years, dictionary learning-based methods [6, 7, 8] have been proposed and promising results have been reported. There are also studies [9] on relations between such approaches and the theory of compressed sensing [10]. In a dictionary-learning-based approach, a compact dictionary is learned from some training image patches and future super-resolution would be based on only this compact dictionary. In a representative algorithm of this kind [6], a pair of dictionaries are learned such that both high-resolution and low-resolution image patches can be transformed into the same sparse representation. Other examples of such approaches include [11, 12, 13]. Most existing such efforts base the learning on the basic K-SVD algorithm [14] or its variants. K-SVD, as a generalized K-means algorithm, is mainly concerned with minimum-mean-squared-error reconstruction, but less concerned with whether the learned dictionary is physically meaningful (and thus potentially optimal in some sense). For example, while the learned dictionary atoms are supposedly some de facto ”basis” image patches, most existing approaches would allow such patches contain negative components and/or allow sparse codes under the dictionary to be negative.

In this paper, aiming at learning dictionaries whose atoms are closer to natural image patches as well as support non-negative combinations to form new image patches (both intuitive and desirable properties for a dictionary-based scheme), we propose a new algorithm for training a dictionary for super-resolution. The key idea and main contributions of this paper are thus:

- We introduce a convexity constraint to the dictionary, which forces the atoms of the dictionary to lie within the convex cone of the training patches, and thus formulate a new dictionary learning problem ”convex dictionary learning” (CDL).
- We propose an optimization algorithm for solving the CDL problem to obtain a solution, and demonstrate

*Efforts partially supported by an ONR grant. The views/conclusions are solely of the authors and do not necessarily reflect ONR’s opinion.

†Efforts partially supported by the Natural Science Foundation of China via Grant 61401108 and also by Natural Science Foundation of Guangxi via Grant 2016GXNSFAA380154.

with experiments the effectiveness of the proposed formulation and algorithm.

The rest of the paper is organized as follows. We review the basic idea of super-resolution and dictionary learning in Sect. 2. Sect. 3 describes our algorithm and the implementation details. Experimental results are given in Sect. 4, and conclusions are stated in Sect. 5.

2. SUPER-RESOLUTION, SPARSE CODING AND DICTIONARY LEARNING

The problem of super-resolution can be described as follows: let $x_h \in \mathbb{R}^h$ be the vector denoting the original high-resolution image, $x_l \in \mathbb{R}^l$ be the vector denoting the low-resolution version of the original image, $H : \mathbb{R}^h \rightarrow \mathbb{R}^h$ and $S : \mathbb{R}^h \rightarrow \mathbb{R}^l$ be the blur and decimation operators respectively, then the relation between x_h and x_l can be written as:

$$x_l = SHx_h \tag{1}$$

Given x_l, the problem of super-resolution is to find $\hat{x_h} \in \mathbb{R}^h$ such that $\hat{x_h} \approx x_h$. Note that the number of columns is larger than the number of rows for SH, and thus the super-resolution problem is ill-posed.

Let X be the input data, which contains M-dimensional N input signals, $i.e.$ $X = (x_1, ..., x_N) \in \mathbb{R}^{M \times N}$. Learning a reconstructive dictionary with K columns, is equivalent to the following optimization problem:

$$\begin{aligned} < D, Y > &= \arg \min_{D,Y} \|X - DY\|_F^2 \\ s.t. \quad & \|y_j\|_0 \leq T \quad \forall j \in \{1, ..., N\} \end{aligned} \tag{2}$$

where $D = (d_1, ..., d_n) \in \mathbb{R}^{M \times K}$ is a over-complete dictionary with $K > M$, $Y = (y_1, ...y_N) \in \mathbb{R}^{K \times N}$ is the sparse representation matrix of input data X, $\|y\|_0$ is the number of non-zero entries in y, T is the sparsity constraint constant, and $\|X - DY\|_F^2$ is the reconstruction error.

To construct D, we need to minimize the reconstruction error and fulfill the sparsity constraints in (2). The optimization problem in (2) can be split into the following two sub problems (3) and (4):

$$\min_Y \|X - DY\|_F^2 \quad s.t. \quad \forall j \in \{1, ..., N\}, \|y_j\|_0 \leq T \tag{3}$$

and

$$D = \arg \min_D \|X - DY\|_F^2 \tag{4}$$

The main idea of dictionary learning is to find a dictionary such that every input can be approximated as a linear combination of the columns of the dictionary. Several dictionary training methods have been proposed, such as method of optimal directions(MOD) [15] and KSVD algorithm [14], the latter is widely used in many real world applications.

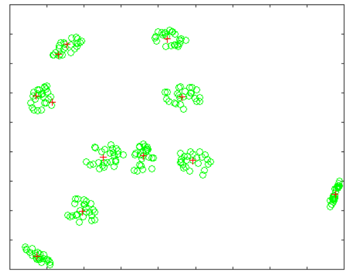

Fig. 1. A synthetic data sets with several clusters. The red"+" are the columns of the dictionary constructed by CDL.

For dictionary-based super-resolution, Yang $etal.$ [7] proposed a coupled dictionary training model:

$$\begin{aligned} \min_{D^H, D^L, y_j} \sum_j \|x_j^H - D^H y_j\|_2^2 + \|x_j^L - D^L y_j\|_2^2 \\ s.t. \quad \|y_j\|_0 \leq T \quad \forall j \in \{1, ..., N\} \end{aligned} \tag{5}$$

where x_j^H and x_j^L are the patches extracted from the high-resolution image and its corresponding feature vector in the low-resolution image at the same location respectively. y_j is the sparse representation for the jth pair of low-resolution and high-resolution patches. The core idea of this approach is, the sparse coefficient of the high-resolution patch, after the process described in Eq. (1), should remain unchanged.

3. CONVEX DICTIONARY LEARNING

We now present the details of our proposed model, which is centered at imposing convexity to dictionary learning.

3.1. Problem Formulation

A baseline dictionary-learning approach like Eq. (2) does not constrain the columns of the dictionary $D = (d_1, ..., d_K)$. We impose a convex constraint on the dictionary such that the column vectors of D lie in the convex cone formed by the column vectors of the input data matrix X, that is:

$$d_j = Xa_j, \quad \|a_j\|_1 = 1, \quad a_j \geq 0 \quad \forall j \in \{1, ..., K\} \tag{6}$$

where $\|\cdot\|_1$ denotes the l_1 norm. We say that d_j is a convex combination of $x_1, ..., x_N$, if there exists $c_i \geq 0, \forall i \in \{1, ..., N\}$ such that $d_j = c_1 x_1 + ... + c_N x_N$ and $\sum_i c_i = 1$. In this case, d_j lies in the convex cone which is formed by some x_i, where c_1, ..., c_N is the barycentric coordinates.

After imposing this constraint, the objective function for dictionary construction can be defined as:

$$\begin{aligned} \min_{D,Y} \|X - DY\|_F^2 \quad s.t. \quad \forall j \in \{1, ...N\} \\ D = XA, A \geq 0, \|a_j\|_1 = 1, \|y_j\|_0 \leq T \end{aligned} \tag{7}$$

	baby	bird	butterfly	head	woman	barbara	flowers	lenna	pepper	zebra
Bicubic	33.94	32.48	24.06	32.90	28..56	26.20	27.32	31.73	30.76	26.69
	0.9589	0.9753	0.9070	0.8897	0.9485	0.8109	0.8832	0.9730	0.9726	0.9249
KSVD	34.61	33.54	25.34	33.27	29.52	26.46	28.09	32.47	31.34	27.88
	0.9618	0.9782	0.9271	0.8938	0.9553	0.8191	0.8938	0.9753	0.9744	0.9406
CDL	**34.98**	**33.78**	**25.61**	**33.51**	**29.87**	**26.72**	**28.40**	**32.82**	**31.47**	**28.33**
	0.9649	**0.9786**	**0.9291**	**0.8981**	**0.9583**	**0.8300**	**0.8993**	**0.9768**	**0.9748**	**0.9460**

Table 1. The PSNR(dB) and SSIM comparsion for bicubic, KSVD, and CDL

where $A = (a_1, ..., a_K) \in \mathbb{R}_+^{N\times K}$ is the convex cone matrix. Since we can rewrite D into XA, Eq. (7) can be simplified to Eq. (8):

$$\begin{aligned} &\min_{A,Y} \|X - XAY\|_F^2 \quad s.t. \quad \forall j \in \{1, ...N\} \\ &A \geq 0, \quad \|a_j\|_1 = 1, \quad \|y_j\|_0 \leq T \end{aligned} \tag{8}$$

Adding the convexity constraint to the dictionary gains us some desired properties for the dictionary. Suppose the data are from some clusters, and the optimization is done correctly, the columns of the dictionary should be close to the centroids of the clusters. Getting those centroids is meaningful, since the training data is part of the population, we are trying to recover the original distribution of the data, where the means (centroids) serve as an important role. Another benefit is that, if x_i and x_j are close enough, they will be in the same cluster, which means their sparse code y_i and y_j will be similar. So, CDL preserves the distance among the data in the sparse representation. From this viewpoint, the convex constraint encourages that there are fewer dominating non-zero entries in each sparse representation. Fig. 1 illustrates these points for CDL on a synthetic data set.

3.2. Optimization

The initialization will be discussed in Sect. 4.1. After the initialization, we update Y and A alternatively until they converge. To approximate the optimal solution, we introduce a matrix multiplicative update rule as follows:

Step 1: Update Y (with others fixed): The minimization problem with respect to Y can be decomposed into N subproblems. The sparse representation Y_j can then be computed by using any non-negative pursuit algorithm:

$$\forall j \in \{1, ..., N\}, \quad \min_{y_j} \|x_j - XAy_j\|_2^2 \quad s.t. \quad \|y_j\|_0 \leq T \tag{9}$$

Step 2: Update A (with others fixed): The minimization problem with respect to A can be solved by a multiplicative rule, which is stated as follows:

$$B_{ij} \leftarrow A_{ij}\sqrt{\frac{[(X^TX)^+Y^T + (X^TX)^-AYY^T]_{ij}}{[(X^TX)^-Y^T + (X^TX)^+AYY^T]_{ij}}} \tag{10}$$

$$A_{ij} \leftarrow \frac{B_{ij}}{\sum_i B_{ij}} \tag{11}$$

Algorithm 1: Algorithm for CDL

Input: X,A_0,Y_0,T

while $J = 1$: *max iteration number* **do**

  update Y using eqt. (9);

  update A using eqt. (10)and (11);

  update $J = J + 1$;

Calculate $D = XA$ and normalize it.

Output: D

Here, M_{ij} is the element located in the ith row and the jth column in M, M^+ and M^- are the positive part and negative part of M respectively, $i.e.$

$$M^+ = (|M| + M)/2 \quad M^- = (|M| - M)/2 \tag{12}$$

Step 3: Increase the iteration counter: J = J + 1

Step 1, 2, 3, should be repeated until convergence. For the update of A, it satisfies the KKT condition and the proof can be found in [16], and thus convergence is guaranteed. The procedure is summarized in Algorithm 1.

4. EXPERIMENTAL RESULTS

We apply our CDL method to single image super-resolution, and compare our algorithm to others using the model described in [7]. In our experiments, we use the same training images used in [7]. The low-resolution images in the training set are generated by the high-resolution images using down-sampling with bicubic interpolation. Eight popular images for super-resolution are chosen as the testing set. We use 5×5 image patch pairs for a up-scaling factor 3, and the number of columns of the dictionary is 256. Part of the codes are provided by [7] and [17].

4.1. Experimental Setting

We describe how to initialize A and Y for CDL. We can pick a random vector for $Y^{(0)}$, uniformly distributed from 0 to 1, and then set $A^{(0)} = (Y^{(0)})^TP$, where b_j equals to the l_2 norm of the jth row of $Y^{(0)}$ and $P = diag(1/b_1, ..., 1/b_K)$. Another way is to do a K-means clustering on X. Suppose we obtained the cluster indicator matrix $C = (c_1, ..., c_k)$, where the elements of C is either 0 or 1. Then, by setting $Y^{(0)} =$

Fig. 2. Visual comparison of SR results on “lenna”, “barbara”. The first, second and last column are from bicubic, KSVD, CDL respectively.

$C + \mu E$, where E is a matrix of all 1s, and μ is a small value ($e.g.$ $\mu = 0.2$). The centroids of the clusters can be used to be the columns of dictionary D, then $\forall j$, $d_j = Xc_j/b_j$ or $D = XCB$ where $B = diag(1/b_1, ..., 1/b_K)$. Follwed by calculating B, we then have $A = CB$. This method is also used in [16]. We use the first method in our implementation.

4.2. Comparison

We compare our model with bicubic interpolation [1], and KSVD [14] under the model described in [7] . In Table 1, we compare the peak signal to noise ratio (PSNR) and structural similarity index (SSIM) for the reconstructed high resolution images. The result shows that our approach is consistently better. As illustrated in Fig. 2, bicubic interpolation generates over-smoothed results, KSVD’s results are a little bit blurry at the complex texture, and some artifacts appear near the edges. Our algorithm gives better results for not just recovering more texture details, but also presenting sharper edges with fewer artifacts.

4.3. Further discussion

KSVD is a dictionary-learning algorithm using the singular value decomposition approach, which makes the dictionary built by KSVD captures as much of the energy of the training data. However, in many real world applications, the testing data may not be identical to the training data, which means KSVD may not be the best approach for reconstructing the testing data. Our approach practically attempts to capture the clusters of the data and thus is less affected by such mismatch between training and testing data.

For time complexity, CDL uses a matrix multiplicative update rule for training, which makes its training time much longer than KSVD. However, since training is typically done offline, this is not a practical concern.

5. CONCLUSION AND FUTURE WORK

We proposed a novel dictionary-learning approach to single-image super-resolution. Adding the convex constraint, which forces the columns of the dictionary to be formed by the columns of the input data, makes the columns of the dictionary close to the centroids of the input clusters. We also presented an iterative update algorithm for finding a solution. Experiments on commonly-used testing images showed that our approach is able to better reconstruct high-resolution images. Future work includes applying our proposed approach to multi-frame image super-resolution while considering joint regularization [18].

6. REFERENCES

[1] H.S. Hou and H. Andrews, "Cubic splines for image interpolation and digital filtering," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 26, no. 6, pp. 508–517, Dec 1978.

[2] S. Farsiu, M.D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *Image Processing, IEEE Transactions on*, vol. 13, no. 10, pp. 1327–1344, Oct 2004.

[3] Michael E. Tipping and Christopher M. Bishop, "Bayesian image super-resolution," in *Advances in Neural Information Processing Systems 15*, S. Becker, S. Thrun, and K. Obermayer, Eds., pp. 1303–1310. MIT Press, 2003.

[4] Hong Chang, Dit-Yan Yeung, and Yimin Xiong, "Super-resolution through neighbor embedding," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, June 2004, vol. 1, pp. I–I.

[5] Jian Sun, Jian Sun, Zongben Xu, and Heung-Yeung Shum, "Image super-resolution using gradient profile prior," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, June 2008, pp. 1–8.

[6] Jianchao Yang, J. Wright, T.S. Huang, and Yi Ma, "Image super-resolution via sparse representation," *Image Processing, IEEE Transactions on*, vol. 19, no. 11, pp. 2861–2873, Nov 2010.

[7] Jianchao Yang, Zhaowen Wang, Zhe Lin, S. Cohen, and T. Huang, "Coupled dictionary training for image super-resolution," *Image Processing, IEEE Transactions on*, vol. 21, no. 8, pp. 3467–3478, Aug 2012.

[8] Julien Mairal, Guillermo Sapiro, and Michael Elad, "Learning multiscale sparse representations for image and video restoration," *Multiscale Modeling & Simulation*, vol. 7, no. 1, pp. 214–241, 2008.

[9] N. Kulkarni, P. Nagesh, R. Gowda, and Baoxin Li, "Understanding compressive sensing and sparse representation-based super-resolution," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 5, pp. 778–789, May 2012.

[10] D.L. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[11] Shenlong Wang, D. Zhang, Yan Liang, and Quan Pan, "Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, June 2012, pp. 2216–2223.

[12] Kaibing Zhang, Xinbo Gao, Dacheng Tao, and Xuelong Li, "Multi-scale dictionary for single image super-resolution," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, June 2012, pp. 1114–1121.

[13] Kui Jia, Xiaogang Wang, and Xiaoou Tang, "Image transformation based on learning dictionaries across image spaces," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 2, pp. 367–380, Feb 2013.

[14] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *Signal Processing, IEEE Transactions on*, vol. 54, no. 11, pp. 4311–4322, Nov 2006.

[15] K. Engan, S.O. Aase, and J.H. Husoy, "Frame based signal compression using method of optimal directions (mod)," in *Circuits and Systems, 1999. ISCAS '99. Proceedings of the 1999 IEEE International Symposium on*, Jul 1999, vol. 4, pp. 1–4 vol.4.

[16] C. Ding, Tao Li, and M.I. Jordan, "Convex and semi-nonnegative matrix factorizations," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 1, pp. 45–55, Jan 2010.

[17] Yifeng Li and Alioune Ngom, "The non-negative matrix factorization toolbox for biological data mining," *Source Code for Biology and Medicine*, vol. 8, no. 1, pp. 1–15, 2013.

[18] Kan Chang, Pak Lun Kevin Ding, and Baoxin Li, "Compressive sensing reconstruction of correlated images using joint regularization," *IEEE Signal Processing Letters*, vol. 23, no. 4, pp. 449–453, 2016.