Progress In Electromagnetics Research M, Vol. 22, 73–89, 2012

IMAGE COMPRESSED SENSING BASED ON DATA-DRIVEN ADAPTIVE REDUNDANT DICTIONARIES

Z. W. Ni, M. X. Zhang, J. Li, and Q. C. Wang*

Department of Computer Science, Xiamen University, Siming South Road 422, Xiamen, Fujian 361005, China

Abstract—Finding sparsifying transforms is an important prerequisite of compressed sensing (CS) theory. It is directly related to the reconstruction accuracy. In this work, we propose a new dictionary learning (DL) algorithm to improve the accuracy of CS reconstruction. In the proposed algorithm, Least Angle Regression (LARS) algorithm and an approximate SVD method (ASVD) are respectively used in the two stages. In addition, adaptive sparsity constraint is used in the sparse representation stage, to obtain sparser representation coefficient, which further improves the dictionary update stage. With these data-driven adaptive dictionaries as sparsifying transforms for image compressed sensing, results of experiments demonstrate noteworthy outperformance in peak signal-to-noisy ratio (PSNR), compared to CS based on dictionaries learned by K-SVD, in the sampling rate of 0.2–0.5. Besides, visual appearance of reconstruction detail at low sampling rate improves, for reducing of 'block' effect.

1. INTRODUCTION

Compressed sensing theory indicates that if sparifying transforms can be found, the original signal can be exactly reconstructed through a set of random linear measurements [1–6]. As the sparsity is the key to achieve accurate CS reconstruction, researchers have studied and developed sparsifying transforms with continuous efforts [7–9]. In fact, the sparser the signal representation is, the more accurately the original signal can be recovered [10, 11].

At present, the study of CS still mostly concentrates in fixed orthogonal basis. Tsaig and Donoho in [7] achieve CS reconstruction

Received 30 September 2011, Accepted 22 November 2011, Scheduled 2 December 2011
* Corresponding author: Qi Cong Wang (qcwang@xmu.edu.cn).

based on orthogonal wavelet transform. However, for lacking of translation and rotation invariance, it is not enough to capture the various features of image. Hence, Cands et al. [10] and Rauhut et al. [11] extend CS to redundant dictionaries. They show that if the signal is sparse under a redundant dictionary, and the combination of the dictionary and some random sampling matrix satisfies Restricted Isometry Property (RIP) [1, 3, 4], then CS can still implement via the existing reconstruction algorithms, such as Basis Pursuit (BP) [12]. Obviously, the sparsity of signal can be enhanced using redundant dictionaries, and recovery of original signal can be achieved from less measurements with higher probability.

Recently, there are studies, which take dictionaries learned by K-SVD algorithm [13] as sparsifying transform to improve CS reconstruction accuracy improvement. Ravishankar et al. [14] use K-SVD for highly undersampled magnetic resonance (MR) image reconstruction, and experiments show increase of PSNR by 4–18 dB compared with algorithm in [9]. Bilgin et al. [15] use a K-SVD patch dictionary in iterative hard thresholding (IHT) and improve the reconstruction signal-to-noise ratio (SNR) for 1.5 dB compared to wavelet-IHT. Xu et al. [16] learn a over-complete data-driven dictionary via K-SVD, specialized for speech signals, to obtain superior performance, compared to CS with redundant discrete cosine transform (RDCT) and redundant discrete wavelet transform (RDWT).

However, K-SVD algorithm, in which greedy algorithms commonly used in the sparse representation stage, such as orthogonal matching pursuit (OMP) [17], easily results the whole dictionary learning algorithm in entrapping into a local optimum. Alternatively, the convex relaxation strategies, such as BP and LASSO [18], are too much slower than OMP, although they can get the sparest solution. And singular value decomposition (SVD) searching for exact solutions in the dictionary update stage is not necessary. Besides, the stopping criteria in the sparse representation stage is fixed in the whole DL process.

Therefore, in this paper, we exploit a novel dictionary learning algorithm, aiming to relieve the above limitations of K-SVD. In the sparse representation stage, LARS is used to obtain sparser coefficient for the consequent dictionary update stage. And for simplification, we use an approximate SVD (ASVD) for dictionary update stage. Furthermore, adaptive sparsity is used by associating the maximal number of atoms in the sparse representation stage with the iterated updated dictionaries. In this work, we use dictionaries learned by this method as sparsifying transforms of image compressed sensing.

The rest of this paper is organized as follows. Section 2 describes the related work in CS theory and dictionary learning algorithm briefly.

The proposed algorithm based on data-driven adaptive redundant dictionary learning is detailed in Section 3. Section 4 demonstrates the performance of our algorithm on examples. In Section 5, we conclude with possible topics for future work.

2. COMPRESSED SENSING THEORY AND DICTIONARY LEARNING

2.1. Compressed Sensing

In this paper, we consider CS with measurement noise. Given a real value signal $x \in R^N$, the CS model can be expressed as

$$\min_{\alpha} \|\Phi\Psi\alpha - y\|_2^2 + \lambda\|\alpha\|_1 \tag{1}$$

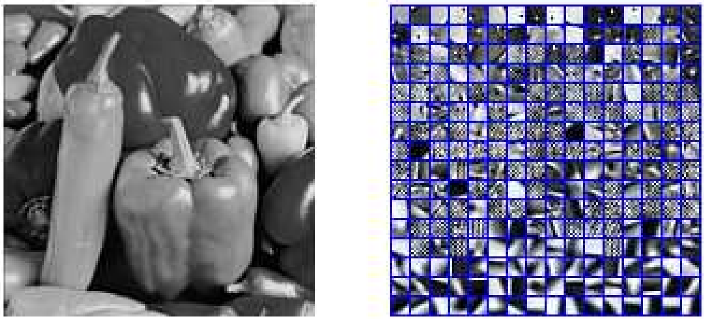
where λ is the Lagrangian multiplier, $\Phi \in R^{m\times N}$ $(m \ll N)$ is the random sampling matrix, y is the corresponding linear measurements, $\Psi \in R^{M\times N}$ is sparsifying transform, and α is sparse coefficient.

A number of heuristic methods can be used to solve this problem. Replace the ℓ_1 norm with ℓ_0 norm, though the problem becomes a NP-hard problem, greedy algorithms, such as OMP, can obtain approximate solution with relatively small computer time. For formulation (1), the minimization problem can also be solved via linear programming algorithms such as BP, least absolute shrinkage method and LASSO. Besides, there exists other sparse approximation algorithms such as the focal underdetermined system solver (FOCUSS) [19] and sparse Bayesian learning [20]. A recent review of the sparse coding algorithms can be found in [21].

In the above model, the sparsifying transform Ψ is one of the two critical parts of CS. From work [4], we know that the representation coefficients of signals will be reconstructed from CS, if they decay like a power-law under a transform. The decay rate of coefficients is a criterion for judging sparsifying transforms ability. Recently, studies in image processing show that superior results have obtained, as signal can be represented sparser under those dictionaries learned from specific samples [22–26].

2.2. Dictionary Learning

For data-driven dictionaries, DL algorithms is the key point. Given a training set $Y = \{y_i\}_{i=1}^N$ which contains N samples, $X = \{x_i\}_{i=1}^N$ is the corresponding coefficient. Then, the DL process is to find a possible optimal dictionary for sparse representation of training samples Y. It



(a) (b)

Figure 1. (a) Image "peppers"; (b) an dictionary learned from (a).

can be expressed as

$$\min_{D,X} \|Y - DX\|_F^2 \quad \text{s.t.} \quad \forall i, \ \|x_i\|_0 \le T_0 \tag{2}$$

where T_0 is the maximum number of atoms used in sparse representation stage. Figure 1(b) is an example of 64×256 dictionary learned by our DL method, here the training set is 8×8 natural image patches randomly extracted from Figure 1(a). In the simulations in Section 4, training sets are patches extracted from variant images different from the test image. In Figure 1(b), each atom in the redundant dictionary is displayed as an 8×8 element.

The DL problem in (2) is also a NP-hard problem. To solve this problem, researchers have proposed many algorithms, such as Maximum Likelihood methods [27], Method of Optimal Directions (MOD) [28], K-SVD and so on. Tosic et al. give a review of these methods in [29]. Among these algorithms, the K-SVD can implement conveniently and effectively.

K-SVD algorithm is implemented by iterating of the following two steps. In the sparse representation step, D is assumed to be fixed. Thus, the minimization problem in (2) can be decomposed into N separate problems in the following formulation

$$\min_{x_i} \|y_i - Dx_i\|_2^2 \quad \text{s.t.} \quad \|x_i\|_0 \le T_0, \quad \text{for} \quad i = 1, 2, \ldots, N \tag{3}$$

these problems is classically solved by OMP in K-SVD. In the dictionary update step, both X and D are assumed to be fixed, and only the kth column d_k in the dictionary D is updated with its corresponding representation coefficients x_T^k. Recalling to the problem in (2), the penalty function becomes

$$\begin{aligned}\|Y - DX\|_F^2 &= \left\| Y - \sum_{j=1}^{K} d_j x_T^j \right\|_F^2 = \left\| \left(Y - \sum_{j \neq k} d_j x_T^j \right) - d_k x_T^k \right\|_F^2 \\ &= \|E_k - d_k x_T^k\|_F^2 \qquad (4)\end{aligned}$$

the solving of (4) is achieved by SVD. The whole K-SVD algorithm iterates between the two stages until convergence, obtaining dictionary D, of which each atom has unit norm. More details about K-SVD can be seen in [13].

3. THE PROPOSED ALGORITHM

Since DL algorithm can always be treated as iterations of sparse representation stage and dictionary update stage, we will follow these two stages to describe the proposed DL algorithm at first.

3.1. Sparse Representation Stage

In this stage, we associate the sparsity constraint with the current dictionary (the iterated updated dictionary) firstly, to obtain an adaptive sparsity constraint. That means replacing T_0 in (3) with T_j. T_j denotes

$$T_j = \left\lceil \frac{1}{2} \left(1 + \frac{1}{\mu(D_j)} \right) \right\rceil \quad \text{for} \quad j = 1, 2, \ldots, P \qquad (5)$$

where $\mu(D)$ is mutual coherence, expressed as

$$\mu(D) := \max_{i \neq j, 1 \leq i, j \leq n} \left\{ \frac{D_i^T D_i}{\|D_i\|_2 \|D_j\|_2} \right\} \qquad (6)$$

It is shown that if inequality (5) holds, the solution X of the minimization problem (3) obtained by OMP will be the sparsest and unique, and it is the same at X of the ℓ_1 alternative obtained by BP [30–32].

In this way, the sparsity upper-bound is adaptive with updated dictionaries, and the reconstruction errors can decrease iteratively. Besides, we replace its ℓ_0 norm in (3) with ℓ_1 norm. Then the problem posed in (3) becomes Lasso's problem

$$\begin{aligned}&\min_{x_i} \{\|y_i - D_j x_i\|_2^2\} \quad \text{s.t.} \quad \|x_i\|_1 \leq T_j, \\ &\text{for} \quad i = 1, 2, \ldots, N; \quad j = 1, 2, \ldots, P \qquad (7)\end{aligned}$$

Here we use LARS with LASSO modification [33], of which computational complexity is close to greedy algorithms to solve the Lasso problem.

Assuming A is a subset of the indices $\{1, 2, \ldots, k\}$, meaning $A \in \{1, 2, \ldots, k\}$, D_{jA} is columns extracted from D_j corresponding to A and let it be equidirectional with Y, X_A is the current solution, and algorithm moves toward the equiangular vectors $u_A = D_{jA}w_A$, in the moving process, the solution increase is

$$X_A(\gamma) = X_A + \gamma w_A \tag{8}$$

where $w_A = (1_A^T(D_{jA}^T D_{jA})^{-1}1_A)^{-\frac{1}{2}}(D_{jA}^T D_{jA})^{-1}1_A$, and γ is step. When signs of X_i and c_i are opposite, then signs between X_i and w_i are opposite, where c_i is correlation between the current covariate and variable Y. As $\gamma_i = -\frac{X_i}{w_i}$, then the opposite sign step appearing earliest is

$$\tilde{\gamma} = \overset{+}{\min_{i \in A}} \left\{ -\frac{X_i}{w_i} \right\} \tag{9}$$

when $\tilde{\gamma} < \hat{\gamma}$, the corresponding $\hat{i}$ should be removed out of A, where $\hat{\gamma}$ is step among equiangular direction u_A.

LARS algorithm and LARS with Lasso modification can be seen in detail in [33].

3.2. Dictionary Update Stage

Given a fixed sparse matrix X, obtained in the above stage, dictionary update can be derived by solving the following problem

$$\min_D \{Y - DX\}_F^2 \tag{10}$$

Typically, the above optimization problem can be solved efficiently using the SVD.

However, exact solver of SVD is not usually required here. The entire K-SVD algorithm only converges to a local minimum rather than a global optimal solution. Therefore, replacing SVD with a faster approximate SVD (ASVD) is a wise choice. ASVD proposed in [34] use a simple iteration between an atom d_k and its corresponding representation coefficient x_T^k to solve problem in (10). The iteration is given as follows

$$\begin{aligned} d_k &= E_k x_k / \|E_k x_k\|_2, \\ x_k &= (E_k)^T d_k \end{aligned} \tag{11}$$

In this way, the result is still very close to the exact SVD solution. Detailed description of the algorithm refers to [34].

Given a $\sqrt{N} \times \sqrt{N}$ large size image I. Divide it into nonoverlapping patches with $\sqrt{n} \times \sqrt{n}$ ($n \ll N$) pixels, and each block is turned into column $I_{ij} \in R^n$, where i, j is the position of the top-left pixel of each patch in the large image respectively. For each patch, the corresponding measurements vector of CS is $v_{ij} = \Phi I_{ij} + \varepsilon_{ij}$, where ε_{ij} is the random noise in measurement process. In the compressed sensing process, we assume that each patch I_{ij} can be represented by a linear combination of $D\alpha_{ij}$, where $\alpha_{ij} \in R^K$ is sparse, $D \in R^{n\times K}$ ($n < K$) (dictionary is redundant and each atom of D has unit norm). The scheme of image compressed sensing based on the proposed DL method is presented in Algorithm 1.

Table 1. Algorithm 1.

Algorithm 1: image compressed sensing based on data-driven adaptive redundant dictionaries
1. Input: training samples Y, initial dictionary D_0, initial sparsity T_0, number of iterations P, measure matrix Φ, linear measurement v_{ij}.
2. Output: reconstructed image $\hat{I}$.
3. use RDCT as the initial dictionary $D = D_0$ and obtain the initial sparsity $T = T_0$ using Equation (5).
4. for $j = 1, \ldots, P$
5. fixed current D_{j-1} and sparsity constraint T_{j-1}, obtain the sparse coefficient X_j corresponding to training samples Y, using Equation (7)
6. given X_j, update dictionary D_j, using Equation (11)
7. compute the sparsity T_j, using Equation (5)
8. end
9. given Φ, v_{ij}, obtain coefficient $\hat{\alpha}_{ij}$ in sparsifying transform $D = D_P$, using Equation (1)
10. obtain the reconstructed image patches, by $\hat{I}_{ij} = D\hat{\alpha}_{ij}$
11. rebuild the reconstructed image $\hat{I}$ form $\hat{I}_{ij}$

4. EXPERIMENTS RESULTS AND ANALYSIS

In this section, we carry out two experiments on natural images. All implementations involved in the experiments are coded in Matlab R2009a. Computations are performed on Window 7 operating system with Intel Core 2 CPU at 2.8 GHz and 2 GB memory.

Like Chen et al. [35], we learn redundant dictionaries using the training samples at first. The four kinds of 64×256 dictionaries being considered here include: (a) LARS + ASVD + ADDAPTIVEL denotes learning dictionaries by the proposed DL algorithm; (b) LARS + ASVD denotes learning dictionaries by algorithm, of which the sparse representation stage adopts LARS and the dictionary update stage adopts ASVD; (c) learning dictionaries by K-SVD algorithm; (d) a dictionary based on scaling and translation of pre-defined basis functions, RDCT. All of the dictionaries in (a)–(c) are initialized by RDCT, and the number of iterations is 10. Besides, for each sample, at most 10 coefficients are used for sparse representation stage in (b) and (c). The K-SVD toolbox used here can be found in [13].

4.1. More Efficient on Sparse Representation

In this experiment, Here, the '1-0001.png' is used as test image, seen in the bottom right corner of Figure 2. The remaining 2413 images in the extended Yale database B are used for randomly extracting the 48260 8×8-pixel training patches, 20 patches from each image. The 192×168 test image is divided into 8×8 nonoverlapping blocks, and each is added with Gaussian random noise, whose the variance is 0.005.

As Figure 2 shows, we can see that all of these four dictionaries can represent the test image well when the sparsity increases to 10 for each nonoverlapping block, where sparsity is the maximum number of non-zeros coefficients obtained by sparse representation with OMP. However, in the cases of lower sparsity, the dictionary learned by the proposed DL algorithm outperforms the other three obviously. Compared (a) with (b), we can find the adaptive sparsity exploited in DL process has great effect on efficient representation of image. See of course, (b) is only a little better than (c). This demonstrates the proposed method can represent image more efficient, and this property can produce superior PSNR during CS reconstruction.

4.2. Improvement on Accuracy in Image Compressed Sensing

This experiment consists of two parts: CS for MR image; CS for standard gray-scale image. The test 512×512 images are divided

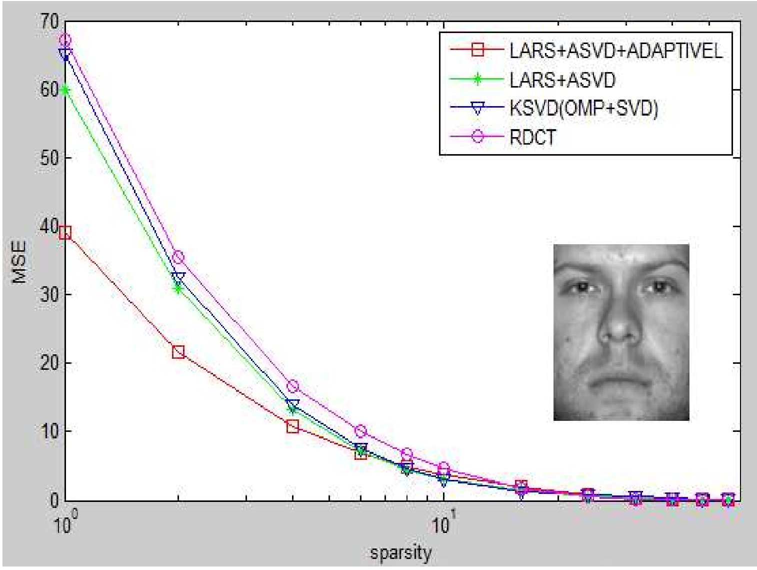


Figure 2. Comparison of the sparse representation recovery mean square error (MSE).

into 8×8 nonoverlapping blocks $\{x_i\}_{i=1,\ldots,N_B}$ as well. Assuming that x_i can be represented by linear combination of $x_i = D\alpha_i$. For each of the N_B patches, the measurement vector is $v_i = \Phi x_i + \varepsilon_i$, where ε_i is Gaussian random noise, of which the variance is 0.005. In this experiment, the elements of Φ are randomly constructed from $N(0,1)$, of course, other similar random measurement matrix in [1] can be used. The PSNR is used as an objective criterion, and visual appearance of reconstruction detail is also considered as a measure. Here, we also consider the four kinds of D in (a)–(d).

In both of the two subsections, the sampling matrices Φ measure the patches randomly, and the sampling rates ($\frac{m}{64}$ in this section) are 0.1, 0.2, 0.3, 0.4 and 0.5 respectively. The CS reconstruction algorithms used here include: BP, LARS, Bayesian Compressive Sensing (BCS) [36], Bayesian compressive sensing using Laplace priors (LAPLACE) [37], and OMP, where BP stops when reaching a maximum iterations number of 20; BCS and LAPLACE stop when the reconstruction error is less than 10^{-8}; and LARS and OMP stop when the used atoms reach 10. The values of PSNR(dB) are averaged over 5 executions.

4.2.1. CS with MR Image

In this subsection, the MR image, seen in the bottom right corner of Figure 3, is used as test image, and the 39988 8×8-pixel training patches are extracted from other 52 MR images different from the test image. These training images are different in resolution and contents.

From Figure 3, compared (b) with (c), the CS reconstruction performance with (b) is better than that with (c), while at most 10 coefficients are used for sparse representation stage in both of them, so we infer that the LARS in sparse representation stage and the ASVD in dictionary update stage is the reason for superiority of (b). And then, compared (a) with (b), the outperformance of (a) is obvious, in fact, (a) is just joins (b) with adaptive sparsity in sparse representation stage, from this, the importance of adaptive sparsity played on dictionary

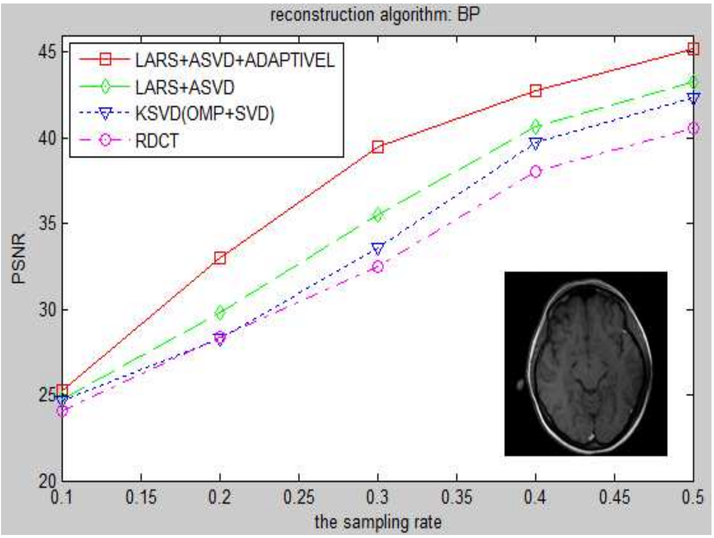


Figure 3. Comparison of CS reconstruction PSNR.

Table 2. PSNR of our method compared with CS based on K-SVD dictionary in various sampling rate.

Method/sampling rate	0.1	0.2	0.3	0.4	0.5
(a) our method	25.23022	32.97324	39.45011	42.76209	45.20174
(c) based on K-SVD	24.707	28.3123	33.5869	39.7406	42.3045
(a)–(c) dB	0.5232	4.6609	5.8632	3.0215	2.8972

learning can be seen. In summary, it is reliable to conclude that our dictionary learning method can find better dictionary for CS.

In Table 2, we compared our method with CS based on K-SVD dictionary in numerical form. It is clear that our method improves the CS reconstruction PSNR about 2.9–5.86 dB compared to K-SVD, when the sampling rate among 0.2 to 0.5. when the sampling rate is very

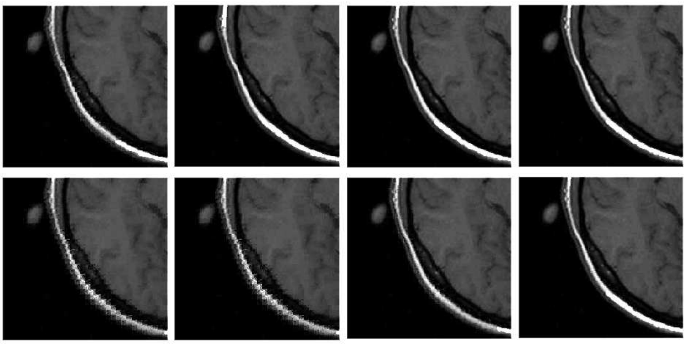


Figure 4. Comparison visual appearance of CS reconstruction details.

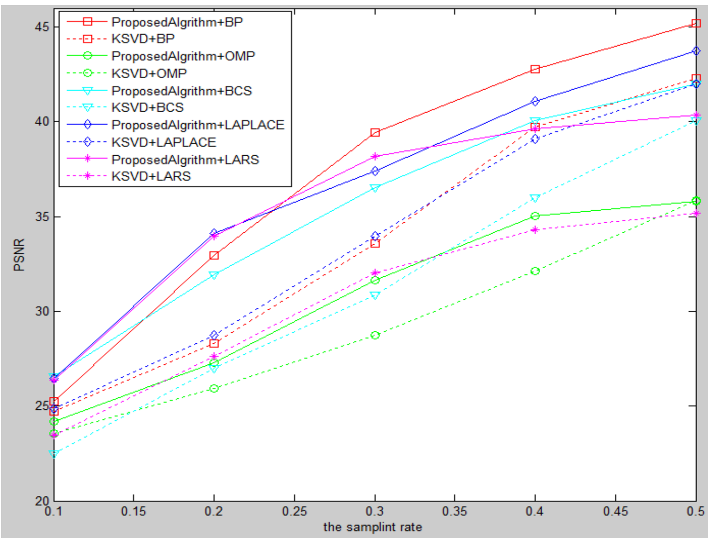


Figure 5. Reconstruction of our method compared to CS based on K-SVD dictionary by BP, OMP, BCS, LAPLACE, and LARS.

low, 0.1, advantage of our method become unconspicuous, however, this is because the sampling rate is too low to well reconstruct for any algorithm.

Figure 4 shows visual appearance of reconstruction details for the left-bottom 1/4 of the test image. The up row is reconstruction of our method, and the bellow is CS based on K-SVD dictionary. From left to right, the sampling rate is 0.2–0.5 respectively. We can find 'block' effect of CS based on K-SVD dictionary, is serious at sampling rate 0.2 and 0.3, while our method can reduce the effect to certain extent.

Furthermore, in Figure 5, we show the superiority of our method compared to CS based on K-SVD dictionary is consistent by various CS reconstruction algorithms. However, because of the limitation of OMP, the results of both method reconstructed by it is not that gratifying. Anyway, dictionary learned by our method do outperform K-SVD dictionary. Besides, the differences of results among various reconstruction algorithms may effect by the stopping criterion. But here, we mainly concentrate on compare between our method and CS based on K-SVD dictionary. So we do not discuss it in detail here.

4.2.2. CS with Standard Gray-scale Image

Similar to the Subsection 4.2.1, in this subsection, we use a standard gray-scale image as test image, seen in the bottom right corner of

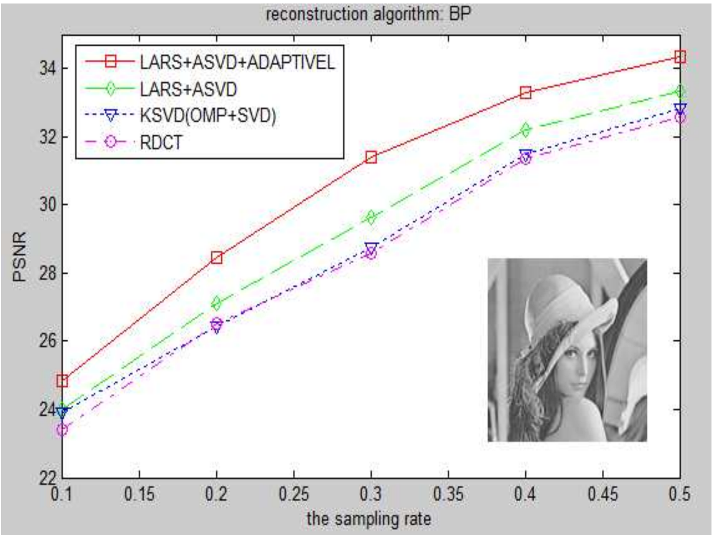


Figure 6. Comparison of reconstruction PSNR.

Figure 6. The 40000 8 × 8-pixel training patches are extracted from 10 standard images different from the test image, here these training images are of same resolution while variant contents.

Figure 6 shows outperformance of our method, compared to CS based on K-SVD dictionary, and RDCT. Here, the results based on K-SVD only get slightly advantage, compared to CS based on RDCT dictionary. We could infer that our dictionary learning algorithm has much more potential of improving a given initial dictionary.

In Table 3, we compare our method with CS based on K-SVD dictionary in numerical form. It is clear that our method improves the CS reconstruction PSNR about 0.93–2.65 dB compared to K-SVD, when the sampling rate among 0.1 to 0.5.

Figure 7 shows visual appearance of reconstruction details for the upper-right 1/4 of the test image. The up row is reconstruction of our method, and the bellow is CS based on K-SVD dictionary. From left to right, the sampling rate is 0.2–0.5 respectively. We can find 'block' effect of CS based on K-SVD dictionary, is a little serious at sampling rate 0.2, while our method can reduce the effect to certain extent.

From Figure 8, we can see that like results in Figure 5, the

Table 3. PSNR of our method compared with CS based on K-SVD dictionary in various sampling rate.

Method/sampling rate	0.1	0.2	0.3	0.4	0.5
(a) our method	24.8599	28.4606	31.4155	33.3128	34.3633
(c) based on K-SVD	23.9284	26.4215	28.7654	31.4739	32.8383
(a)–(c) dB	0.9315	2.0391	2.6501	1.8389	1.525

Figure 7. Comparison visual appearance of CS reconstruction details.

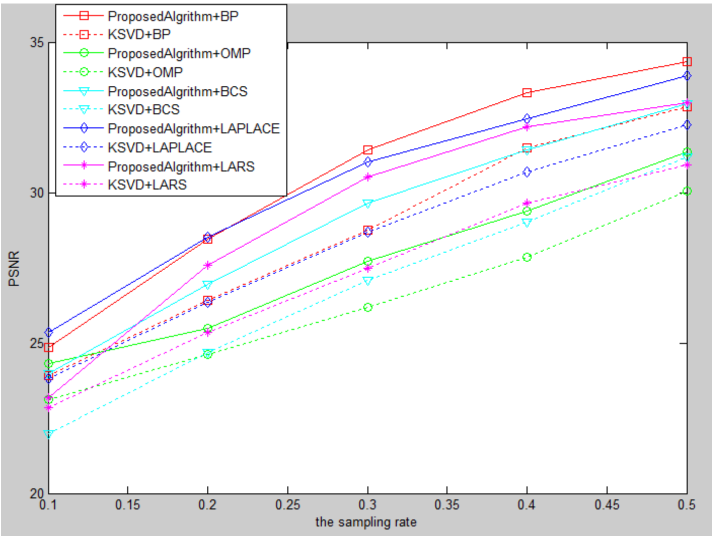


Figure 8. Reconstruction of our method compared to CS based on K-SVD dictionary by BP, OMP, BCS, LAPLACE, and LARS.

superiority of our method compared to CS based on K-SVD dictionary is consistent by various CS reconstruction algorithms. Though there are much differences among various reconstruction algorithms.

In a word, the results in this subsection agree with the ones in the above subsection, meaning that our method does have its outperformance.

5. CONCLUSION

In this paper, we present a novel method for learning patch-sparse data-driven adaptive redundant dictionaries to improve the accuracy of CS reconstruction. In the proposed DL algorithm, LARS with LASSO modification and adaptive sparsity constraint are used in the sparse representation stage, and ASVD is used in the following dictionary update stage. Experiments show that the CS reconstruction performance is improved both in PSNR and visual appearance, compared to CS based on K-SVD dictionaries. Since the image is treated as blocks, 'block' effect is clear at low sampling rate. In the future, we will try hard to solve this problem. Besides, we will apply the dictionary learning method to face recognition and CSMRI, and the work is ongoing.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (No. 61001143).

REFERENCES

1. Candes, E. J., J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, Vol. 52, No. 2, 489–509, 2006.
2. Donoho, D., "Compressed sensing," *IEEE Trans. Inf. Theory*, Vol. 52, No. 4, 1289–1306, 2006.
3. Candes, E. J. and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, Vol. 51, No. 12, 4203–4215, 2005.
4. Candes, E. J. and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, Vol. 25, No. 2, 21–30, 2008.
5. Davenport, M., M. Duarte, Y. C. Eldar, and G. Kutyniok, *Introduction to Compressed Sensing*, Cambridge University Press, 2011.
6. Fornasier, M. and H. Rauhut, *Compressive Sensing*, Springer, 2011.
7. Tsaig, Y. and D. L. Donoho, "Extensions of compressed sensing," *Signal Processing*, Vol. 86, No. 3, 549–571, 2006.
8. Gan, L., "Block compressed sensing of natural images," *Proc. Int. Conf. on Digital Signal Processing*, 403–406, Cardiff, UK, 2007.
9. Lustig, M., D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, Vol. 58, No. 6, 1182–1195, 2007.
10. Candes, E. J., Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Applied and Computational Harmonic Analysis*, Vol. 31, No. 1, 59–73, 2010.
11. Rauhut, H., K. Schnass, and P. Vandergheynst, "Compressed sensing and redundant dictionaries," *IEEE Trans. Inf. Theory*, Vol. 54, No. 5, 2210–2219, 2008.
12. Chen, S. S., D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, Vol. 43, No. 1, 129–159, 2001.

13. Aharon, M., M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Processing*, Vol. 54, No. 11, 4311–4322, 2006.
14. Ravishankar, S. and Y. Bresler, "MR image reconstruction from highly undersampled K-space data by dictionary learning," *IEEE Trans. Medical Imaging*, Vol. 30, No. 5, 1028–1041, 2011.
15. Bilgin, A., Y. Kim, F. Liu, and M. S. Nadar, "Dictionary design for compressed sensing MRI," *Proc. ISMRM*, 2010.
16. Xu, T. T., Z. Yang, and X. Shao, "Adaptive compressed sensing of speech signal based on data-driven dictionary," *Conf. 15-th Asia-Pacific*, 2009.
17. Pati, Y. C., R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," *Proc. AsilomarSS*, Vol. 1, 40–44, 1993.
18. Tibshirani, R., "Regression shrinkage and selection via the lasso," *J. Royal. Statist. Soc. B*, Vol. 58, No. 1, 267–288, 1996.
19. Gorodnitsky, I. and B. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Processing*, Vol. 45, No. 3, 600–616, 1997.
20. Wipf, D. P. and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Processing*, Vol. 52, No. 8, 2153–2164, 2004.
21. Tropp, J. A. and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE (Special Issue on Applications of Sparse Representation and Compressive Sensing)*, Vol. 98, No. 6, 948–958, 2010.
22. Elad, M. and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, Vol. 15, No. 12, 3736–3745, 2006.
23. Protter, M. and M. Elad, "Image sequence denoising via sparse and redundant representations," *IEEE Trans. Image Process.*, Vol. 18, No. 1, 27–36, 2009.
24. Mairal, J., M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Trans. Image Process.*, Vol. 17, No. 1, 53–69, 2008.
25. Mairal, J., G. Sapiro, and M. Elad, "Learning multiscale sparse representations for image and video restoration," *SIAM Multiscale Model. Simulat.*, Vol. 7, No. 1, 214–241, 2008.

26. Bryt, O. and M. Elad, "Compression of ficial images using the K-SVD algorithm," *J. Visual Communication and Image Representation*, Vol. 19, No. 4, 270–283, 2008.
27. Olshausen, B. A. and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?" *Vis. Res.*, Vol. 37, No. 23, 3311–3325, 1997.
28. Engan, K., S. Aase, and J. Hakon Husoy, "Method of optimal directions for frame design," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP'99)*, 1999.
29. Tosic, I. and P. Frossard, "Dictionary learning," *IEEE Signal Process. Mag.*, Vol. 28, No. 2, 27–38, 2011.
30. Donoho, D. L. and M. Elad, "Optimally sparse representation in general (non-orthogonal) dictionaries via minimization," *Proc. Nat. Aca. Sci.*, Vol. 100, No. 5, 2197–2202, 2003.
31. Gribonval, R. and M. Nielsen, "Sparse representation in unions of bases," *IEEE Trans. Inf. Theory*, Vol. 49, No. 12, 3320–3325, 2003.
32. Tropp, J. A., "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, Vol. 50, No. 10, 2231–2242, 2004.
33. Efron, B., T. Hastie, I. Johnston, and R. Tibshirani, "Least angle regression," *Ann. Statist.*, Vol. 32, No. 2, 407–499, 2004.
34. Rubinstein, R., M. Zibulevsky, and M. Elad, "Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit," *Technical Report-CS Technion.*, 2008.
35. Chen, Y., X. Ye, and F. Huang, "A novel method and fast algorithm for MR image reconstruction with significantly undersampled data," *Inverse Problems Image*, Vol. 4, No. 2, 223–240, 2010.
36. Ji, S., Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Processing*, Vol. 56, No. 6, 2346–2356, 2008.
37. Babacan, S. D., R. Molina, and A. K. Katsaggelos, "Bayesian compressive sensing using Laplace priors," *IEEE Trans. Image Process.*, Vol. 19, No. 1, 53–63, 2010.