BOOSTED MULTI-SCALE DICTIONARIES FOR IMAGE COMPRESSION

Mansour Nejati[1], Shadrokh Samavi[1], Nader Karimi[1], S. M. Reza Soroushmehr[2,3], Kayvan Najarian[2,3,4]

[1]ECE Department, Isfahan University of Technology, Iran
[2]Department of Emergency Medicine, University of Michigan, Ann Arbor, USA
[3]University of Michigan Center for Integrative Research in Critical Care (MCIRCC), USA
[4]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, USA

ABSTRACT

Sparse representations over redundant dictionaries have shown to produce high quality results in various signal and image processing tasks. Recent advancements in learning of the sparsifying dictionaries have made image compression based on sparse representation a promising field. In this paper, we present a boosted dictionary learning framework to construct an ensemble of complementary specialized dictionaries for sparse image representation. Boosted dictionaries along with a competitive sparse coding can provide us with more efficient sparse representations. Based on the proposed ensemble model, we then develop a new image compression algorithm using boosted multi-scale dictionaries learned in the wavelet domain. Our algorithm is evaluated for compression of natural images. Experimental results demonstrate that the proposed algorithm has better rate-distortion performance as compared with several competing compression methods including analytic and learned dictionary schemes.

Index Terms— Image compression, boosted dictionary learning, sparse representation, wavelet.

1. INTRODUCTION

Compression of natural images is realized by capturing and exploiting redundancies found in these images. The most widely used image compression technique is transform-based coding where quantization and encoding are applied on the transform coefficients of the image [1]. This approach has been extensively investigated in the past three decades leading to powerful compression standards including JPEG [2] and JPEG2000 [3]. The widely used JPEG image compression standard is based on discrete cosine transform (DCT) while its successor, JPEG2000 standard, employs wavelet transform to obtain more compact representations. There are several other wavelet-based compression algorithms such as the well-known Set Partitioning in Hierarchical Trees (SPIHT) algorithm [4].

These compression methods are general-purpose, in the sense that they are not tailored to specific classes of images. They employ an analytic dictionary of basis elements over which the image is known to be compressible. However, dictionary design has evolved over the past decades. The latest trend is to use learned and data-adaptive dictionaries [5]. Sparse representation over the learned dictionaries instead of analytically predefined ones has been shown to produce impressive results in various image processing tasks [6]. Targeted at a specific class of images, the learned dictionaries have also shown promising results in several recent works on compression of facial images [7,8], fingerprint images [1], and synthetic aperture radar (SAR) images [9,10]. Compression of general images using the learned dictionaries has also been investigated in [11-13].

In this paper, we target natural image compression using dictionary ensembles. We first present a Boosted Dictionary Learning (BDL) framework for compact image representation. In this framework, we use a boosting strategy to train a set of dictionaries sequentially, where each dictionary is optimized for the training samples having high representation errors in the previous ones. Given this ensemble of different specialized dictionaries, the most fitted dictionary is adaptively selected for each input data to be sparsely approximated. Based on the proposed boosting framework, we then develop a new image compression scheme in which the BDL is applied in the analysis domain of the Wavelet transform to create boosted multi-scale dictionaries. The experimental results on natural images demonstrate that our algorithm outperforms competing compression methods including both analytic and learned dictionary schemes.

The rest of the paper is organized as follows. In Section 2, we elaborate on the details of our boosted dictionary learning approach. Section 3 introduces the proposed image compression scheme based on boosted dictionaries. Experimental results are presented in Section 4 and we conclude the paper in Section 5.

2. BOOSTED DICTIONARY LEARNING

The dictionary plays a critical role in a successful sparse representation modeling and learned overcomplete dictionaries have been popular in recent years. In image processing applications it is common to train dictionaries for sparse representation of small image patches collected from a number of images. Recurrence of these local regions in natural images allows building of effective representative dictionaries. However, by learning based on a generic image dataset, we may not accurately capture the appearance of all image patches using a single, universal, and compact dictionary. This means that a trained dictionary could produce better sparse approximations for some training image patches as compared to some other ones.

To alleviate this problem, we present a boosted dictionary learning (BDL) approach to train a sparsifying dictionary ensemble consisting of multiple over-complete dictionaries with complementary representational powers. Fig. 1(a) illustrates the block diagram of the BDL. As shown in this figure, the BDL applies a given base learning algorithm repeatedly in series of boosting rounds $l = 1, \dots, L$ to learn a set of dictionaries $\{\mathbf{D}_l\}_{l=1}^{L}$ sequentially. Each round proceeds in three stages of dictionary training, data partitioning, and dictionary refinement. Given a set of N training image patches, $\mathbf{Y} = \{\mathbf{y}_j \in \mathbb{R}^n\}_{j=1}^{N}$, the training stage of each round l is conducted as a joint optimization problem with respect to the dictionary $\mathbf{D} \in \mathbb{R}^{n\times K}$ $(K > n)$ and sparse representations $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbb{R}^{K\times N}$:

$$(\widetilde{\mathbf{D}}_l, \widetilde{\mathbf{X}}_l) = \underset{\mathbf{D},\mathbf{X}}{\operatorname{argmin}}\, f(\mathbf{D}, \mathbf{X}, \mathbf{Y}_l), \quad \forall l = 1, \dots, L \tag{1}$$

where $\mathbf{Y}_l \subseteq \mathbf{Y}$ denotes the training data for round l, and f is objective function of the base learning algorithm. In reconstructive tasks such as our case, f is generally composed by a data fitting term and a sparsity-inducing norm [14]. Let us denote Λ_l as the index set of training samples used for learning $\widetilde{\mathbf{D}}_l$. At the beginning of boosting, Λ_l includes indices of all training samples (i.e. $\Lambda_1 = \{1,2, \dots, N\}$), but it is modified in the subsequent rounds. In fact, the second stage of each round partitions the set $\mathbf{Y}_l$ into disjoint subsets $\widetilde{\mathbf{Y}}_l$ and $\mathbf{Y}_{l+1}$. The data partitioning is accomplished based on the representation error of sparse coefficients obtained for the training set. For this purpose, we compute the sparse representations $\{\mathbf{x}_{l,j}\}_{j\in\Lambda_l}$ of $\mathbf{Y}_l = \{\mathbf{y}_j\}_{j\in\Lambda_l}$ with respect to $\widetilde{\mathbf{D}}_l$ under the same sparsity level fixed to a small value τ_0. This is done using orthogonal matching pursuit (OMP) method [15]. Let $\mathbf{e}_l = \{e_{l,j}\}_{j\in\Lambda_l}$ be the set of residual errors calculated as:

$$e_{l,j} = \left\|\mathbf{y}_j - \widetilde{\mathbf{D}}_l\mathbf{x}_{l,j}\right\|_2^2, \quad \forall j \in \Lambda_l \tag{2}$$

The set $\mathbf{e}_l$ is sorted in ascending order to get a set $\mathbf{e}_l^s = \{e_{l,1}^s, \dots, e_{l,|\Lambda_l|}^s\}$. Then, the index set Λ_{l+1} of training samples for next round of boosted learning is determined as:

$$\Lambda_{l+1} = \left\{ j \in \Lambda_l : e_{l,j} \geq e^s_{l,\lfloor \frac{N}{L} \rfloor} \right\} \tag{3}$$

where $\lfloor x \rfloor$ returns the integer part of x. Also, the complementary set of Λ_{l+1} in Λ_l, denoted by $\overline{\Lambda}_{l+1} = \{i \in \Lambda_l : i \notin \Lambda_{l+1}\}$

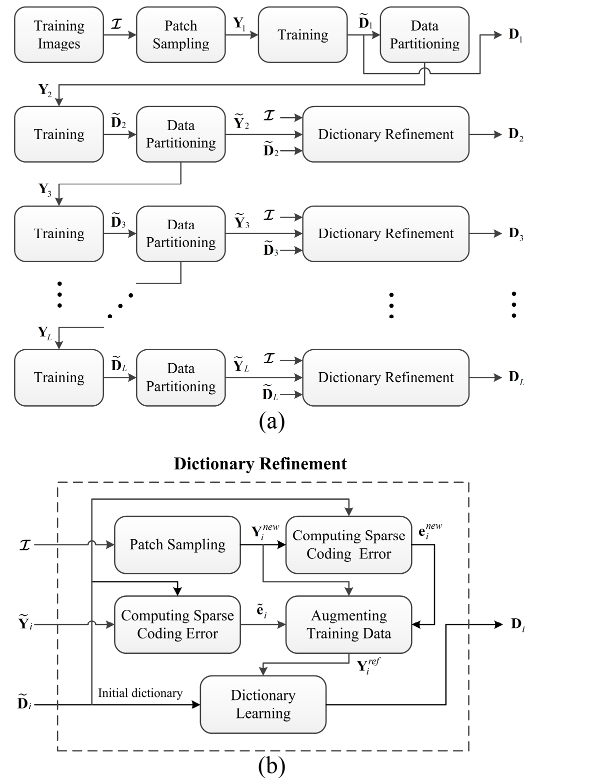


Fig. 1. (a) Steps of the proposed boosted dictionary learning scheme. (b) Block diagram of dictionary refinement stage.

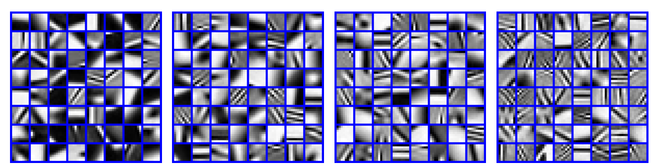


Fig. 2. Example of dictionaries obtained by our boosting approach with four rounds of boosting trained on 8×8 image patches.

, specifies the subset $\widetilde{\mathbf{Y}}_l$ which is presented at the input of the dictionary refinement stage. The refinement stage for l-th round is conducted by enrichment of the training set $\widetilde{\mathbf{Y}}_l$ with new samples and refinement of dictionary $\widetilde{\mathbf{D}}_l$. The goal of refinement is to focus on specific group of samples and optimizing dictionary for representation of them. This in turn increases sparsity and improves reconstruction performance of dictionary on that group. In this stage, we first extract new samples $\mathbf{Y}_l^{new} = \{\mathbf{y}_{l,j}^{new}\}_{j=1}^{N}$ from training images and obtain their representations $\{\mathbf{x}_{l,j}^{new}\}_{j=1}^{N}$ over $\widetilde{\mathbf{D}}_l$. Let $\tilde{\mathbf{e}}_l = \{e_{l,j} : j \in \overline{\Lambda}_{l+1}\}$ be the error set corresponding to $\widetilde{\mathbf{Y}}_l$. Based on $\tilde{\mathbf{e}}_l$ and residual error of new samples, $e_{l,j}^{new}$, the refinement set $\mathbf{Y}_l^{ref}$ is produced as follows.

$$\mathbf{Y}_l^{ref} = \widetilde{\mathbf{Y}}_l \cup \{\mathbf{y}_{l,j}^{new} : \min(\tilde{\mathbf{e}}_l) \leq e_{l,j}^{new} \leq \max(\tilde{\mathbf{e}}_l), j = 1{:}N\} \tag{4}$$

Using $\widetilde{\mathbf{D}}_l$ as an initial dictionary, $\mathbf{Y}_l^{ref}$ is used to train a refined dictionary $\mathbf{D}_l$ (see Fig. 1(b)).

After completion of all boosting rounds, the ensemble dictionaries $\{\mathbf{D}_l\}_{l=1}^{L}$ are obtained. Fig. 2 demonstrates an example of a trained ensemble using BDL approach learned with $L = 4$ on a set of 8×8 image patches. In this example, K-SVD algorithm [16] is employed as the base learning algorithm in each boosting round. As can be seen in Fig. 2, BDL produces dictionaries which are progressively adapted to image patches with more complex content.

3. IMAGE COMPRESSION ALGORITHM

3.1. Boosted Multi-scale Dictionary Ensemble

Targeted at compact representation of natural images, we obtain data-adaptivity by training the dictionary ensembles using the BDL approach. In our proposed compression method, we adopt the Wavelet transform as the first layer of image sparsification and apply the learning algorithm on Wavelet coefficient patches. In this way, spatial correlations among Wavelet coefficients are exploited to reduce some of the remaining redundancies among the coefficients. Moreover, due to the multi-scale properties of the Wavelet decomposition, a multi-scale dictionary can be learned in this way [17]. It is known that different Wavelet bands possess different directional correlations. Thus, the band-specific dictionaries are trained to push sparsity further by using highly specialized dictionaries. If S indicates the number of Wavelet decomposition levels, $3S$ band-specific dictionary ensembles $\mathbb{D}_b = \{\mathbf{D}_l^b\}_{l=1}^L, b = 1,2,\ldots,3S$ are trained for detail bands of Wavelet transform. These dictionaries are obtained in an off-line process and stored both in the encoder and decoder.

3.2. Sparse Coding

In the compression stage, the learned dictionary ensembles are used to encode Wavelet coefficient patches. More precisely, the input image is firstly decomposed using a S-level Wavelet transform into $3S + 1$ bands including $3S$ detail bands plus an approximation image. Each detail band is sliced into non-overlapping patches which are sparse coded by using the corresponding dictionary ensemble. The sparse coding is accomplished in a competitive way by generating a group of competing representations and choosing the best one based on their representation accuracy. Let $\mathbf{y}$ be the input image and $[\mathbf{W}_A\mathbf{y}]^b$ denote the coefficient band b of its Wavelet decomposition using Wavelet analysis operator $\mathbf{W}_A$. The competitive coding of the j-th patch from the band b, denoted by $[\mathbf{W}_A\mathbf{y}]_j^b$, with respect to the dictionary ensemble $\mathbb{D}_b = \{\mathbf{D}_l^b\}_{l=1}^L$ is started by solving the error-constrained sparse coding problem:

$$\mathbf{x}_{1,j}^b = \underset{\mathbf{x}}{\operatorname{argmin}} \ \|\mathbf{x}\|_0 \ \text{ s.t. } \left\|[\mathbf{W}_A\mathbf{y}]_j^b - \mathbf{D}_1^b\mathbf{x}\right\|_2^2 \le \epsilon^2 \qquad (5)$$

where the error threshold ϵ varies with the compression rate. Only the first dictionary in each boosting ensemble is globally learned using total training data. The sparsity levels of other competing representations are determined based on this global dictionary. Assuming an ensemble $\mathbb{D}_b = \{\mathbf{D}_l^b\}_{l=1}^L$ of L dictionaries for the band b, the competitive coding method allows each patch to select its optimal dictionary from the ensemble as:

$$\gamma_j^b = \underset{1 \le l \le L}{\operatorname{argmin}} \left\|[\mathbf{W}_A\mathbf{y}]_j^b - \mathbf{D}_l^b\mathbf{x}_{l,j}^b\right\|_2^2 \qquad (6)$$

where γ_i^b is the index of optimal dictionary for the coefficient patch $[\mathbf{W}_A\mathbf{y}]_j^b$.

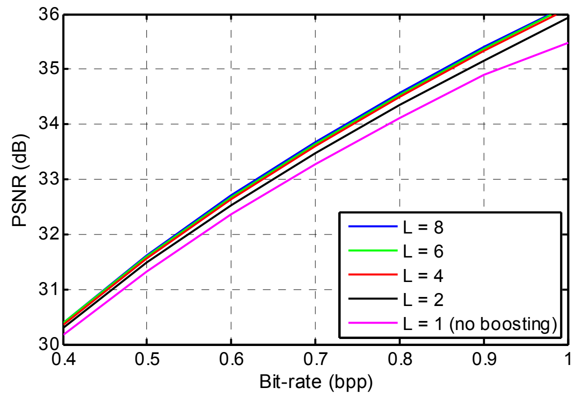


Fig. 3. Rate-distortion performance of our method using BDL with varying number of boosting rounds L.

3.3. Quantization and Entropy Coding

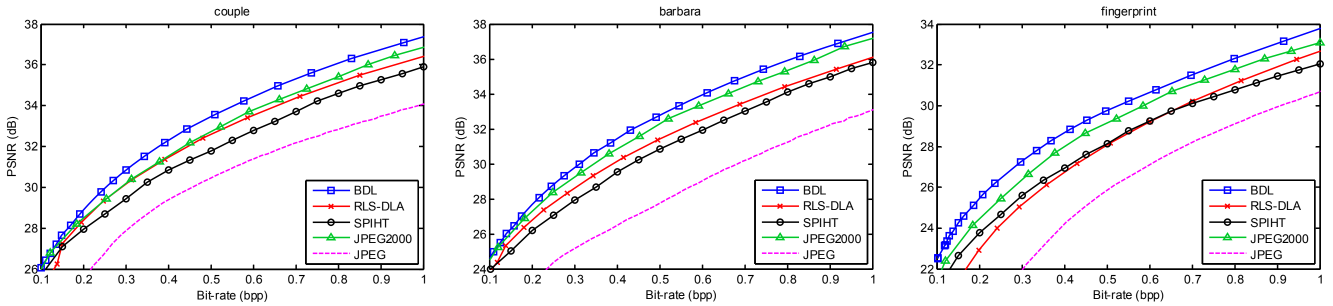
Once the sparse representation of all non-overlapping patches of all detail bands are obtained, they are arranged as columns of sparse matrix $\mathbf{X} \in \mathbb{R}^{K \times M}$, where M is the total number of patches. This matrix is quantized using a uniform quantizer to form $\mathbf{X}_q$. We put the nonzero values of $\mathbf{X}_q$ into one sequence and the running differences of their position indices into another sequence. These two sequences, along with the set $\mathbf{\Gamma}$ of the optimal dictionary indices are then encoded using arithmetic entropy coding.

The approximation coefficients of Wavelet transform at the coarsest scale are also uniformly quantized and entropy coded. Quantization is performed using the same quantization step size δ as for the sparse coefficients $\mathbf{X}$. To adjust the value of δ, an iterative process is done until a pre-specified quantization-induced PSNR loss is reached. We empirically found that a δ giving a PSNR loss in the range of [0.9-1.0] dB, is an appropriate value for it. To compress the quantized approximation coefficients, we use the adaptive prediction technique [18].

4. EXPERIMENTS

4.1. Assessment of Boosted Learning

To verify the effectiveness of the proposed BDL in providing more efficient sparse representations, we apply it in our image compression scheme and evaluate the rate-distortion performance with respect to the number of



(a) (b) (c)

Fig. 4. Rate-distortion performance of the proposed method compared to JPEG2000 [3], JPEG [2], SPIHT [4] and RLS-DLA [11] on three standard test images: (a) *Couple*, (b) *Barbara*, (c) *Fingerprint*.

Table 1. PSNR comparison of the proposed compression method with competing ones at three different bit-rates. Best results are in bold.

Rate (bpp)	0.2				0.6				1.0			
Images	JPEG2000	SPIHT	RLS-DLA	BDL	JPEG2000	SPIHT	RLS-DLA	BDL	JPEG2000	SPIHT	RLS-DLA	BDL
Barbara	27.30	26.18	26.78	**27.69**	33.39	31.92	32.54	**33.94**	37.17	35.80	36.08	**37.51**
Boat	29.16	28.61	28.94	**29.52**	34.19	33.42	33.95	**34.52**	36.73	35.95	36.39	**36.97**
Bridge	24.71	24.41	24.64	**24.97**	28.38	27.94	28.60	**28.98**	31.25	30.41	31.26	**31.76**
Couple	28.51	27.93	28.43	**28.91**	33.78	32.79	33.54	**34.46**	36.85	35.88	36.41	**37.37**
F.print	24.48	23.77	22.99	**25.48**	30.17	29.24	29.20	**30.66**	33.09	32.04	32.67	**33.78**
Hill	29.79	29.35	29.73	**30.16**	33.90	33.27	33.92	**34.48**	36.47	35.70	36.46	**37.03**
Lena	33.02	32.73	32.93	**33.44**	38.04	37.58	37.95	**38.51**	40.42	39.95	40.25	**40.85**
Pirate	27.37	27.14	27.43	**27.78**	32.13	31.54	32.03	**32.55**	34.99	34.50	34.67	**35.30**
Average	28.04	27.52	27.73	**28.49**	33.00	32.21	32.72	**33.51**	35.87	35.03	35.52	**36.32**

boosting rounds. In this experiment, we use a set of 100 randomly chosen images from the Berkeley segmentation dataset [19] where 30 images are employed for training dictionaries and the rest for testing. These images are transformed using two-level 9/7 Wavelet decomposition. The train set is prepared by random sampling of 8×8 patches from the Wavelet detail bands of training images. For the whole set of training images, 80000 patches are extracted and vectorized for each band.

Under different settings for L, we train a dictionary ensemble for each detail band with $K = 512$ using K-SVD [16] as the base learning algorithm. The boosted ensembles are then employed in our compression algorithm to compress the test images. Fig. 3 shows the rate-distortion curves resulted from these ensembles, where each curve is averaged over the test images. As can been seen, boosted learning (i.e. $L > 1$) improves the performance compared to the traditional learning (i.e. $L = 1$), and the improvement increases with L. However, with incrementing of L, the gain becomes smaller due to the extra cost imposed for transmission of dictionary indices $\mathbf{\Gamma}$. In the following, this parameter is fixed to $L = 6$.

4.2. Comparison to other Compression Methods

We compare the proposed compression algorithm with four competing methods: JPEG, JPEG2000, SPIHT [4], and RLS-DLA [11] which uses a learned dictionary in the Wavelet domain. The performance evaluation is conducted on a set of standard test images shown in Table 1. Construction of boosted dictionaries follows the same procedure described in the previous section. Fig. 4 shows comparison of rate-distortion curves for three images of test set. As can be observed, our compression method outperforms the other ones particularly for rates above 0.2 bpp. The PSNR results on test images at three distinct rates are reported in Table 1. Compared to the RLS-DLA, our method based on boosted learning provides an average PSNR gain of about 0.8 dB at low and moderate bit-rates.

In terms of computational cost, our method takes about 3.8 seconds for compressing a 512×512 image at 0.2 bpp using a non-optimized MATLAB implementation on a 3.40 GHz Intel Core i7 CPU. This is comparable to the compression time of RLS-DLA method (3.5s). Decompression time for both methods is less than 1s.

5. CONCLUSION

In this paper, we presented a boosting approach to dictionary learning for the purpose of compact image representation. Using a competitive sparse coding, with a set of boosted dictionaries, improves the quality of sparse image approximations. Based on this approach, we developed a new image compression algorithm using boosted dictionaries learned in the Wavelet domain. Assessment of rate-distortion performance on a set of natural images, confirmed the effectiveness of the proposed boosted dictionary learning approach.

6. REFERENCES

[1] G. Shao, Y. Wu, Y. A, X. Liu, and T. Guo, “Fingerprint compression based on sparse representation,” *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 489–501, Feb. 2014.

[2] W. B. Pennebaker and J. L. Mitchell, *JPEG still image data compression standard*, Springer, New York, 1993.

[3] A. Skodras, C. Christopoulos, and T. Ebrahimi, “The JPEG 2000 still image compression standard,” *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 36–58, 2001.

[4] A. Said and W. A. Pearlman, “A new fast and efficient image codec based on set partitioning in hierarchical trees,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 3, pp. 243–250, June 1996.

[5] R. Rubinstein, A. M. Bruckstein, and M. Elad, “Dictionaries for sparse representation modeling,” *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1045–1057, June 2010.

[6] M. Elad, M. Figueiredo, and Y. Ma, “On the role of sparse and redundant representations in image processing,” *Proceedings of the IEEE*, vol. 98, no. 6, pp. 972–982, June 2010.

[7] O. Bryt and M. Elad, “Compression of facial images using the KSVD algorithm,” *Journal of Visual Communication and Image Representation*, vol. 19, no. 4, pp. 270–283, 2008.

[8] J. Zepeda, C. Guillemot, and E. Kijak, “Image compression using sparse representations and the iteration-tuned and aligned dictionary,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 5, pp. 1061– 1073, 2011.

[9] X. Zhan, R. Zhang, D. Yin, and C. Huo, “SAR image compression using multiscale dictionary learning and sparse representation,” *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 5, pp. 1090–1094, Sept. 2013.

[10] X. Zhan, R. Zhang, D. Yin, A. Hu and W. Hu, “Remote sensing image compression based on double-sparsity dictionary learning and universal trellis coded quantization,” in *Proc. IEEE International Conference on Image Processing*, Melbourne, VIC, Sept. 2013, pp. 1665–1669.

[11] K. Skretting and K. Engan, “Image compression using learned dictionaries by RLS-DLA and compared with K-SVD,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Prague, May 2011, pp. 1517–1520.

[12] I. Horev, O. Bryt, and R. Rubinstein, “Adaptive image compression using sparse dictionaries,” in *Proc. 19th International Conference on Systems, Signals and Image Processing*, Vienna, April 2012, pp. 592–595.

[13] K. S. Gurumoorthy, A. Rajwade, A. Banerjee, and A. Rangarajan, “A method for compact image representation using sparse matrix and tensor projections onto exemplar orthonormal bases,” *IEEE Transactions on Image Processing*, vol. 19, no. 2, pp. 322–334, Feb. 2010.

[14] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. Huang, and S. Yan, “Sparse representation for computer vision and pattern recognition,” *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1031–1044, June 2010.

[15] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, “Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition,” in *Conference Record of the 27th Asilomar Conference on Signals, Systems and Computers*, Nov. 1993, pp. 40–44.

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

[17] B. Ophir, M. Lustig, and M. Elad, “Multi-scale dictionary learning using wavelets,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 5, pp. 1014–1024, Sept. 2011.

[18] I. Matsuda, N. Ozaki, Y. Umezu and S. Itoh, “Lossless coding using variable block-size adaptive prediction optimized for each image,” in *Proc. of 13th European Signal Processing Conference EUSIPCO-05*, Sept. 2005.

[19] Berkeley segmentation dataset: www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/