Computing and Informatics, Vol. 38, 2019, 151–172, doi: 10.31577/cai_2019_1_151

LOSSY COMPRESSIVE SENSING BASED ON ONLINE DICTIONARY LEARNING

İrem Ülkü

Department of Electrical and Electronics Engineering
Çankaya University
Eskişehir Yolu 29. Km
06790 Ankara, Turkey
e-mail: iremulku@cankaya.edu.tr

Ersin Kizgut

Instituto Universitario de Matemática Pura y Applicada (IUMPA)
Universitat Politècnica de València
E-46071 Valencia, Spain
e-mail: erkiz@upv.es

Abstract. In this paper, a lossy compression of hyperspectral images is realized by using a novel online dictionary learning method in which three dimensional datasets can be compressed. This online dictionary learning method and blind compressive sensing (BCS) algorithm are combined in a hybrid lossy compression framework for the first time in the literature. According to the experimental results, BCS algorithm has the best compression performance when the compression bit rate is higher than or equal to 0.5 bps. Apart from observing rate-distortion performance, anomaly detection performance is also tested on the reconstructed images to measure the information preservation performance.

Keywords: Hyperspectral imaging, compression algorithms, dictionary learning, sparse coding

Mathematics Subject Classification 2010: 68U10, 94A08

1 INTRODUCTION

Hyperspectral images are composed of hundreds of contiguous narrow (generally $0.010\,\mu$m) spectral bands from the visible region (0.4–0.7 μm) to the near-infrared region (about $2.4\,\mu$m) of the electromagnetic spectrum. Hyperspectral images have a huge image size. Therefore, to cope with storage or transmission issues, and to match the available transmission bandwidth in the downlink operation, the hyperspectral image compression is compulsory. Compression can be realized as lossless or lossy. Lossy compression algorithms can reach high compression ratios while experiencing information loss. Quality metrics should capture the degradation which occurs in the image.

Classification of lossy and lossless compression methods is canonically fourfold: prediction-based [24, 31], transformation-based [9, 28], vector quantization (VQ)-based [32], and sparse representation-based [18]. One of the transformation-based algorithms is the principal component analysis (PCA). The PCA algorithm realizes the decorrelation of spectral bands. The improved version of this algorithm is compressive-projection principal component analysis (CPPCA) algorithm [11].

Sparse representation-based methods appear to distinguish among others with its scheme. Rather than using a pre-defined dictionary, such methods learn it directly from the observed data. Data-dependent dictionaries are gathered using dictionary learning [6, 27, 41, 48]. Two different learning schemes can be considered, a batch method which uses the whole training set in the learning process at each iteration and an online learning method which processes one sample from the entire training set at each iteration in an alternating fashion. By using the singular value decomposition (SVD), K-SVD algorithm is developed which can be given as a typical example of batch methods [7]. In the work [21], online dictionary learning algorithm is proposed.

Using dictionary learning in the lossy hyperspectral image compression algorithms is quite common [18, 38, 40]. In the literature [36, 37], it is shown that online dictionary learning algorithm is more effective in processing large datasets with sequentially arriving samples such as hyperspectral images. This sparse representation process finds the sparsest solution, which means solving the non-deterministic polynomial-time hard (NP-hard) ℓ_0-norm minimization problem [10].

Sparse representation algorithms are analyzed in three categories [44, 48]. These are greedy pursuit algorithms, ℓ_p-norm regularization based algorithms, and Bayesian inference algorithms [33, 45]. The most popular greedy pursuit algorithms are the matching pursuit (MP) algorithm [22], the orthogonal matching pursuit (OMP) algorithm [35], the generalized OMP (gOMP) algorithm [39] and the compressive sampling matching pursuit (CoSaMP) algorithm [25].

We may think of ℓ_p-norm regularization algorithms as of two kinds depending on the value of p, namely, for $p \geq 1$ and $0 < p < 1$. In $p \geq 1$ category, only the ℓ_1-norm minimization is sufficiently sparse [48]. The ℓ_1-norm minimization algorithms are divided into three such as constraint based, proximity based, and homotopy based optimization algorithms. The constraint based optimization algorithms category includes the truncated Newton based interior-point method (TNIPM) algorithm [20],

the alternating direction method of the multipliers (ADMM) algorithm [5] and the active-set algorithm with either a primal or dual type [12]. The ADMM algorithm is used to solve the least absolute shrinkage and selection operator (LASSO) problem. The dual active-set algorithm is used to solve the basis pursuit (BP) problem.

The proximity based optimization algorithms category covers the sparse reconstruction by separable approximation (SpaRSA) algorithm [42], the general iterative shrinkage and thresholding (GIST) algorithm [15], the Shotgun algorithm [13] and the augmented Lagrangian method (ALM) algorithm which consists of the primal ALM (PALM) and dual ALM (DALM) [43].

Basic homotopy based algorithms are the LASSO homotopy algorithm [8] and the basis pursuit denoising (BPDN) homotopy algorithm [3].

The generalized iterated shrinkage algorithm (GISA) and the focal underdetermined system solver (FOCUSS) algorithm [16] are analyzed under the non-convex ℓ_p-norm ($0 < p < 1$) regularization algorithms.

The Bayesian inference algorithms category includes the Bayesian compressive sensing projected Landweber based on three-dimensional bivariate shrinkage plus 3D wavelet packet transform (BCS PL-3DBS + 3DWPT) algorithm [17] and the sparse Bayesian learning (SBL) algorithm [34] increase the performance when OBD-BCS algorithm is used. This expectation is reasonable given the rate-distortion performance results, since OBD-BCS algorithm outperforms the others.

Model-based CS algorithms aim to integrate the structured sparsity models into CS algorithms [4]. An algorithm called JSM-2 is a model-based CS algorithm [25].

The blind compressed sensing (BCS) algorithm solves the CS problem without prior knowledge of the sparsity basis [46]. In this case, to guarantee the unique solution, three constraints are considered on the sparsity basis [14]. The algorithm used in the process is called an orthonormal block diagonal BCS (OBD-BCS) algorithm. Each iteration consists of an OMP algorithm and singular value decomposition (SVD) algorithm. There are a handful of studies [29, 23] which use BCS for hyperspectral image reconstruction purpose. In this paper, however, BCS is utilized only in the solution of the sparse coding equation. It is, indeed, not a part of dictionary learning, but rather a tool for finding the sparse coefficients. After finding the sparse coefficients, online dictionary learning is performed as usual. Therefore, previously the BCS algorithm is not implemented with online dictionary learning method. The main contributions of this study are as follows:

1. Different sparse representation algorithms are used to compress hyperspectral images based on online dictionary learning. The compression performances of these algorithms are compared with the performances of the state-of-the-art lossy compression algorithms. This study used the results from the previous studies and therefore only the best performing sparse representation algorithms are included.
2. This is the first time that the BCS algorithm is used in conjunction with the online dictionary learning method for hyperspectral image compression purposes.

3. The anomaly detection technique is applied to further test the information preservation performance of the lossy hyperspectral image compression.

The lossy compression of the hyperspectral images based on online dictionary learning is presented in Section 2. The results are presented in Section 3. The conclusions are given in Section 4.

2 LOSSY COMPRESSION OF HYPERSPECTRAL IMAGES BASED ON DICTIONARY LEARNING

In this section, online dictionary learning based sparse coding on hyperspectral image compression is explained. Sparse coding models the data as the sparse linear combination of the dictionary elements. Dictionary learning is based on learning the dictionary to adapt it to specific data. The online dictionary learning method relies on stochastic approximations and it is suitable for large scale datasets such as hyperspectral images [21]. In this study, the iterative online dictionary learning algorithm is used, which minimizes the surrogate function of the empirical cost under particular constraints at each iteration [21].

2.1 Notation and Problem Statement

In the analysis, the number of bands in the hyperspectral image is represented by n_b, the number of lines in the hyperspectral image is represented by nl, the number of samples in the hyperspectral image is represented by ns, and the number of columns in the dictionary is denoted k. The initial dictionary is expressed as $\mathbf{D}_0 \in \mathbb{R}^{n_b \times k}$. The auxiliary matrices for updating the dictionary are denoted $\mathbf{A}_0 \in \mathbb{R}^{k \times k}$ and $\mathbf{B}_0 \in \mathbb{R}^{n_b \times k}$. The number of iterations is represented by T, the error is expressed as $\mathbf{E} \in \mathbb{R}^{k \times 1}$, the regularization parameter is denoted $\lambda \in \mathbb{R}$, and the sparse coefficients are shown by $\boldsymbol{\alpha} \in \mathbb{R}^k$.

In the dictionary learning process, optimization is performed on the empirical cost by considering a finite training set $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_T]$ in $\mathbb{R}^{n_b \times T}$ [26]. The empirical cost is given as

$$f_T(\mathbf{D}) := \frac{1}{T} \sum_{i=1}^{T} l(\mathbf{x}_i, \mathbf{D}), \tag{1}$$

where $\mathbf{D} \in \mathbb{R}^{n_b \times k}$ represents the dictionary and l expresses the loss function. This loss function corresponds to the optimal value of ℓ_1 norm sparse coding [21] given by the equation

$$l(\mathbf{x}_t, D) := \min_{\boldsymbol{\alpha} \in \mathbb{R}^k} \frac{1}{2} \|\mathbf{x}_t - \mathbf{D}\alpha_t\|_2^2 + \lambda \|\alpha_t\|_1 \tag{2}$$

where λ represents the regularization parameter, x_t expresses the training sample at iteration t and α_t defines the coefficient set at iteration t. In (2), ℓ_1 regularization ensures the sparsity.

A convex set of matrices C is defined to constraint arbitrarily large elements in $\mathbf{D} = [\mathbf{d}_1 \dots \mathbf{d}_k]$ as well as arbitrarily small values of α_t. This convex set C is given

$$C := \{\mathbf{D} \in \mathbb{R}^{n_b \times k} : \|\mathbf{d}_j\| \leq 1, \forall j = 1, \dots k\}. \tag{3}$$

In the optimization, the minimization of the empirical cost $f_T(\mathbf{D})$ with respect to $\mathbf{D}$ is not a convex operation. According to this issue, the process is modified as a joint optimization problem. The modified optimization problem is convex when the sparse coefficients $\mathbf{\Gamma} := [\alpha_1, \dots, \alpha_T] \in \mathbb{R}^{k \times T}$ are fixed, while the optimization is performed with respect to $\mathbf{D}$, and when $\mathbf{D}$ is fixed while the optimization is performed with respect to sparse coefficients $\mathbf{\Gamma}$. This joint optimization problem is as follows:

$$\min_{\mathbf{D} \in C, \mathbf{\Gamma} \in \mathbb{R}^{k \times T}} \frac{1}{T} \sum_{i=1}^{T} \left(\frac{1}{2} \|\mathbf{x}_i - \mathbf{D}\alpha_i\|_2^2 + \lambda \|\alpha_i\|_1 \right). \tag{4}$$

Equation (4) is solved as a convex optimization problem such that $\mathbf{D}$ is minimized when $\mathbf{\Gamma}$ is fixed, and $\mathbf{\Gamma}$ is minimized when $\mathbf{D}$ is fixed, respectively. Instead of minimizing the empirical cost $f_T(\mathbf{D})$, minimizing the expected cost $f(\mathbf{D})$ is much more computationally efficient. This expected cost is given as

$$f(\mathbf{D}) := E_x[l(\mathbf{x}, \mathbf{D})] = \lim_{T \to \infty} f_T(\mathbf{D}) \tag{5}$$

where the unknown probability distribution of the data is utilized to find out the expectation. In the literature, it has been proved that the equality in (5) converges with the probability one [21].

For large scale data sets such as hyperspectral images, stochastic gradient algorithms provide a better rate of convergence [21]. Therefore, in this study, dictionary learning is realized by using projected first order stochastic gradient descent algorithm. According to this algorithm, dictionary $\mathbf{D}$ is updated sequentially and is shown as [2].

$$\mathbf{D}_t = \prod_C [\mathbf{D}_{t-1} - \frac{\rho}{t} \nabla_D \, l(\mathbf{x}_t, \mathbf{D}_{t-1})] \tag{6}$$

where $\mathbf{D}_t$ represents the optimal dictionary at iteration t, ρ presents the gradient step, and $\prod_C$ shows the orthogonal projector on C. It is assumed that the training set $\mathbf{X}$ has i.i.d. samples of the unknown distribution of the particular data [21].

2.2 Algorithm

In this study, an algorithm which consists of two parts is used. These two parts, namely dictionary learning and dictionary update, are solved alternately. The sparse coding equation is solved by using $\mathbf{x}_t$ from the current iteration, and $\mathbf{D}_{t-1}$ from the previous iteration. When α_t is found, the following $\widehat{f}_t(\mathbf{D})$ function is minimized over

set C to obtain an updated dictionary $\mathbf{D}_t$:

$$\widehat{f}_t(\mathbf{D}) := \frac{1}{t}\sum_{i=1}^{t}\frac{1}{2}\|\mathbf{x}_i - \mathbf{D}\alpha_i\|_2^2 + \lambda\|\alpha_i\|_1 \tag{7}$$

where α_i values are obtained. In the literature, it has been proved that the empirical cost $f_t(\mathbf{D})$ and the function $\widehat{f}_t(\mathbf{D})$, which is quadratic in the $\mathbf{D}$ converge is almost surely to the same limit [21]. Therefore, the function $\widehat{f}_t$ is the surrogate for the function f_t. For the large values of t, the function $\widehat{f}_t$ is close to $\widehat{f}_{t-1}$ function. In these circumstances, $\mathbf{D}_t$ is also close to $\mathbf{D}_{t-1}$ such that it is effective to use $\mathbf{D}_{t-1}$ as a warm restart for finding $\mathbf{D}_t$. At each iteration Algorithm 1 finds the sparse coefficients, while Algorithm 2 uses these sparse coefficients to update the current dictionary. Using various different sparse representation algorithms for the solution of the sparse coding equation in Algorithm 1, best performing algorithm can be determined, enabling a comparison between state-of-the-art algorithms. Online dictionary learning, which is the main implementation in Algorithm 2 will be used for all scenarios.

2.2.1 Algorithm 1

In Algorithm 1 the sparse coding equation is solved. Equation (2) is called sparse coding equation. The value of λ is set to 0.1 while T equals 200.

Algorithm 1 Dictionary Learning

1: Construct random initial dictionary $\mathbf{D}_0$
2: Set initial values $\mathbf{A}_0$ and $\mathbf{B}_0$ matrices to zero
3: **for** $t = 1$ to T **do**
4: Choose $\mathbf{x}_t \in \mathbb{R}^{n_b}$ randomly from the image.
5: Solve sparse coding equation.
6: Update $\mathbf{A}_t = \mathbf{A}_{t-1} + \boldsymbol{\alpha}_t\boldsymbol{\alpha}_t^T$, $\mathbf{B}_t = \mathbf{B}_{t-1} + \mathbf{x}_t\boldsymbol{\alpha}_t^T$.
7: Find $\mathbf{D}_t$ using Algorithm 2.
8: **end for**
9: Obtain learned dictionary $\mathbf{D}_t$.

2.2.2 Algorithm 2

In Algorithm 2 dictionary is updated by utilizing the block-coordinate descent with $\mathbf{D}_{t-1}$ as a warm restart. Equation (7) is called as the dictionary update equation. Algorithm 1 and Algorithm 2 are applied in an alternating fashion which is the online learning strategy. Various algorithms are used to solve sparse coding equation (cf. Table 2).

Algorithm 2 Dictionary Update

Calculate $\mathbf{D}_t$ in dictionary update equation
repeat
  for $j = 1$ to k **do**
    Find jth column of $\mathbf{D}_t$, where $\mathbf{D} = [\mathbf{d}_1 \dots \mathbf{d}_k] \in \mathbb{R}^{n_b \times k}$, $\mathbf{A} = [\mathbf{a}_1 \dots \mathbf{a}_k] \in \mathbb{R}^{k \times k}$ and $\mathbf{B} = [\mathbf{b}_1 \dots \mathbf{b}_k] \in \mathbb{R}^{n_b \times k}$
    $\mathbf{u}_j := \frac{1}{A(j,j)}(\mathbf{b}_j - \mathbf{D}\mathbf{a}_j) + \mathbf{d}_j$
    $\mathbf{d}_j = \dfrac{1}{\max(\|\mathbf{u}_j\|_2, 1)}\mathbf{u}_j$
    $E_j = \sqrt{\sum_{n_b} |\mathbf{d}_j^t - \mathbf{d}_j^{t-1}|^2}$
  end for
  $E = \frac{1}{k}\sum_{j=1}^{k} E_j$
until $E <$ Treshold
Use $\mathbf{D}$ in Algorithm 1.

3 RESULTS

The online dictionary learning based hyperspectral image compression is applied by using AVIRIS and Hyperion datasets for all the different sparse representation algorithms [19]. The compression performances of these algorithms are compared with the performances of the state-of-the-art lossy compression algorithms such as BCS PL-3DBS + 3DWPT and CPPCA. BCS PL-3DBS + 3DWPT and CPPCA algorithms are not based on learning while the remaining ones are employed by an online learning scheme. The quality metric tool is the Peak Signal-to-Noise Ratio (PSNR). The bit rate r is calculated in terms of the bits per sample (bps), and the formulation is as follows

$$r = \frac{z}{n_b}(b_d), \quad z < k \tag{8}$$

where z represents the number of sparse coefficients, k defines the size of the dictionary, n_b is the number of bands, and b_d represents the bit depth.

3.1 Datasets

The information about the AVIRIS and Hyperion datasets which are used in this study are given in Table 1 [19].

3.2 AVIRIS Datasets Results

Low Altitude, Lunar Lake, and the Jasper Ridge are used as the AVIRIS datasets (cf. Table 1). In Table 2, the compression performances of different sparse representation algorithms are shown. The quality metric tool, which reflects the compression performance, is PSNR in terms of dB. The PSNR values are calculated against

the compression ratios in terms of the bps. The state-of-the-art algorithms BCS PL-3DBS + 3DWPT and CPPCA, which are given in Table 2, are used for the comparison [17]. The highest three PSNR values per each compression ratio are marked in boldface. If Table 2 is analyzed at the highest compression ratio of 0.5 bps, only the OBD-BCS algorithm involves among the algorithms with the best three PSNR values for all datasets.

AVIRIS HYPERSPECTRAL DATA					
Name	No. Samples	No. Lines	No. Bands	Bit Depth	Year
Jasper Ridge	614	2 587	224	16	1997
Lunar Lake	614	1 432	224	16	1997
Low Altitude	614	3 689	224	16	1996
HYPERION HYPERSPECTRAL DATA					
Name	No. Samples	No. Lines	No. Bands	Bit Depth	Year
Lake Monona	256	3 176	242	12	2009
Mt. St. Helens	256	3 242	242	12	2009
Erta Ale	256	3 187	242	12	2010
SALINAS-A HYPERSPECTRAL DATA					
	No. Samples	No. Lines	No. Bands	Bit Depth	Year
	83	86	204	12	1998
PAVIA UNIVERSITY HYPERSPECTRAL DATA					
	No. Samples	No. Lines	No. Bands	Bit Depth	Year
	200	200	103	12	2002
INDIANA HYPERSPECTRAL DATA					
	No. Samples	No. Lines	No. Bands	Bit Depth	Year
	145	145	220	12	1992

Table 1. Detailed information of AVIRIS, Hyperion, Salinas-A, Pavia and Indiana hyperspectral datasets

3.3 Hyperion Datasets Results

The Erta Ale, Mt. St. Helens, and Lake Monona images are used as Hyperion datasets (cf. Table 1). In Figures 1, 2 and 3, the PSNR values of these datasets against 0.1, 0.3, and 0.5 bps compression ratios for all sparse representation algorithms, are given. The PSNR values are expressed in terms of dB, and they are plotted against the compression ratios in terms of bps. The corresponding compression ratios of the algorithms with highest three PSNR values are shown in circles.

As seen from Figures 1, 2, and 3, at the highest compression ratio of 0.5 bps, the SpaRSA algorithm appears among the best three algorithms for all the datasets, while the OBD-BCS algorithm is situated among the best three algorithms for the Mt. St. Helens and Lake Monona datasets. Therefore, at high compression

	Lunar Lake Image													
	Sparse Representation Algorithms													
BPS	BCS PL_3DBS + 3DWPT	BP (Dual active set)	gOMP	LASSO (ADMM)	CPPCA	SpaRSA	GIST	BPDN (Homo-topy)	GISA $p = 0.4$	OBD-BCS	SBL	FOCUSS	Shotgun	JSM-2
0.1	54.74	**59.96**	59.79	59.59	47.47	59.88	59.85	59.82	59.74	**60**	**59.97**	59.93	59.79	59.62
0.3	61.74	**70.16**	**70.28**	68.85	60.98	69.78	69.79	69.04	67.64	**69.99**	68.15	68.52	69.73	67.74
0.5	67.08	**73.24**	72.68	**73.52**	70.01	72.82	73.21	72	71.9	**73.56**	72.74	73.09	71.81	68.9
	Jasper Ridge Image													
	Sparse Representation Algorithms													
BPS	BCS PL_3DBS + 3DWPT	BP (Dual active set)	gOMP	LASSO (ADMM)	CPPCA	SpaRSA	GIST	BPDN (Homo-topy)	GISA $p = 0.4$	OBD-BCS	SBL	FOCUSS	Shotgun	JSM-2
0.1	**61.34**	59.55	58.37	59.54	48.43	59.51	**59.68**	59.57	59.58	59.22	**59.6**	59.54	59.57	59.51
0.3	69.38	**73.85**	73.84	73.34	72.19	**73.89**	73.62	68.58	69.41	**73.97**	72.07	73.16	69.87	71.19
0.5	72.62	**76.55**	74.92	75.2	**76.82**	75.07	75.37	71.81	71.57	75.45	75.01	74.19	74.47	72.1
	Low Altitude Image													
	Sparse Representation Algorithms													
BPS	BCS PL_3DBS + 3DWPT	BP (Dual active set)	gOMP	LASSO (ADMM)	CPPCA	SpaRSA	GIST	BPDN (Homo-topy)	GISA $p = 0.4$	OBD-BCS	SBL	FOCUSS	Shotgun	JSM-2
0.1	56.78	59.41	59.4	59.3	30.2	**59.47**	58.83	59.32	59.28	**59.44**	58.6	59.37	59.42	**59.43**
0.3	64.21	69.23	70.01	70.67	**71.31**	70.69	70.15	69.56	68.45	70.57	**70.74**	**70.83**	70.73	69.42
0.5	69.95	71.71	71.14	**73.17**	**76.4**	72.49	72.24	72.44	70.46	**72.54**	72.46	72.17	72.27	71.26

Table 2. Compression performance comparison between sparse representation algorithms and state-of-the-art compression algorithms [17]

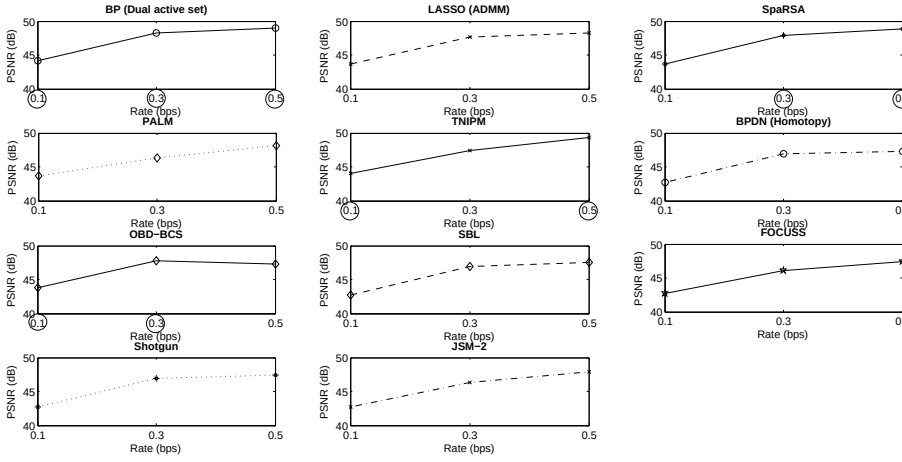


Figure 1. Compression performances of sparse representation algorithms for Erta Ale image (cf. Table 1)

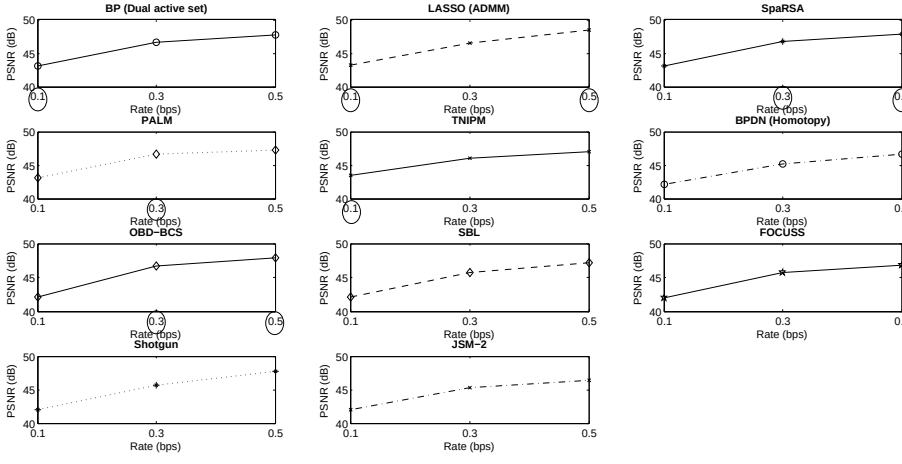


Figure 2. Compression performances of sparse representation algorithms for Mt. St. Helens image (cf. Table 1)

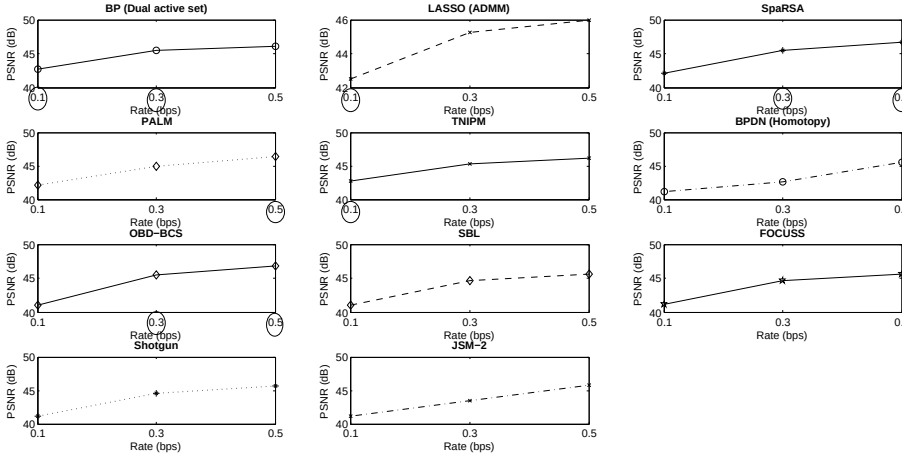


Figure 3. Compression performances of sparse representation algorithms for Lake Monona image (cf. Table 1)

ratios, the SpaRSA and OBD-BCS algorithms show better compression performances.

3.4 Comparison with Several HCS Methods

In literature, a novel reweighted Laplace prior based hyperspectral compressive sensing (RLPHCS) method named as RLPHCS_Cov outperforms several state-of-the-art HCS algorithms [47]. This compression method is not based on learning. For further comparison, compression performance of the algorithm RLPHCS_Cov is compared to that of the sparse representation algorithms based on online dictionary learning.

The signal to noise ratio (SNR) is fixed at 20dB. Pavia University and Indiana datasets are used (cf. Table 1). Figures 4 and 5 show PSNR curves of different algorithms at various bps levels when Pavia University and Indiana datasets are used, respectively. Online dictionary learning (ODL) and hyperspectral compressive sensing (HCS) abbreviations are used.

Figures 4 and 5 indicate that the reconstruction performance of the OBD-BCS (ODL) algorithm is superior to that of the other algorithms at 0.5 bps level. Although for 0.5 bps compression level the OBD-BCS (ODL) algorithms is better for both datasets, setting the compression ratio to moderate levels such as 0.3 bps yields better RLPHCS_Cov performance for Pavia University dataset.

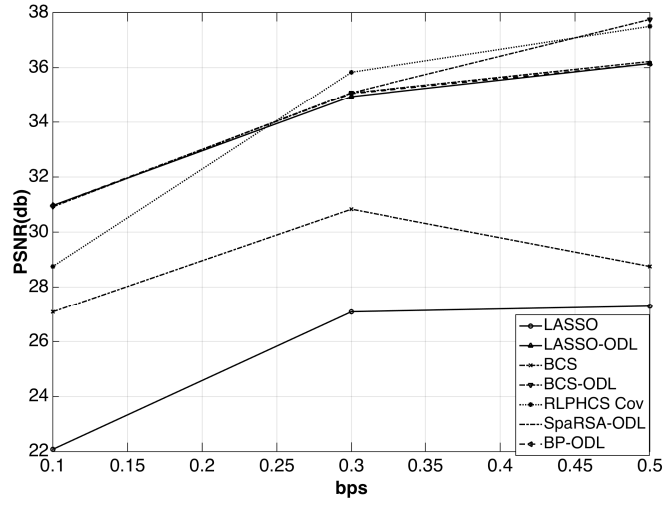


Figure 4. The reconstruction performances of different methods for Pavia University dataset when SNR is 20 db

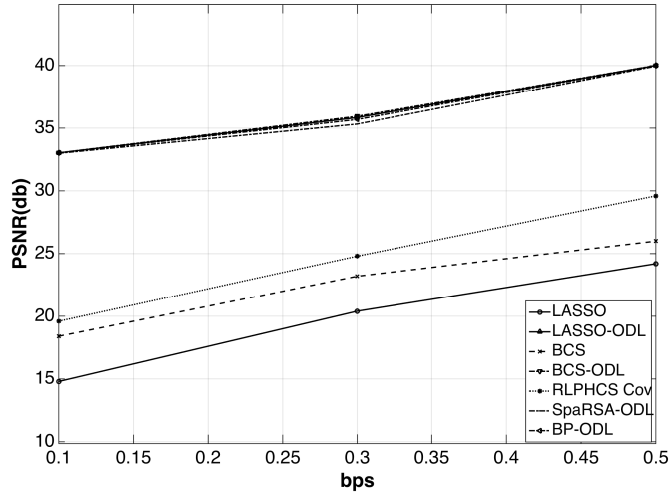


Figure 5. The reconstruction performances of different methods for Indiana dataset when SNR is 20 db

3.5 Compression Performance Analysis of OBD-BCS Algorithm

According to Table 2, the OBD-BCS algorithm involves among the top three algorithms at the highest compression ratio of 0.5 bps for all the datasets.

In Figures 1, 2, and 3, the OBD-BCS algorithm belongs to the top three algorithms with the highest PSNR values at the highest ratio of 0.5 bps for the Mt. St. Helens and Lake Monona datasets.

The results indicate that the OBD-BCS algorithm shows a better compression performance when the compression ratio gets higher. OBD-BCS algorithm is being considered as a compressive sensing framework. However, in this study it is only used in Algorithm 1 to solve the sparse coding equation. Since the OBD-BCS algorithm itself includes a dictionary learning process, learning is applied not just in Algorithms 2 for online dictionary learning, but also in Algorithm 1 while finding the sparse coefficients. It is expected that using these more accurate sparse coefficients in online dictionary learning will increase the performance when OBD-BCS algorithm is used. This expectation is reasonable given the rate-distortion performance results, since OBD-BCS algorithm outperforms the others.

3.6 Anomaly Detection Application

The anomaly detection is applied to make a further comparison between various sparse representation methods. It is a useful tool for assessing the information preservation ability of these methods. Reed-Xiaoli (RX) anomaly detection algorithm is used [30].

Spectral signature which belongs to the input signal is compared with the mean values of each spectral band by using Mahalanobis distance,

$$\delta_{RX}(\mathbf{x}_i) = (\mathbf{x}_i - \mathbf{M})^T \mathbf{Cov}^{-1}(\mathbf{x}_i - \mathbf{M}) \tag{9}$$

where $\mathbf{x}_i \in \mathbb{R}^{n_b}$, $\mathbf{M}$ represents the mean of each spectral band and $\mathbf{Cov}$ indicates the spectral covariance matrix. Covariance matrix $\mathbf{Cov}$ is as follows:

$$\mathbf{Cov} = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{x}_i - \mathbf{M})(\mathbf{x}_i - \mathbf{M})^T \tag{10}$$

where $N = nl \times ns$ and $i = 1, \ldots N$.

Anomalous region is assumed to be present if $\delta_{RX}(\mathbf{x}_i) \geq \eta$ condition is satisfied, where η represents the threshold value. The most appropriate threshold value is the one that is obtained from the desired false alarm probability. Anomaly detection is applied on Salinas-A and Low Altitute hyperspectral datasets (cf. Table 1) [1]. Sparse representation based on online dictionary learning algorithms such as BP by using dual active set algorithm, LASSO by using ADMM algorithm, SpaRSA and OBD-BCS are utilized.

The anomaly detection results are illustrated in Figure 10 for Salinas-A dataset. First, the anomaly detection is applied on the original hyperspectral dataset whose results are presented in Figure 10 a). The desired anomaly is marked with a circle. Figure 10 b)–e) depict anomaly detection results for OBD-BCS, BP by using dual active set, SpaRSA and LASSO algorithms at 0.5, 0.3 and 0.1 bps levels, respectively. None of the algorithms is able to detect the desired anomaly at 0.1 bps bit rate. Among the anomaly detection results at 0.5 bps bit rate, OBD-BCS algorithm seems to provide the best performance.

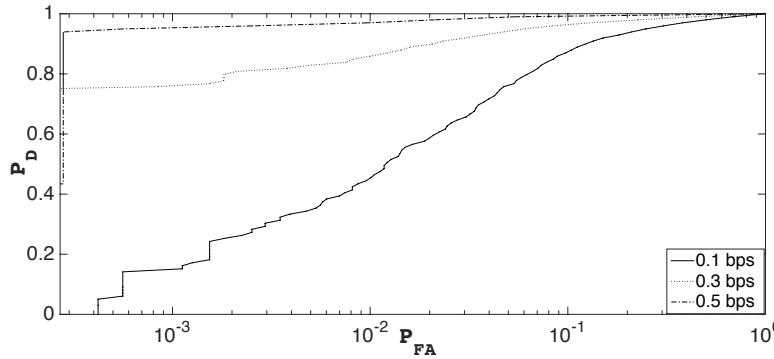


Figure 6. ROC Semilog curves for Salinas-A dataset at 0.1, 0.3 and 0.5 bps by using OBD-BCS

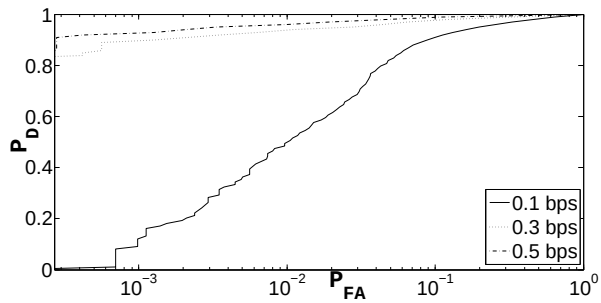


Figure 7. ROC Semilog curves for Salinas-A dataset at 0.1, 0.3 and 0.5 bps by using LASSO algorithm

The PSNR values of each sparse representation algorithms are also presented in Table 3 for 0.1, 0.3 and 0.5 bit rates in such a way to further strengthen the anomaly detection results obtained in Figure 10. The two highest PSNR values are marked in boldface.

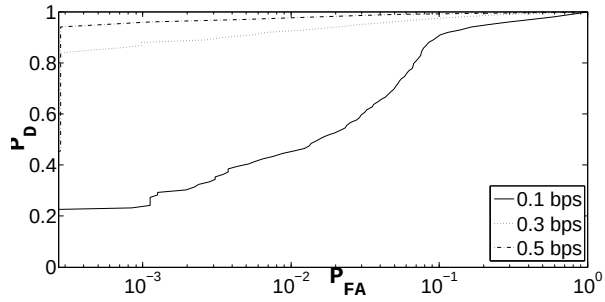


Figure 8. ROC Semilog curves for Salinas-A dataset at 0.1, 0.3 and 0.5 bps by using BP by using dual active set algorithm

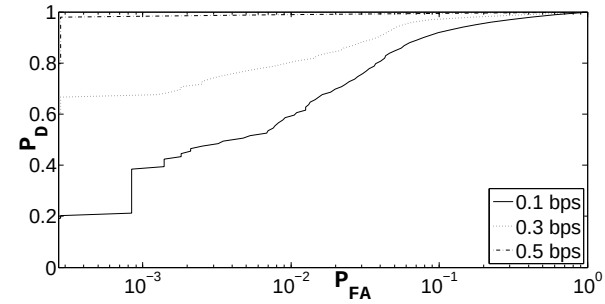


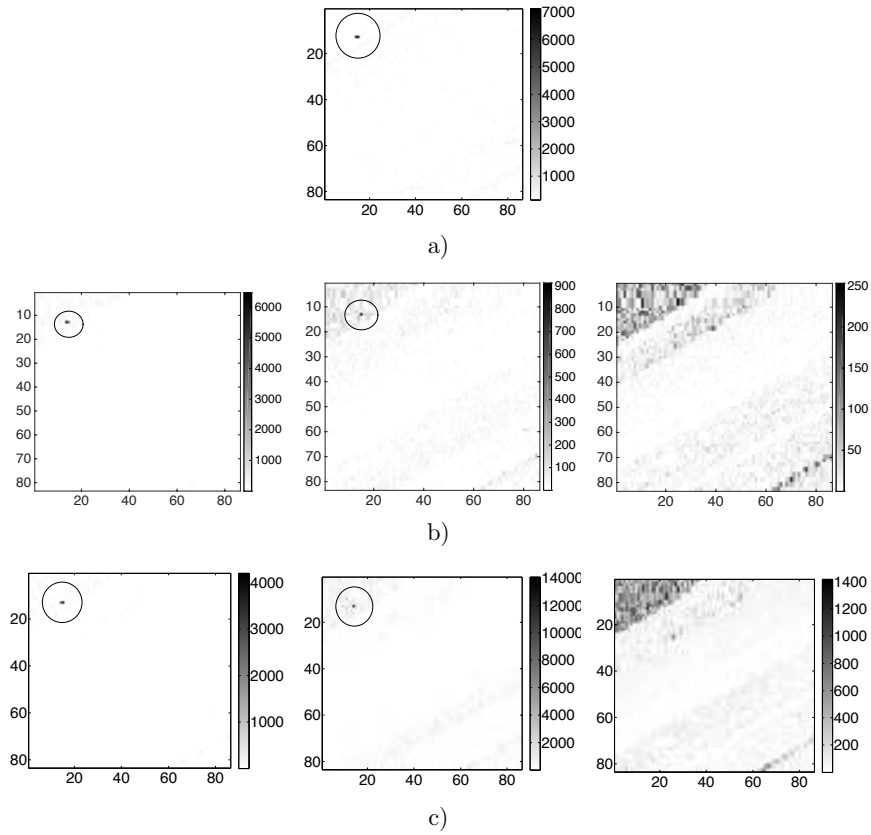
Figure 9. ROC Semilog curves for Salinas-A dataset at 0.1, 0.3 and 0.5 bps by using SpaRSA

Anomaly detection performances of different sparse representation algorithms can be assessed using receiver operating characteristic (ROC) curves. The ROC curves plot detection probability versus false alarm probability. ROC curves are plotted with a logarithmic x axis for better illustration.

Figure 6 shows the ROC Semilog curves of OBD-BCS algorithm at 0.1 bps, 0.3 bps and 0.5 bps rates when Salinas-A hyperspectral dataset is used. The probability of detection is denoted by PD and the probability of false alarm is denoted by PFA. Anomaly detection result at 0.5 bps rate is significantly better than those of the 0.3 bps and 0.1 bps levels.

The ROC Semilog curves of BP by using dual active set algorithm at 0.1, 0.3 and 0.5 bps bit rates are depicted in Figure 8. For the Salinas-A dataset, the ROC Semilog curves of SpaRSA and LASSO algorithm at various bit rates are illustrated in Figures 9 and 7, respectively.

In order to further evaluate the ROC curves, the area under curve (AUC) is employed as a performance metric that can be obtained by calculating the area



under the ROC curve. Calculated AUC values are presented in Table 3. In Table 3 for Salinas-A dataset, it can be seen that the best result is from OBD-BCS at 0.5 bps bit rate which is 0.9945.

Results in Table 3 and Figures 6–7 demonstrate that the detection performance of OBD-BCS algorithm is better than that of the other algorithms for the case where the bit rate is high such as 0.5 bps. The illustrations in Figure 10 also suggest that the detection performance of OBD-BCS algorithm is the best of all at 0.5 bps bit rate.

According to the values in Table 3, OBD-BCS algorithm is among the best two algorithms in terms of PSNR values at 0.5, 0.3 and 0.1 bps rates for Low-Altitude dataset. Particularly at 0.5 bps level, OBD-BCS has PSNR value of 73.56 which is the highest. The superiority of OBD-BCS algorithm at 0.5 bps rate is supported by the results in Table 3 for Low-Altitude dataset. At 0.5 bps, OBD-BCS algorithm has the highest AUC value which is 0.9943.

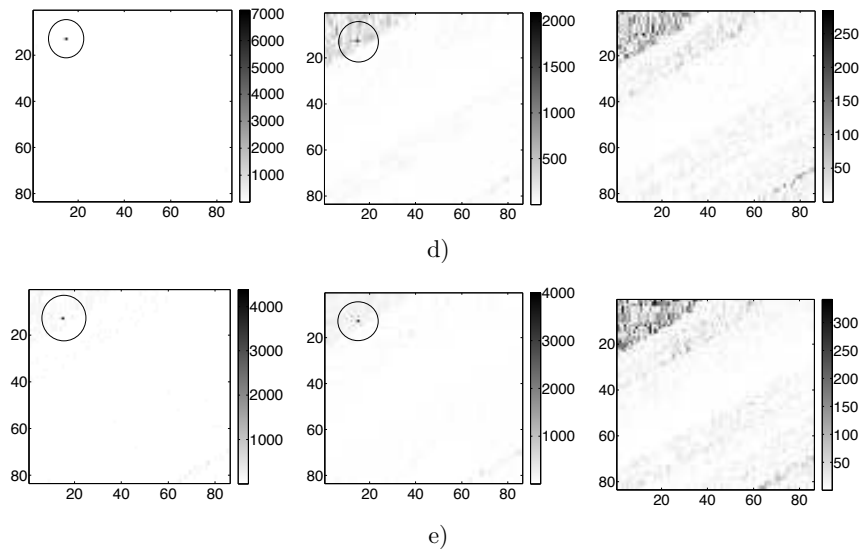


Figure 10. RX anomaly detection results of the Salinas-A hyperspectral image: a) original image, b) OBD-BCS with 0.5 bps, 0.3 bps and 0.1 bps, c) BP with 0.5 bps, 0.3 bps, and 0.1 bps, d) SpaRSA with 0.5 bps, 0.3 bps, and 0.1 bps, e) LASSO with 0.5 bps, 0.3 bps, and 0.1 bps

4 CONCLUSION

Sparse representation algorithms from many different categories are applied for the purpose of hyperspectral image compression based on online dictionary learning. The hyperspectral compression performance of these sparse representation algorithms are analyzed by further analyzing the OBD-BCS algorithm. By analyzing the results of all the datasets, the OBD-BCS algorithm shows the best compression performance at high compression ratios. At a 0.5 bps ratio, it involves among the best three algorithms at most for all the datasets. Other algorithms with good compression performances at high ratios are BP by using dual active set, the LASSO by using ADMM, and the SpaRSA algorithms.

According to the anomaly detection results, compressed image at bit rates of 0.5 bps or higher can be used as an estimate of the original hyperspectral image. Anomaly detection or similar real-world applications can be applied on the compressed hyperspectral image instead of the original one. Anomaly detection results further prove that the OBD-BCS algorithm has a better information preservation performance than that of the other algorithms as the bit rate gets higher.

	Salinas-A							
	Sparse Respresentation Algorithms							
BPS	BP		OBD-BCS		LASSO		SpaRSA	
	PSNR	AUC	PSNR	AUC	PSNR	AUC	PSNR	AUC
0.1	36.62	0.96	**36.67**	0.9464	**36.65**	0.9741	36.58	0.9424
0.3	41.54	0.9934	**41.89**	0.9929	41.16	0.9799	**42.61**	0.992
0.5	43.95	0.9943	**43.98**	0.9945	43.74	0.9928	**43.96**	0.9943
	Low Altitude							
	Sparse Respresentation Algorithms							
BPS	BP		OBD-BCS		LASSO		SpaRSA	
	PSNR	AUC	PSNR	AUC	PSNR	AUC	PSNR	AUC
0.1	**59.96**	0.9917	**60**	0.9887	59.59	0.9906	59.88	0.9896
0.3	**70.16**	0.9932	**69.99**	0.9936	68.85	0.9932	69.78	0.9914
0.5	73.24	0.9942	**73.56**	0.9943	**73.52**	0.9931	72.82	0.9937

Table 3. PSNR values of sparse representation algorithms

Acknowledgements

The authors would like to thank the anonymous reviewers for their helpful and constructive comments that greatly contributed to improving the final version of the paper. The authors would also like to thank Prof. Dr. Halil T. Eyyuboğlu for useful suggestions and comments. This research was partially supported by the Turkish Scientific and Technical Research Council.

REFERENCES

[1] Hyperspectral Remote Sensing Scenes. `http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes`, accessed March 2018.

[2] Aharon, M.—Elad, M.: Sparse and Redundant Modeling of Image Content Using an Image-Signature-Dictionary. SIAM Journal on Imaging Sciences, Vol. 1, 2008, No. 3, pp. 228–247, doi: 10.1137/07070156X.

[3] Asif, M. S.—Romberg, J.: Dynamic Updating for ℓ_1 Minimization. IEEE Journal of Selected Topics in Signal Processing, Vol. 4, 2010, No. 2, pp. 421–434, doi: 10.1109/JSTSP.2009.2039174.

[4] Baraniuk, R. G.—Cevher, V.—Duarte, M. F.—Hegde, C.: Model-Based Compressive Sensing. IEEE Transactions on Information Theory, Vol. 56, 2010, No. 4, pp. 1982–2001, doi: 10.1109/TIT.2010.2040894.

[5] Boyd, S.—Parikh, N.—Chu, E.—Peleato, B.—Eckstein, J.: Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. Foundations and Trends® in Machine Learning, Vol. 3, 2011, No. 1, pp. 1–122, doi: 10.1561/2200000016.

[6] Charles, A. S.—Olshausen, B. A.—Rozell, C. J.: Learning Sparse Codes for Hyperspectral Imagery. IEEE Journal of Selected Topics in Signal Processing, Vol. 5, 2011, No. 5, pp. 963–978, doi: 10.1109/jstsp.2011.2149497.

[7] Chen, G.—Needell, D.: Compressed Sensing and Dictionary Learning. Proceedings of Symposia in Applied Mathematics, Vol. 73, 2016, pp. 210–241, doi: 10.1090/psapm/073.

[8] Donoho, D. L.—Tsaig, Y.: Fast Solution of ℓ_1-Norm Minimization Problems when the Solution May Be Sparse. Online, Stanford University, Stanford, CA, 2006, `https://statistics.stanford.edu/sites/g/files/sbiybj6031/f/2006-18.pdf`.

[9] Dragotti, P. L.—Poggi, G.—Ragozini, A. R. P: Compression of Multispectral Images by Three-Dimensional SPIHT Algorithm. IEEE Transactions on Geoscience and Remote Sensing, Vol. 38, 2000, No. 1, pp. 416–428, doi: 10.1109/36.823937.

[10] Elad, M.: Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing. Springer, New York, 2010, doi: 10.1007/978-1-4419-7011-4.

[11] Fowler, J. E.: Compressive-Projection Principal Component Analysis. IEEE Transactions on Image Processing, Vol. 18, 2009, No. 10, pp. 2230–2242, doi: 10.1109/tip.2009.2025089.

[12] Friedlander, M. P.—Saunders, M. A.: A Dual Active-Set Quadratic Programming Method for Finding Sparse Least-Squares Solutions. Online, University of British Columbia, BC, Canada, 2012.

[13] Fu, W. J.: Penalized Regressions the Bridge versus the LASSO. Journal of Computational and Graphical Statistics, Vol. 7, 1998, No. 3, pp. 397–416, doi: 10.1080/10618600.1998.10474784.

[14] Gleichman, S.—Eldar, Y. C.: Blind Compressed Sensing. IEEE Transactions on Information Theory, Vol. 57, 2011, No. 10, pp. 6958–6975, doi: 10.1109/tit.2011.2165821.

[15] Gong, P.—Zhang, C.—Lu, Z.—Huang, J. Z.—Ye, J.: A General Iterative Shrinkage and Thresholding Algorithm for Non-Convex Regularized Optimization Problems. Proceedings of the 30th International Conference on Machine Learning (ICML '13), Vol. 28, 2013, JMLR, pp. 37–45.

[16] Gorodnitsky, I. F.—Rao, B. D.: Sparse Signal Reconstruction from Limited Data Using FOCUSS: A Weighted Minimum Norm Algorithm. IEEE Transactions on Signal Processing, Vol. 45, 1997, No. 3, pp. 600–616, doi: 10.1109/78.558475.

[17] Hou, Y.—Zhang, Y.: Effective Hyperspectral Image Block Compressed Sensing Using Thress-Dimensional Wavelet Transform. 2014 IEEE Geoscience and Remote Sensing Symposium (IGARSS), 2014, pp. 2973–2976, doi: 10.1109/IGARSS.2014.6947101.

[18] Huo, C.—Zhang, R.—Yin, D.—Wu, Q.—Xu, D.: Hyperspectral Data Compression Using Sparse Representation. 2012 4th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPER), IEEE, 2012, pp. 1–4, doi: 10.1109/WHISPERS.2012.6874259.

[19] Kiely, A. B.—Klimesh, M. A.: Exploiting Calibration-Induced Artifacts in Lossless Compression of Hyperspectral Imagery. IEEE Transactions on Geoscience and Remote Sensing, Vol. 47, 2009, No. 8, pp. 2672–2678, doi: 10.1109/tgrs.2009.2015291.

[20] Kim, S. J.—Koh, K.—Lustig, M.—Boyd, S.—Gorinevsky, D.: An Interior-Point Method for Large-Scaleℓ_1-Regularized Least Squares. IEEE Journal of Selected Topics in Signal Processing, Vol. 1, 2007, No. 4, pp. 606–617, doi: 10.1109/JSTSP.2007.910971.

[21] Mairal, J.—Bach, F.—Ponce, J.—Sapiro, G.: Online Learning for Matrix Factorization and Sparse Coding. Journal of Machine Learning Research, Vol. 11, 2010, pp. 19–60, doi: 10.1145/1553374.1553463.

[22] Mallat, S. G.—Zhang, Z.: Matching Pursuits with Time-Frequency Dictionaries. IEEE Transactions on Signal Processing, Vol. 41, 1993, No. 12, pp. 3397–3415, doi: 10.1109/78.258082.

[23] Martin, G.—Bioucas-Dias, J.—Plaza. A.: B-HYCA: Blind Hyperspectral Compressive Sensing. 2015 IEEE International Geoscience and Remote Sensing Symposim (IGARSS'15), Milan, Italy, July, 2015, doi: 10.1109/IGARSS.2015.7326410.

[24] Mielikainen, J.—Toivanen, P.: Lossless Compression of Hyperspectral Images Using a Quantized Index to Lookup Tables. IEEE Geoscience Remote Sensing Letters, Vol. 5, 2008, No. 3, pp. 474–478, doi: 10.1109/LGRS.2008.917598.

[25] Needell, D.—Tropp, J. A.: CoSaMP: Iterative Signal Recovery from Incomplete and Inaccurate Samples. Applied and Computational Harmonic Analysis, Vol. 26, 2009, No. 3, pp. 301–321, doi: 10.1016/j.acha.2008.07.002.

[26] Olshausen, B. A.—Field, D. J.: Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1? Vision Research, Vol. 37, 1997, No. 23, pp. 3311–3325, doi: 10.1016/S0042-6989(97)00169-7.

[27] Du, P.—Xue, Z.—Li, J.—Plaza, A.: Learning Discriminative Sparse Representations for Hyperspectral Image Classification. IEEE Journal of Selected Topics in Signal Processing, Vol. 9, 2015, No. 6, pp. 1089–1104, doi: 10.1109/JSTSP.2015.2423260.

[28] Du, Q.—Fowler, J. E.: Hyperspectral Image Compression Using JPEG2000 and Principal Component Analysis. IEEE Geoscience and Remote Sensing Letters, Vol. 4, 2007, No. 2, pp. 201–205, doi: 10.1109/LGRS.2006.888109.

[29] Rajwade, A.—Kittle, D.—Tsai, T.-H.—Brady, D.—Carin, L.: Coded Hyperspectral Imaging and Blind Compressive Sensing. SIAM Journal on Imaging Sciences, 2013, pp. 782–812, doi: 10.1137/120875302.

[30] Reed, S. I.—Yu, X.: Adaptive Multiple-Band CFAR Detection of an Optical Pattern with Unknown Spectral Distribution. IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 38, 1990, No. 10, pp. 1760–1770, doi: 10.1109/29.60107.

[31] Ricci, M.—Magli, E.: Predictor Analysis for Onboard Lossy Predictive Compression of Multispectral and Hyperspectral Images. Journal of Applied Remote Sensing, Vol. 7, 2013, No. 1, Art. No. 074591, doi: 10.1117/1.JRS.7.074591.

[32] Ryan, M. J.—Arnold, J. F.: The Lossless Compression of AVIRIS Images by Vector Quantization. IEEE Transactions on Geoscience and Remote Sensing, Vol. 35, 1997, No. 5, pp. 546–550, doi: 10.1109/36.581964.

[33] Song, H.—Wang, G.: Sparse Signal Recovery via ECME Tresholding Pursuits. Mathematical Problems in Engineering, Vol. 2012, 2012, Art. No. 478931, 22 pp., doi: 10.1155/2012/478931.

[34] Tipping, M. E.: Sparse Bayesian Learning and the Relevance Vector Machine. Journal of Machine Learning Research, Vol. 1, 2001, pp. 211–244.

[35] Tropp, J. A.—Gilbert, A. C.: Signal Recovery from Random Measurements via Orthogonal Matching Pursuit. IEEE Transactions on Information Theory, Vol. 53, 2007, No. 12, pp. 4655–4666, doi: 10.1109/tit.2007.909108.

[36] Ülkü, İ.—Kizgut, E.: Hyperspectral Compressive Sensing Based on Online Dictionary Learning. Imaging and Applied Optics 2017, Optical Society of America, 2017, Art. No. ITh4E.1, doi: 10.1364/ISA.2017.ITh4E.1.

[37] Ülkü, İ.—Kizgut, E.: Large-Scale Hyperspectral Image Compression via Sparse Representations Based on Online Learning. International Journal of Applied Mathematics and Computer Science, Vol. 28, 2018, No. 1, pp. 197–207, doi: 10.2478/amcs-2018-0015.

[38] Ülkü, İ.—Töreyin, B. U.: Sparse Representations for Online-Learning-Based Hyperspectral Image Compression. Applied Optics, Vol. 54, 2015, No. 29, pp. 8625–8631, doi: 10.1364/AO.54.008625.

[39] Wang, J.—Kwon, S.—Shim, B.: Generalized Orthogonal Matching Pursuit. IEEE Transactions on Signal Processing, Vol. 60, 2012, No. 12, pp. 6202–6216, doi: 10.1109/TSP.2012.2218810.

[40] Wang, Z.—Nasrabadi, N. M.—Huang, T. S.: Spatial-Spectral Classification of Hyperspectral Images Using Discriminative Dictionary Designed by Learning Vector Quantization. IEEE Transactions on Geoscience and Remote Sensing, Vol. 52, 2014, No. 8, pp. 4808–4822, doi: 10.1109/tgrs.2013.2285049.

[41] Wang, H.—Celik, T.: Sparse Representation-Based Hyperspectral Image Classification. Signal, Image and Video Processing, Vol. 12, 2018, No. 5, pp. 1009–1017, doi: 10.1007/s11760-018-1249-1.

[42] Wright, J.—Yang, A. Y.—Ganesh, A.—Sastry, S. S.: Robust Face Recognition via Sparse Representation. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 31, 2009, No. 2, pp. 210–227, doi: 10.1109/tpami.2008.79.

[43] Yang, A. Y.—Sastry, S. S.—Ganesh, A.—Ma, Y.: Fast ℓ_1-Minimization Algorithms and an Application in Robust Face Recognition: A Review. 2010 IEEE International Conference on Image Processing (ICIP), 2010, pp. 1849–1852, doi: 10.1109/ICIP.2010.5651522.

[44] Yang, J.—Peng, Y.—Xu, W.—Dai, Q.: Ways to Sparse Representation: An Overview. Science in China Series F: Information Sciences, Vol. 52, 2009, No. 4, pp. 695–703, doi: 10.1007/s11432-009-0045-5.

[45] Zayyani, H.—Babaie-Zadeh, M.—Jutten, C: An Iterative Bayesian Algorithm for Sparse Component Analysis in Presence of Noise. IEEE Transactions on Signal Processing, Vol. 57, 2009, No. 11, pp. 4378–4390, doi: 10.1109/tsp.2009.2025154.

[46] Zayyani, H.—Korki, M.—Marvasti, F: Dictionary Learning for Blind One Bit Compressed Sensing. IEEE Signal Processing Letters, Vol. 23, 2016, No. 2, pp. 187–191, doi: 10.1109/lsp.2015.2503804.

[47] Zhang, L.—Wei, W.—Tian, C.—Li, F.—Zhang, Y.: Exploring Structured Sparsity by a Reweighted Laplace Prior for Hyperspectral Compressive Sensing. IEEE Transactions on Image Processing, Vol. 25, 2016, No. 10, pp. 4974–4988, doi: 10.1109/tip.2016.2598652.

[48] Zhang, Z.—Xu, Y.—Yang, J.—Li, X.—Zhang, D.: A Survey of Sparse Representation: Algorithms and Applications. IEEE Access, Vol. 3, 2015, pp. 490–530, doi: 10.1109/ACCESS.2015.2430359.

İrem Ülkü received her M.Sc. degree in electrical and electronics engineering from Middle East Technical University in 2013, and Ph.D. degree in electronics and communication engineering from Çankaya University in 2017. Her research interests include hyperspectral image processing, dictionary learning, compressive sensing and sparse coding.

Ersin Kizgut received his Ph.D. degree in mathematics from Middle East Techical University in 2016. He is recently a postdoctoral researcher at Polytechnic University of Valencia. His research interests include applied mathematics, computer science, operator theory, complex analysis, and topological vector spaces.