Adaptive Spectral–Spatial Compression of Hyperspectral Image With Sparse Representation

Wei Fu, *Student Member, IEEE*, Shutao Li, *Senior Member, IEEE*, Leyuan Fang, *Member, IEEE*, and Jón Atli Benediktsson, *Fellow, IEEE*

***Abstract*— Sparse representation (SR) can transform spectral signatures of hyperspectral pixels into sparse coefficients with very few nonzero entries, which can efficiently be used for compression. In this paper, a spectral–spatial adaptive SR (SSASR) method is proposed for hyperspectral image (HSI) compression by taking advantage of the spectral and spatial information of HSIs. First, we construct superpixels, i.e., homogeneous regions with adaptive sizes and shapes, to describe HSIs. Since homogeneous regions usually consist of similar pixels, pixels within each superpixel will be similar and share similar spectral signatures. Then, the spectral signatures of each superpixel can be simultaneously coded in the SR model to exploit their joint sparsity. Since different superpixels generally have different performances of SR, their rate–distortion performances in the sparse coding will be different. To achieve the best possible overall rate–distortion performance, an adaptive coding scheme is introduced to adaptively assign distortions to superpixels. Finally, the obtained sparse coefficients are quantized and entropy coded and constitute the final bitstream with the coded superpixel map. The experimental results over several HSIs show that the proposed SSASR method outperforms some state-of-the-art HSI compression methods in terms of the rate–distortion and spectral fidelity performances.**

***Index Terms*— Compression, hyperspectral image (HSI), sparse representation (SR), superpixel.**

I. Introduction

HYPERSPECTRAL images (HSIs) generally consist of more than one hundred spectral bands. Since HSIs can provide abundant spectral and spatial information [1]–[4], research studies on HSI processing have become more and more popular, e.g., classification of ground materials [5], [6], target detection [7], [8], and unmixing [9], [10]. Although image analysis can benefit from the abundant data of HSIs, the vast data may create a heavy burden for the storage and transmission of HSIs. Thus, how to effectively compress HSIs is becoming an important problem among HSI applications.

Manuscript received May 6, 2016; revised August 22, 2016; accepted September 21, 2016. This work was supported in part by the National Natural Science Fund of China for Distinguished Young Scholars under Grant 61325007 and in part by the National Natural Science Fund of China for International Cooperation and Exchanges under Grant 61520106001.

W. Fu, S. Li, and L. Fang are with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China (e-mail: fuweihunandaxue@gmail.com; shutao_li@hnu.edu.cn; fangleyuan@gmail.com).

J. A. Benediktsson is with the Faculty of Electrical and Computer Engineering, University of Iceland, 101 Reykjavik, Iceland (e-mail: benedikt@hi.is).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2016.2613848

There are two types of HSI compression methods, lossless compression and lossy compression. For lossless compression, the key is to eliminate redundancy of the data without loss of information [11]. However, the best compression ratios (CRs) achieved by lossless methods are of the order of 3:1 [12]. In contrast, lossy compression [13] loses some information to achieve higher CRs than lossless compression. Therefore, lossy compression is a promising research topic in terms of high CRs, and is the focus of this paper.

During the past decades, frequency domain-based compression approaches have been successively used in 2-D digital images and show promising performance, e.g., the Joint Photographic Experts Group (JPEG) standard [14] and the JPEG2000 standard [15]. JPEG transforms images by discrete cosine transform (DCT). JPEG2000 mainly uses discrete wavelet transform (DWT) instead of DCT and can usually achieve higher CRs than the JPEG. In addition, considering the superiority of DWT in digital image analysis, Pearlman *et al.* [16] propose a set-partitioned embedded block (SPECK) compression method. In detail, they transform the original image into blocks (i.e., wavelet coefficients) by the DWT and then recursively split the blocks into subblocks to find significant coefficients. The SPECK has been extended to a 3-D version of SPECK (i.e., the 3-D SPECK) for HSI compression in [17]. The 3-D-SPECK separately applies a 3-D DWT to small portions of hyperspectral bands and then extends the method of splitting blocks and finding significant coefficients to the 3-D case. In addition, the extension of JPEG2000 to HSI compression has drawn wide attention. An idea of using JPEG2000 for HSI compression is that each spectral band is separately compressed by JPEG2000. However, the band-by-band JPEG2000 method exploits only the spatial redundancy in each spectral band without consideration of the spectral redundancy in HSIs. In order to achieve better compression performance, Rucker *et al.* [18] propose the DWT + JPEG2000 method, which first applies a wavelet-based decorrelating transform in the spectral direction and then uses JPEG2000 to compress the transformed data in the spatial direction. Du *et al.* [19], [20] modify the DWT + JPEG2000 method [18] using principal component analysis (PCA) algorithm [21] for spectral decorrelation and introduce the PCA + JPEG2000 method. The PCA + JPEG2000 method is widely used and can provide superior compression performance over other HSI compression methods, including the DWT + JPEG2000 method.

Furthermore, the PCA + JPEG2000 method has been shown to do well in preserving useful information in HSIs for unmixing, anomaly detection, and classification.

Recently, sparse representation (SR) theory has been demonstrated to be a powerful tool in digital signal processing [22], [23]. Generally, natural signals are considered to be sparse and can be represented by sparse coefficients (vectors) over a dictionary in the SR model. The sparse coefficients contain only a few nonzero entries, so the information of original signals is naturally compressed into the few nonzero values of sparse coefficients. The SR compression method has been applied to 2-D image compression in [24] by sparsely representing nonoverlapping image patches over an offline dictionary. The obtained sparse coefficients are quantized to integers and then transformed into a bitstream by the entropy coding technique.

The SR compression method can be extended to compression of HSIs, since spectral signatures of HSI pixels can be sparsely represented according to [25]–[27]. The basic method is to separately compress each HSI pixel by sparsely representing the pixel's spectral signature over a suitable dictionary. Unlike the SR-based 2-D image compression [24], which learns an offline dictionary (data-independent) for various images, the HSI compression requires a data-dependent dictionary due to the large variety of HSIs. Subdictionary-based dictionary construction method [28], which has a lower complexity and leads to sparser representation, can be used for SR compression. The above SR method for HSIs compression can be called the pixelwise SR (PWSR) method. However, the PWSR method ignores the spatial correlation of HSI pixels, i.e., neighboring pixels usually represent the same material and share the similar spectral signatures. In [29] and [30], it is shown that simultaneous SR of correlative signals (i.e., similar vectors) with the row-sparse coefficients can recover signals with higher sparsity levels (i.e., sparser coefficients) compared with the SR of the single signal. This is because the former incorporates contextual information of signals into the model. This is meaningful for compression since sparser coefficients mean less nonzero data, and thus leads to a higher CR. Therefore, the spatial correlation of HSI pixels is important for the SR compression, since neighboring HSI pixels may probably be correlative signals and can be represented by sparser coefficients to improve the compression performance.

Different HSI pixels share different spatial correlations, which are determined by the spatial structure of HSIs [31]. In order to exactly extract the spatial correlations of HSIs, superpixels [32]–[36], which are defined as local homogeneous regions, have been introduced into HSI processing. In this paper, we couple the SR with the superpixel technique and propose a spectral–spatial adaptive SR (SSASR) method to effectively take advantage of the spatial correlation for compression. Applying the superpixel technique to the HSI compression has the following advantages.

1) Superpixels have adaptive sizes and shapes according to the structural information of HSIs and are generally homogeneous. Therefore, the HSI pixels in the superpixel share the similar spectral signatures and thus are correlative. Simultaneous SR can exploit the contextual information of HSI pixels in superpixels and achieve better compression performance.
2) Some superpixels may belong to regions of interest that should be compressed with lower CRs. Some superpixels may belong to backgrounds that can be compressed with higher CRs. Since superpixels are independent of each other, their compression tasks can also be independent. As a consequence, it is very convenient to make custom compression for different superpixels according to the interest.
3) The number of superpixels is much less than the number of pixels, so superpixel-based processes may be more efficient than pixel-based processes. In addition, the nonzero positions of sparse coefficients are recorded for each superpixel rather than for each pixel in the SR. That can significantly reduce the cost of encoding the sparse coefficients and thus improve the CRs.

With these advantages, the proposed SSASR method first constructs superpixels to effectively convey the spectral and spatial information of HSIs instead of adopting HSI pixels. Next, superpixels are represented in the SR model with sparse coefficients, which is called sparse coding. Since superpixels are of different sizes and shapes and represent different image details, they generally have different performances of SR. As a result, the rate–distortion performances of different superpixels in the sparse coding will be different. To achieve the possible best overall rate–distortion performance in sparse coding, an adaptive coding scheme is introduced by adaptively assigning distortions to superpixels. Except for the sparse coefficients, the label map of superpixels, which records the results of superpixel segmentation and the dictionary, must be encoded to bits and involved in the final bitstream. The label map occupies a high proportion of bits in the final bitstream, especially when the CRs are high (i.e., the total number of bits is limited). To better compress the superpixel map, we relabel the superpixels to improve the compression of entropy coding. Finally, the nonzero positions and the quantized nonzero values of the sparse coefficients are entropy coded to bits. The uniform quantizer [37] and Huffman coding [38] are used for quantization and entropy coding, respectively. The experimental results demonstrate the superiority of the proposed SSASR method over the PWSR method and some state-of-the-art HSI compression methods in terms of not only the rate–distortion performance but also spectral fidelity.

The remainder of this paper is organized as follows. The SSASR compression method for HSIs is proposed in Section II. Section III shows the experimental results and discussions. Finally, the results of this paper are summarized in Section IV.

II. Spectral–Spatial Adaptive Sparse Representation Compression

The proposed SSASR method first constructs superpixels to exploit the spatial–spectral information of HSIs. Then, superpixels are coded to sparse coefficients in the adaptive sparse coding scheme. Moreover, the label map of superpixels is encoded to reduce the abundance. Finally, the sparse

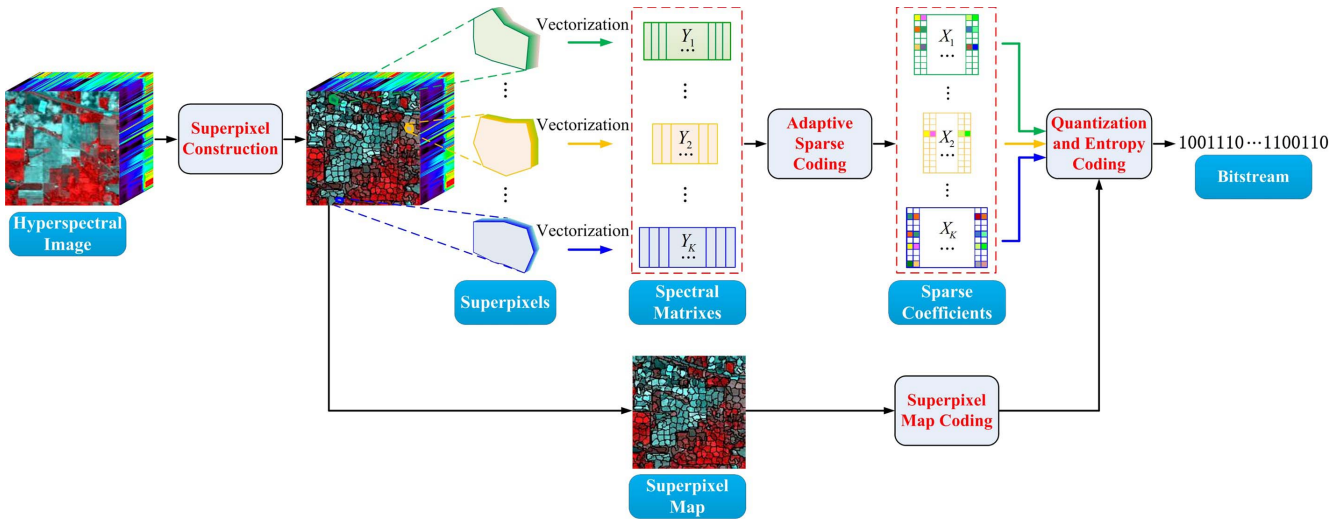


Fig. 1. Proposed SSASR compression method for HSIs.

coefficients are quantized and entropy coded to bitstream. The proposed compression method is shown in Fig. 1.

A. Construction of Superpixels

Superpixels are defined as a set of nonoverlapping regions [36], which can be represented by

$$\forall i, j \in K \quad \begin{cases} \mathbf{Y}_i \cap \mathbf{Y}_j = \emptyset \\ \mathbf{Y}_i \cup \mathbf{Y}_j = \mathbf{Y} \end{cases} \tag{1}$$

where $\mathbf{Y}_k = \{\mathbf{y}_{k,1}\ \mathbf{y}_{k,2}\ \ldots\ \mathbf{y}_{k,n_k}\}$ $(\sum_{k=1}^{K} n_k = mn)$ denotes the kth superpixel with n_k pixels and K is the total number of superpixels in image $\mathbf{Y}$. The localized K-means clustering scheme [32], which can be conveniently used for various applications by defining different distance measurements between samples, is used in this paper. With the purpose of clustering correlative HSI pixels that share similar spectral signatures, the distance measurement between HSI pixels is defined based on the spectral similarity by

$$d(\mathbf{y}_a, \mathbf{y}_b) = \left\| \frac{\mathbf{y}_a}{\|\mathbf{y}_a\|_2} - \frac{\mathbf{y}_b}{\|\mathbf{y}_b\|_2} \right\|_2 \tag{2}$$

where $\mathbf{y}_a$ and $\mathbf{y}_b$ are the spectral signatures of two arbitrary HSI pixels and $d(\mathbf{y}_a, \mathbf{y}_b)$ is the distance between $\mathbf{y}_a$ and $\mathbf{y}_b$. The distance in (2) mainly measures the similarity of spectral shapes since the normalization of the spectral signatures in (2) can reduce the effect of spectral brightness. After the K initial clustering centers are uniformly chosen in the spatial domain of HSIs, similar HSI pixels are recursively searched around each clustering center within a fixed size window. The implementation of the superpixels construction is shown in Algorithm 1.

B. Adaptive Sparse Coding of Superpixels

It has been proved that simultaneous representation of correlative HSI pixels, which are usually neighboring and share similar spectral signatures, can improve the performance of SR [36], [39], [40]. Superpixels can cluster correlative HSI pixels, and so spectral signatures of each superpixel can be simultaneously represented in the SR model. Let $\mathbf{Y}_k$ denote the spectral signature matrix of the kth superpixel. The SR of the kth superpixel can be represented as

$$\begin{aligned} &\min\ \|\mathbf{X_k}\|_{\text{row},0} \\ &\text{s.t. } \mathbf{Y}_k = \mathbf{D}\mathbf{X}_k + \mathbf{N} \end{aligned} \tag{3}$$

where $\mathbf{X}_k$ is the sparse coefficient matrix, $\|\mathbf{X_k}\|_{\text{row},0}$ denotes the number of nonzero rows in $\mathbf{X}_k$, $\mathbf{D}$ is the dictionary, and $\mathbf{N}$ is the possible noise. As dictionary $\mathbf{D}$ is available, the sparse coefficients $\mathbf{X}_k$ can be obtained by solving the relaxed optimization problem [41]

$$\begin{aligned} \hat{\mathbf{X}}_k &= \arg\min \|\ \mathbf{X}_k\ \|_{\text{row},0} \\ &\text{s.t. } \|\ \mathbf{Y}_k - \mathbf{D}\mathbf{X}_k\ \|_F^2 \le \delta_k \end{aligned} \tag{4}$$

Algorithm 1 Superpixel Construction

Input: 1). The hyperspectral image (size of $M \times N \times B$); 2). The number of the superpixels K and maximum number of iteration I;

Output: The label map $\mathbf{L}$ (size of $M \times N$) of superpixels;

Initialize label map of superpixels $\mathbf{L} = \mathbf{0}$ and the number of iteration $q = 0$;
Initialize K seed points $\{\hat{\mathbf{y}}_k^0\}_{k=1,2,\ldots,K}$;
Calculate window size of $2\sqrt{MN/K} \times 2\sqrt{MN/K}$;
while $q < I$ **do**
  for Each HSI pixel $\mathbf{y}$ **do**
    Search seeds $\{\hat{\mathbf{y}}_{k_v}^q\}_{(v\in V, k_v\in\{1,2,\ldots,K\})}$ around $\mathbf{y}$ within the window (V is the numberof searched seed points);
    Calculate distances $\{d(\mathbf{y}, \hat{\mathbf{y}}_{k_v}^q)\}_{v\in V}$ between $\mathbf{y}$ and V seed points by equation (2);
    $\mathbf{L}(\mathbf{y}) = \arg\min_{k_v}\{d(\mathbf{y}, \hat{\mathbf{y}}_{k_v})\}_{v\in V}$;
  end for
  Update the kth ($k = 1, 2, \ldots K$) seed point by averaging the spectral signatures and spatial positions of pixels whose labels in $\mathbf{L}$ are equal to k;
  $q \leftarrow q + 1$;
end while

where δ_k is the target error [42] for the kth superpixel to control the distortion of data. With the given error δ_k, the simultaneous orthogonal matching pursuit (SOMP) [43] method can be used to effectively solve (4). To efficiently and effectively construct dictionary $\mathbf{D}$, the PCA subdictionary-based method [28] is adopted in this paper. That is, we first sample a few correlative HSI pixels (usually are three pixels) from each superpixel, and then cluster those pixels by the K-means method. Finally, PCA is applied to each cluster and the projections over the first P principle components are used as atoms of subdictionary. All subdictionaries constitute the entire dictionary $\mathbf{D}$.

It is very convenient to make custom compression of interested superpixels by defining different target errors. However, in this paper, we focus on achieving the possible optimal overall rate–distortion performance, i.e., rate is minimized for a given overall data distortion (overall target error). Rate is quantified using the bits per pixel per band (bpppb), which is positively related to the number of nonzero sparse coefficients (more nonzero coefficients are needed to be encoded with more bits and thus lead to higher rates, and vice versa)

$$\text{Rate} \propto \sum_k n_k \|\mathbf{X}_k\|_{\text{row},0} \tag{5}$$

where n_k is the number of pixels in the kth superpixel and $\sum_k n_k \|\mathbf{X}_k\|_{\text{row},0}$ denotes the total number of nonzero entries in the sparse coefficients of the kth superpixel. Therefore, the minimizing rate is equal to minimizing $\sum_k n_k \|\mathbf{X}_k\|_{\text{row},0}$. Let $s_k = n_k \|\mathbf{X}_k\|_{\text{row},0}$ denote the number of nonzero coefficients for the kth superpixel. Let $\delta_k(s_k)$ represent the target error as a function of s_k, which is obtained by solving (4). The problem of minimizing rate for a given overall target distortion δ_t can be described as

$$\begin{aligned} &\min \sum_{k=1}^{K} s_k \\ &\text{s.t.} \left| \sum_{k=1}^{K} \delta_k(s_k) - \delta_t \right| < \xi \end{aligned} \tag{6}$$

where ξ denotes the tolerant error and is a very small constant. Generally, different superpixels have different performances of SR, which means that the curve of function $\delta_k(s_k)$ varies for different values of k. The examples of three superpixels (from the *Indian Pines* image) shown in Fig. 2 reveal this phenomenon. As a consequence, the problem in (6) is nondeterministic polynomial time hard. To approximately solve (6), a greedy algorithm-based sparse coding scheme is proposed. The sparse coding scheme iteratively encodes superpixels from coarse to fine. First, all superpixels are sparsely represented with a large target error. Then, the overall error $\sum_{k=1}^{K} \delta_k(s_k)$ is gradually decreased with the step length of $\frac{\delta_t}{I}$ as the number of iterations increases, where I is the maximum number of iterations. In each iteration, the superpixel, which has the smallest increase of $\sum_{k=1}^{K} s_k$ with per unit reduction of $\sum_{k=1}^{K} \delta_k(s_k)$, is greedily selected. In detail, the gain ratio of s_k for the kth superpixel is first calculated as

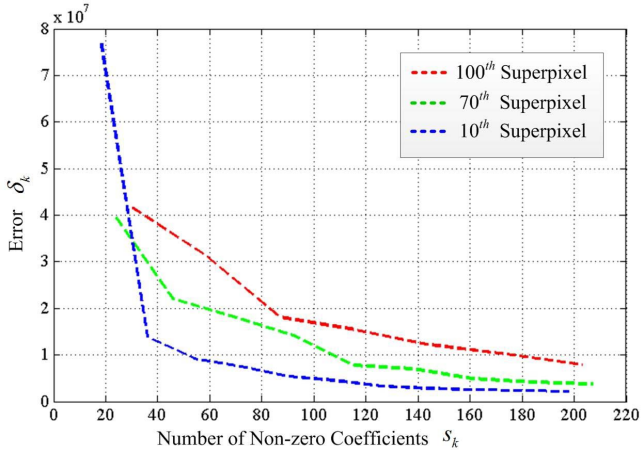


Fig. 2. Representation errors of SR for three superpixels (from the *Indian Pines* image) with different numbers of nonzero coefficients.

$$\frac{\Delta s_k^i}{\Delta \delta_k^i} = \frac{s_k^{i-1} - s_k^i}{\delta_k^{i-1} - \delta_k^i} \tag{7}$$

where i denotes the ith iteration. Then, the superpixel with the minimum gain ratio is chosen by

$$\hat{k}^i = \arg\min_k \frac{\Delta s_k^i}{\Delta \delta_k^i} \quad (k = 1, 2, \ldots, K). \tag{8}$$

This greedy criterion is a natural generalization of the standard greedy algorithm [44]. The chosen superpixel is further encoded with a smaller target error and is responsible for the decrease in the overall error. Finally, when the condition of $|\sum_{k=1}^{K} \delta_k(s_k) - \delta_t| < \xi$ is reached, the iteration stops and the results of the adaptive sparse coding are obtained. The implementation of the adaptive sparse coding is shown in Algorithm 2.

C. Coding of the Superpixel Map

As a recorder of the superpixel segmentation results, the superpixel map (label map) should be included in the final compression results (bitstream) of the HSI. The entropy coding (i.e., Huffman coding [45]) is used to compress and transform the superpixel map to bits without losing any information. However, the number of bits for encoding the superpixel map is large. Especially, when CR is very high, the available number of bits is limited. A large number of bits used for encoding the superpixel map can result in fewer bits used for sparse coding, and thus reduce the compression performance. Furthermore, as the number of superpixels increases, the superpixel map becomes more complex and the cost of encoding the map is increased. To effectively compress the superpixel map, we relabel the superpixels to improve the compression of entropy coding.

For the entropy coding, the minimum number of bits to compress the superpixel map $\mathbf{L}$ can be approximately measured based on the entropy of signals

$$H(\mathbf{L}) = -MN \sum_{k=1}^{K} \frac{n_k}{MN} \log_2 \frac{n_k}{MN} \tag{9}$$

Algorithm 2 Adaptive Sparse Coding

Input: 1). The superpixels $\{\mathbf{Y}_k\}_{k=1,2,\ldots,K}$; 2). The trained dictionary $\mathbf{D}$; 3). The overall target distortion δ_t;
Output: The sparse coefficients $\{\mathbf{X}_k\}_{k=1,2,\ldots,K}$;
1: Calculate step length δ_t/I;
2: Set $\delta_k^0 = 2\delta_t n_k/MN$, $\delta_k^1 = \delta_0 - \Delta\delta$, and obtain corresponding sparse coefficients $\mathbf{X}_k^0$, $\mathbf{X}_k^1$;
3: Calculate $s_k^0 = n_k\|\mathbf{X}_k^0\|_{row,0}$ and $s_k^1 = n_k\|\mathbf{X}_k^1\|_{row,0}$;
4: Calculate $\frac{\Delta s_k^1}{\Delta\delta_k^1}$ by (7);
5: $\mathbf{X}_k \leftarrow \mathbf{X}_k^0$ and $i \leftarrow 1$;
6: **while** $|\sum_k \delta_k^i - \delta_t| > \xi$ **do**
7: Obtain $\hat{k}^i$ by (8);
8: Solve $\mathbf{X}_{\hat{k}^i}^i = \arg\min\|\mathbf{X}_{\hat{k}^i}\|_{row,0}\ s.t.\ \|\mathbf{Y}_{\hat{k}^i} - \mathbf{D}\mathbf{X}_{\hat{k}^i}\|_F^2 < \delta_{\hat{k}^i}^i$;
9: Renew $\delta_{\hat{k}^i}^{i+1} = \delta_{\hat{k}^i}^i - \Delta\delta$, $s_{\hat{k}^i}^i = n_{\hat{k}^i}\|\mathbf{X}_{\hat{k}^i}^i\|_{row,0}$;
10: Solve $\mathbf{X}_{\hat{k}^i}^{i+1} = \arg\min\|\mathbf{X}_{\hat{k}^i}\|_{row,0}\ s.t.\ \|\mathbf{Y}_{\hat{k}^i} - \mathbf{D}\mathbf{X}_{\hat{k}^i}\|_F^2 < \delta_{\hat{k}^i}^{i+1}$;
11: Renew $s_{\hat{k}^i}^{i+1} = n_{\hat{k}^i}\|\mathbf{X}_{\hat{k}^i}^{i+1}\|_{row,0}$;
12: Calculate $\frac{\Delta s_{\hat{k}^i}^{i+1}}{\Delta\delta_{\hat{k}^i}^{i+1}}$ by (7);
13: $\mathbf{X}_{\hat{k}^i} \leftarrow \mathbf{X}_{\hat{k}^i}^i$ and then $i \leftarrow i+1$;
14: **end while**

where n_k is the number of pixels with the label of k. Since $H(\mathbf{L})$ is positively related to K and negatively related to n_k, we can reduce $H(\mathbf{L})$ by reducing K and increasing n_k. Relabeling superpixels is a good method to simultaneously reduce K and increase n_k. The rule is to assign the same label to the nonadjacent superpixels as much as possible and make sure that the adjacent superpixels are assigned different labels. As a consequence, many nonadjacent superpixels share the same label, which means that K is decreased and n_k is increased, and thus reduce the number of bits for encoding the superpixel map (i.e., improving the CRs of entropy coding). At the same time, the results of the superpixel segmentation are preserved without any losses. Fig. 3 demonstrates the results of encoding the relabeled map and the original map by Huffman coding. In Fig. 3, the test data is the Indian Pines image and the number of superpixels increases from 200 to 1100. As the number of superpixels increases, the bits of encoding the original map approximately increase from 150 000 to 220 000. However, the bits of encoding the relabeled map remain fairly stable around 60 000. Therefore, the proposed method can significantly reduce the cost of encoding the superpixel map and is robust for the number of superpixels. The implementation details of the relabeling for superpixels can be seen in Algorithm 3.

D. Quantization, Entropy Coding, and Decompression

The final compression results are in the form of bitstream, which consist of three parts: the nonzero values of sparse coefficients, the positions of nonzero rows in sparse coefficients, and the coded superpixel map. The classical Huffman coding [45] is used to transform the above data to bits.

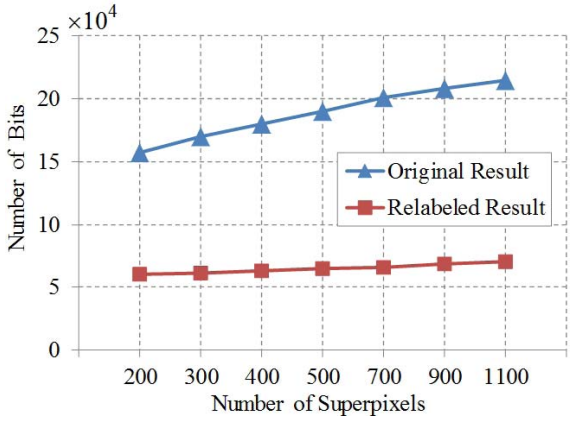


Fig. 3. Number of bits for coding the superpixel map by the original map and the relabeled map.

Algorithm 3 Superpixel Map Coding

Input: 1). The label map $\mathbf{L}$ of superpixels; 2). The candidate labels $\mathbf{L}_c = \{1, 2, \ldots, L\}$;
Output: The relabeled map $\hat{\mathbf{L}}$ of superpixels;
1: Extract the coordinates Ω_k of the kth superpixel;
2: Initialize relabeled map of superpixels $\hat{\mathbf{L}} = \mathbf{0}$;
3: Select the initial superpixel $k = 1$ and assign the initial label by $\hat{\mathbf{L}}(\Omega_1) = 1$;
4: **for** $k = 2 : K$ **do**
5: Search adjacent superpixels around the kth superpixel in $\hat{\mathbf{L}}$;
6: Record their labels in the set $\mathbf{L}_r$ by removing the repetitive labels;
7: Calculate the label of the kth superpixel by $\hat{\mathbf{L}}(\Omega_k) = \min\{\mathbf{L}_c - \mathbf{L}_c \cap \mathbf{L}_r\}$;
8: Clear $\mathbf{L}_r$;
9: **end for**

Algorithm 4 Decompression

Input: The bitstream of sparse coefficients, coded superpixel map, and dictionary $\mathbf{D}$;
Output: The reconstructed hyperspectral image $\hat{\mathbf{Y}}$;
1: Decode the bitstream by Huffman decoding;
2: Requantize and recover the sparse coefficients $\hat{\mathbf{X}}_k$ and dictionary $\mathbf{D}$;
3: Decode superpixel map and obtain the spatial coordinates Ω_k of the kth superpixel;
4: **for** $k = 1 : K$ **do**
5: Calculate the spectral signatures of the kth superpixel by: $\hat{\mathbf{Y}}(\Omega_k) = \mathbf{D} \cdot \hat{\mathbf{X}}_k$;
6: **end for**

Since the nonzero values of sparse coefficients are not integers, they should be quantized before the Huffman coding. In this paper, the uniform quantizer is used for quantization. Except for the inverse quantization and the Huffman decoding, the decompression of the proposed SSASR method mainly contains the SR reconstruction process, which is shown in Algorithm 4.

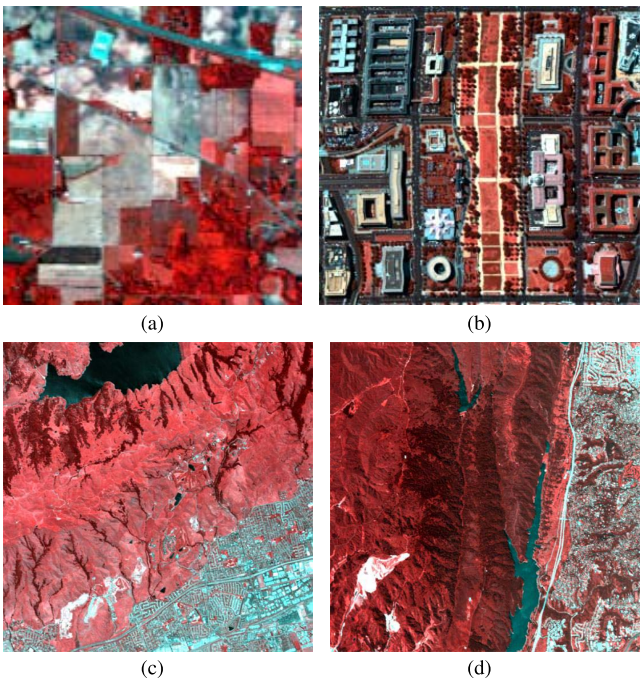


Fig. 4. False color composites of four HSIs. (a) *Indian Pines* image (145 x 145). (b) *Washington DC* image (280 x 307). (c) *Moffett* image (512 x 512). (d) *Jasper Ridge* image (512 x 512).

III. Experimental Results and Analysis

In this section, experimental data sets and setup are first introduced. Then, the compression performances of the proposed SSASR method are evaluated. Next, we analyze the effect of compression on HSI classification. Finally, the computational complexity of the proposed method is analyzed.

A. Experimental Data Sets and Setup

Four hyperspectral data sets are used: the *Indian Pines* image, the *Washington DC* image, the *Moffett* image, and the *Jasper Ridge* image. The *Indian Pines* image and *Washington DC* image are chosen due to the availability of labeled ground truth, which will be useful for analyzing the effect of compression on classification. The *Moffett* image and *Jasper Ridge* image are generally used for HSI compression [20]. The detailed descriptions of these HSIs are shown in Fig. 4.

1) Indian Pines Image: The image is collected by the Airborn/Visible Infrared Imageing Specrometer (AVIRIS) sensor over the agricultural Indian pine test site in Northwestern Indiana. This HSI consists of 224 spectral bands across the spectral range from 0.4 to 2.5 μm and each band contains 145×145 pixels with a spatial resolution of 20 m. The 24 water absorption and noise bands are removed in the experiment, i.e., bands 1, 33, 97, 107–111, 153–167, and 224 [19]. The false color composite of the *Indian Pines* image is shown in Fig. 4(a).

2) Washington DC Image: The image is acquired by the Hyperspectral Digital Image Collection Experiment sensor over the Washington DC Mall. The spectral range of this image is from 0.4 to 2.4 μm and the number of spectral bands is 191. The spatial size of this image is 280×307. Fig. 4(b) shows the false color composite of the *Washington DC* image.

3) Moffett and Jasper Ridge Images: The two images are two popular 16-bit radiance data sets of the AVIRIS sensor. They have the same spectral range, which is from 0.4 to 2.5 μm, and the number of spectral bands is 224. The spatial sizes of the two images are both 512×512. The false composites of the two HSIs are, respectively, shown in Fig. 4(c) and (d).

4) Methods for Comparison: To further demonstrate the effectiveness of the proposed SSASR method, we compare the SSASR method with other classical or state-of-the-art compression methods, including the JPEG2000 with band independent fixed rate (JPEG-BIFR) method [11], the 3-D-SPECK method [17], and the PCA + JPEG2000 method [19]. In particular, the SSASR method is also compared with the SR compression method without considering the spatial information, which is called the PWSR method. The JPEG-BIFR method first uses JPEG2000 to compress each band of HSIs by fixing the same target rate for each band and then distributes the target rate among the bands with a simple bit allocation method. The 3-D-SPECK method applies a 3-D DWT to HSIs and then recursively splits the obtained blocks to find the significant coefficients. The PCA + JPEG2000 method uses the PCA algorithm for the spectral decorrelation of HSIs and then adopts the JPEG2000 for the compression in the spatial domain. The JPEG-BIFR method and the PCA + JPEG2000 method are implemented with Kakadu Software.[1]

5) Parameter Setting: For the proposed SSASR method, the number K of the superpixels plays an important role in exploiting the spatial information. Thus, it must be carefully determined. In this paper, the numbers of the superpixels for the *Indian Pines* image, the *Washington DC* image, the *Moffett* image, and *Jasper Ridge* image are set to 600, 2500, 5000, and 5000, respectively. The discussion of this parameter will be demonstrated in the following section. For superpixel construction, the iteration number I is fixed to 10, which can make superpixels construction algorithm be converged [32]. The size of entire dictionary $\mathbf{D}$ is 1.5 times of the spectral bands' number B. For subdictionary construction, dissimilar pixels in training data are automatically picked up and regarded as the initial cluster centers in the K-means algorithm. The spectral angle distance (SAD) is used to measure the similarity of two pixels. When two pixels' SAD is larger than a threshold α, they are recognized as dissimilar pixels. Initial cluster centers are chosen based on the rule that any two initial cluster centers' SAD should be larger than α. In our experiments, α is set to an empirical value of 25°. Let C represent the number of chosen initial cluster centers. Since C is also the number of subdictionaries, the number of atoms in each subdictionary will be $P = 1.5B/C$. The dictionary for the PWSR method contains $1.5B$ atoms, which are HSI pixels uniformly sampled from image. The parameters of other compression methods including the JPEG-BIFR method, the 3-D-SPECK method, and the PCA + JPEG2000 method are set to the default values in [11], [17], [19], respectively.

[1] http://www.kakadusoftware.com

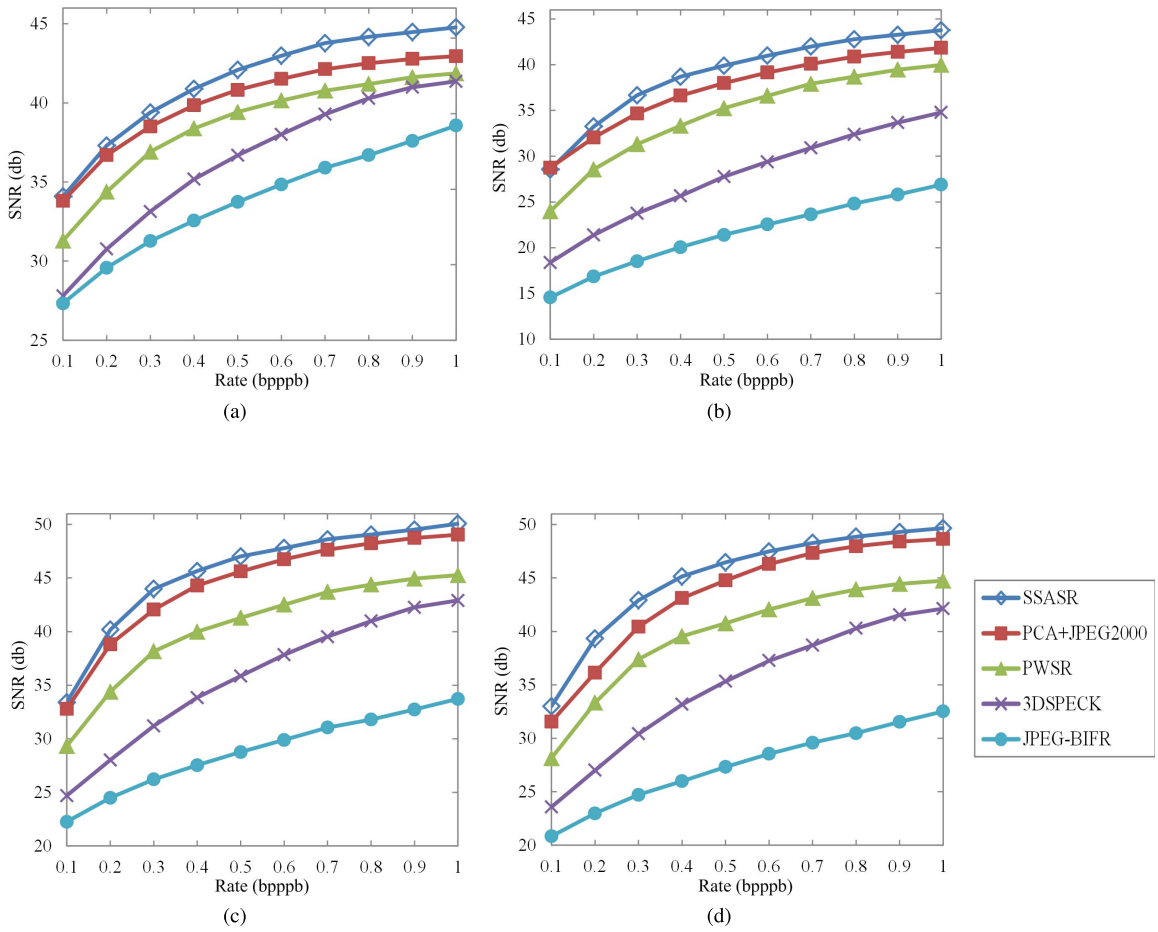


Fig. 5. Rate–distortion performances of various compression methods over four HSIs. (a) *Indian Pines* image. (b) *Washington DC* image. (c) *Moffett* image. (d) *Jasper Ridge* image.

B. Compression Performance Comparisons

In this section, the compression performances of the proposed SSASR method are mainly evaluated by the widely used rate–distortion performance. In addition, the spectral fidelity of the SSASR method is analyzed as in [20].

1) Rate–Distortion Performance: The rate–distortion performance, which is a very important indicator for HSI compression [19], [20], is analyzed. HSIs are compressed and then reconstructed by decompressing. The signal-to-noise ratios (SNRs) between the original image and the reconstructed images is calculated to measure the distortion of HSIs. The SNR is calculated by

$$\mathrm{SNR} = 10\log_{10}\frac{\| \mathbf{Y} \|_F^2}{\|\mathbf{Y} - \hat{\mathbf{Y}}\|_F^2} \tag{10}$$

where $\hat{\mathbf{Y}}$ denotes the reconstructed hyperspectral data. The rates are quantified using the bpppb and are ranged from 0.1 to 1 bpppb. Note that the smaller rates mean the larger CRs. Then, based on the concept that larger SNRs indicate less distortion of the data, the proposed SSASR method is compared with other compression methods in terms of SNR values. The quantitative results over four test HSIs are shown in Fig. 5. As can be observed in Fig. 5, the proposed SSASR method achieves the highest SNRs for most rates, which shows the superiority over compared methods, i.e., the JPEG-BIFR method [11], 3-D SPECK method [17], PWSR method, and PCA + JPEG2000 method [19]. Moreover, the improvement of SSASR method over the PWSR method is large, which demonstrates the effectiveness of exploiting the spectral and spatial redundancy for compression. The PCA + JPEG2000 method also has very high SNRs compared with other methods (e.g., JPEG-BIFR, 3-D-SPECK, and PWSR). Especially when rates are small (e.g., 0.1 bpppb), the PCA + JPEG2000 method is comparable to the proposed SSASR method since the gaps of SNRs are very limited. As rates increase, the SSASR method obviously outperforms the PCA + JPEG2000 method in terms of SNR values. Therefore, the proposed SSASR method provides a good rate–distortion performance.

Next, we conduct the experiments to analyze the effect of the number K of the superpixels on the proposed SSASR method. In the experiments, the number of the superpixels is ranged from small values to large values. The corresponding SNRs with respect to different rates (i.e., 0.1, 0.5, and 1.0 bpppb) are recorded and used to analyze the effect on the SSASR method. The visual results are shown in Fig. 6. As can be seen in Fig. 6, as K increases, SNRs fist grow to the stable and optimal values and then gradually drop.

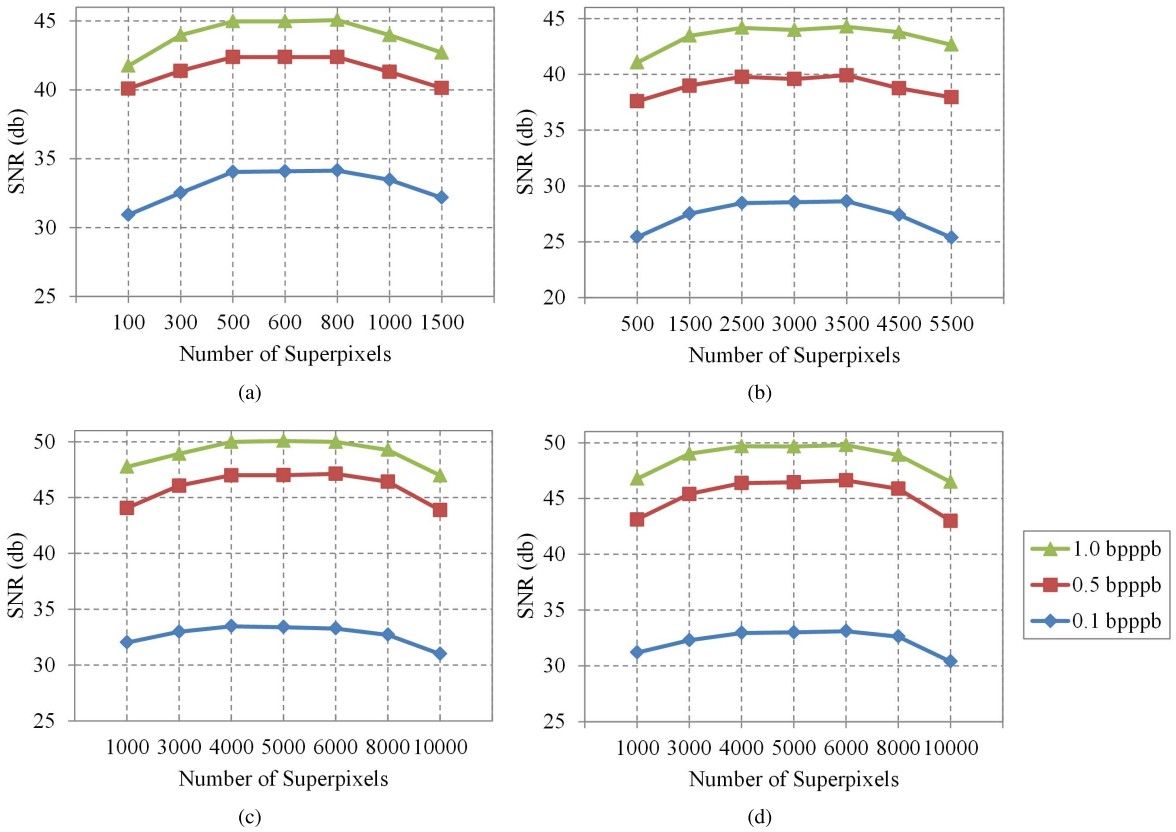


Fig. 6. Rate–distortion performances of various compression methods over four HSIs. (a) *Indian Pines* image. (b) *Washington DC* image. (c) *Moffett* image. (d) *Jasper Ridge* image.

For the four HSIs, the stable SNRS are obtained when the numbers of superpixels are ranged from 500 to 800, 2500 to 3500, 4000 to 6000, and 4000 to 6000, respectively. The reasons why the SNR reduces with too small and too large K are shown below. A too small K means that each superpixel contains many pixels. This can make the superpixel more heterogeneous and thus degrade the performance of joint sparse coding (i.e., simultaneously SR). As a result, HSIs are distorted more seriously after compression. A too large K means each superpixel contains less pixels and will result in the degradation of the SSASR method to the PWSR.

2) Spectral Fidelity Analysis: The rate–distortion performance is used to evaluate the overall data fidelity. Except for the overall data fidelity, we also care about the spectral fidelity since spectral information is very important for HSI applications, e.g., classification and target detection. To analyze the spectral fidelity, we conduct experiments by directly measuring the distortion of all HSI pixels' spectra. The *Indian Pines* image and *Washington DC* image are used as test images. In the experiments, the spectral distortion between original HSI pixels and corresponding reconstruction pixels (obtained after decompressing) are measured by the SADs. The SAD between original pixel $\mathbf{y}$ and reconstruction pixel $\hat{\mathbf{y}}$ can be denoted by

$$\mathrm{SAD}(\mathbf{y}, \hat{\mathbf{y}}) = \cos^{-1} \frac{\mathbf{y}^T \hat{\mathbf{y}}}{\|\mathbf{y}\|_2 \cdot \|\hat{\mathbf{y}}\|_2}. \tag{11}$$

The smaller value of SAD means that the spectra of two pixels are more similar, and thus indicates that the distortion of pixels' spectra is smaller. The mean and standard deviation of SADs over all pixel pairs (original pixel and reconstructed pixel) are calculated to indicate the average spectral distortion of HSIs. The obtained mean and standard over two test HSIs are shown in Table I. In order to further show the effectiveness of the proposed SSASR method, we compare the results of SSASR method with the results of other compression methods (i.e., JPEG-BIFR [11], 3-D-SPECK [17], PCA + JPEG2000 [19], and PWSR). All mean and standard values of SADs with respect to various compression methods are also shown in Table I. As can be observed in Table I, in most cases, the proposed SSASR method has the smallest values of μ_{SAD} and σ_{SAD}, which means that the average spectral distortion of the proposed SSASR method is smaller than other compression methods. Therefore, the proposed SSASR method has a good performance of spectral fidelity.

C. Effect of Compression on Classification

Classification is an important application of HSIs. In order to analyze the effect of compression on classification, we conduct experiments in this section. Similar to other HSI compression works [11], [19], [46], the pixelwise SVM [47] with RBF kernel is used as the classifier. The classifier is used to classify the reconstruction HSIs (obtained after decompressing) and original HSIs, respectively. The obtained classification results can be used for analysis. In the experiments, the *Indian Pines* image and *Washington DC* image are used as test images since

TABLE I

MEAN AND STANDARD DEVIATION OF SADs OVER ENTIRE HSIs WITH RESPECT TO VARIOUS COMPRESSION METHODS: JPEG-BIFR [11], 3-D-SPECK [17], PCA + JPEG2000 [19], PWSR, AND SSASR. THE SMALLEST VALUES FOR EACH CASE ARE GIVEN IN BOLD

	Rate(bpppb)	0.1		0.2		0.3		0.4		0.5	
		μ_{SAD}	σ_{SAD}	μ_{SAD}	σ_{SAD}	μ_{SAD}	σ_{SAD}	μ_{SAD}	σ_{SAD}	μ_{SAD}	σ_{SAD}
Indian Pines	JPEG-BIFR	2.02°	0.93°	1.61°	0.68°	1.37°	0.51°	1.19°	0.43°	1.07°	0.36°
	PCA+JPEG2000	1.13°	0.67°	0.79°	0.26°	0.67°	0.20°	0.60°	0.16°	0.54°	0.11°
	3D-SPECK	2.05°	0.46°	1.46°	0.34°	1.14°	0.25°	0.90°	0.19°	0.78°	0.16°
	PWSR	1.75°	0.36°	1.29°	0.25°	0.97°	0.17°	0.79°	0.11°	0.69°	0.083°
	SSASR	**0.81°**	**0.19°**	**0.67°**	**0.17°**	**0.51°**	**0.099°**	**0.43°**	**0.075°**	**0.36°**	**0.054°**
Washington DC	JPEG-BIFR	8.16°	4.18°	6.61°	3.36°	5.86°	2.95°	5.16°	2.57°	4.60°	2.16°
	PCA+JPEG2000	1.73°	1.04°	1.25°	0.72°	1.00°	0.47°	0.85°	0.37°	0.76°	0.33°
	3D-SPECK	6.15°	2.65°	4.65°	1.97°	3.67°	1.47°	3.08°	1.22°	2.58°	0.99°
	PWSR	3.12°	1.71°	1.91°	0.89°	1.42°	0.59°	1.13°	0.47°	0.97°	0.34°
	SSASR	**1.81°**	**0.98°**	**0.96°**	**0.42°**	**0.69°**	**0.28°**	**0.56°**	**0.22°**	**0.52°**	**0.20°**
	Rate(bpppb)	0.6		0.7		0.8		0.9		1.0	
		μ_{SAD}	σ_{SAD}	μ_{SAD}	σ_{SAD}	μ_{SAD}	σ_{SAD}	μ_{SAD}	σ_{SAD}	μ_{SAD}	σ_{SAD}
Indian Pines	JPEG-BIFR	0.95°	0.29°	0.84°	0.24°	0.78°	0.22°	0.72°	0.20°	0.65°	0.17°
	PCA+JPEG2000	0.49°	0.090°	0.49°	0.086°	0.48°	0.082°	0.48°	0.081°	0.48°	0.073°
	3D-SPECK	0.69°	0.11°	0.59°	0.099°	0.52°	0.089°	0.50°	0.084°	0.48°	0.080°
	PWSR	0.66°	0.073°	0.62°	0.060°	0.57°	0.056°	0.51°	0.049°	0.49°	**0.043°**
	SSASR	**0.36°**	**0.054°**	**0.34°**	**0.049°**	**0.33°**	**0.047°**	**0.33°**	**0.047°**	**0.32°**	0.046°
Washington DC	JPEG-BIFR	4.02°	1.91°	3.67°	1.72°	3.30°	1.53°	3.02°	1.34°	2.70°	1.16°
	PCA+JPEG2000	0.70°	0.26°	0.64°	0.22°	0.61°	0.22°	0.60°	0.22°	0.60°	0.21°
	3D-SPECK	2.16°	0.81°	1.87°	0.68°	1.63°	0.60°	1.38°	0.50°	1.22°	0.43°
	PWSR	0.90°	0.31°	0.59°	0.21°	0.67°	**0.14°**	0.61°	**0.13°**	0.44°	0.15°
	SSASR	**0.46°**	**0.17°**	**0.42°**	**0.16°**	**0.40°**	0.15°	**0.37°**	0.14°	**0.36°**	**0.14°**

the two HSIs have labeled pixels (i.e., ground reference). In the pixelwise classification, 10% and 4% of the labeled pixels in the *Indian Pines* image and the *Washington DC* image are, respectively, used for training, and the rest labeled pixels are test samples. The classification results with respect to various compression methods are shown in Fig. 7. The classification results over original HSIs are used as referential results and represented by the red dotted lines.

Since pixelwise SVM classification is mainly based on spectral information, the distortion of spectral information by HSI compression can change the classification results. Therefore, as the change of classification results is large, it means that HSI compression seriously distorts spectral information and thus largely influences classification performance. As reported in Fig. 7(a), all compression methods distort spectral information and thus have lower classification accuracies than referential accuracies, especially when rates are small (from 0.1 to 0.3 bpppb). However, overall accuracies (OAs) of the SSASR method are consistently closer to referential OAs compared with those of other methods. The experiments of spectral fidelity analysis in the last section have shown that the spectral distortion of the SSASR method is smaller. As a result, the change of classification results by the SSASR compression is smaller. Therefore, the influence of the SSASR compression on classification is smaller. Unlike the *Washington DC* image, which contains many detailed regions, the *Indian Pines* image mainly consists of large homogeneous regions and contains less detailed regions. Therefore, some compression methods, which smooth HSIs, can improve classification accuracies over the *Indian Pines* image. For example, Fig. 7(b) shows that the JPEG-BIFR method has very high classification accuracies when rates are small (i.e., 0.1 to 0.3 bpppb). The classification accuracies of the JPEG-BIFR method are even much higher than the referential accuracies, which means that the change of classification results by the JPEG-BIFR method is large.

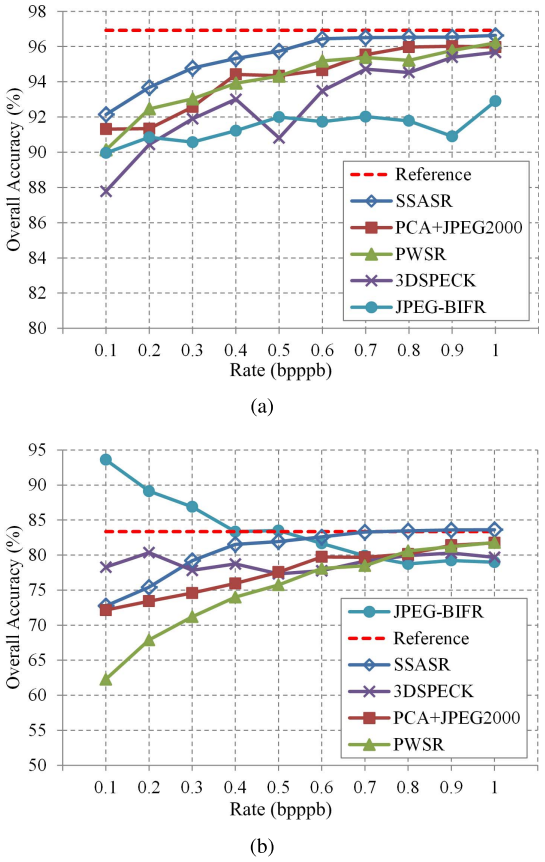


Fig. 7. OA (%) with respect to various compression methods over two HSIs. (a) *Washington DC* image. (b) *Indian Pines* image.

Therefore, the influence of the JPEG-BIFR compression on classification is large. As can be seen in Fig. 7(b), the classification results of the proposed SSASR method are closer

TABLE II
EXECUTION TIME(S) OF THE *Indian Pines* IMAGE COMPRESSION OBTAINED BY JPEG-BIFR [11], 3-D-SPECK [17], PCA + JPEG2000 [19], PWSR, AND SSASR

Rate (bpppb)	C/C++			Matlab	
	3D-SPECK	JPEG-BIFR	PCA+ JPEG2000	PWSR	SSASR
0.1	2.03	1.88	2.28	6.19	7.13
0.2	2.24	1.74	2.06	8.61	7.61
0.3	2.58	1.68	2.12	12.2	8.57
0.4	2.84	2.08	2.20	16.9	9.35
0.5	3.21	1.97	2.43	21.3	10.1
0.6	3.52	2.06	2.54	28.5	11.0
0.7	3.72	1.93	2.87	36.1	12.3
0.8	4.07	2.04	2.90	43.0	13.9
0.9	4.35	1.98	2.82	50.8	15.3
1.0	4.67	1.91	3.06	60.7	17.1

to the referential results in most cases, which indicates that the SSASR compression's influence on classification is smaller.

D. Computational Complexity

The complexity of the proposed SSASR algorithm is analyzed in this section. Let $M \times N \times B$ represent the size of an HSI. According to [32], the complexity of the superpixel construction is $O(MNI_sB)$, where I_s means the number of iterations required for convergence.

Dictionary construction in this paper has the complexity of $O(KCI) + O((K/C)^3)$, where I is the number of iterations for the K-means algorithm. $O(KCI)$ is the complexity of the K-means algorithm and $O((K/C)^3)$ is the complexity of subdictionary construction. The adaptive sparse coding of superpixels mainly consists of two parts, the initial coding and the iterative coding of superpixels. The SOMP algorithm of solving (4) is the main step in both parts. The complexity of the SOMP algorithm is $O(B^3)$. In the initial coding, the SOMP algorithm is separately applied to K superpixels, so the complexity of this part is $O(KB^3)$. In the iterative coding, the SOMP algorithm runs one time for each iteration. Thus, the complexity of iterative coding is $O(I_cB^3)$, where I_c denotes the number of iterations.

After the adaptive sparse coding, the main computation burden lies on the Huffman coding. The complexity of Huffman coding is $O(\eta MN \log(\eta MN))$, where ηMN is the overall number of nonzero values with respect to all sparse coefficients. Therefore, the total complexity of the proposed SSASR algorithm is $O(MNI_sB) + O(KCI) + O((K/C)^3) + O(KB^3) + O(I_cB^3) + O(\eta MN \log(\eta MN))$.

According to [17], except for 3-D DWT and the entropy coding processes, only the most basic operations, e.g., memory access, bit shifts, additions, and comparisons, are required in the 3-D-SPECK method. For the PCA + JPEG2000 method, the computation burden mainly lies on the implementation of JPEG2000. The execution times of various compression methods are shown in Table II. To save space, Table II shows the results on the *Indian Pines* image. The execution time in Table II is obtained based on the laptop, which has an Intel Core CPU of 2.50 GHz and a 16-GB RAM. The PWSR and SSASR methods are implemented on the MATLAB platform (Huffman coding is implemented by C/C++ code). With superpixel strategy, the SSASR method is more efficient than the PWSR method, which is a pixel-level process, and thus has less execution time in Table II. The 3-D SPECK method, JPEG-BIFR method, and PCA + JPEG2000 method are implemented by C/C++ code. Furthermore, the JPEG-BIFR and PCA + JPEG2000 methods are based on Kakadu Software, which is an optimized software of JPEG2000. Therefore, they have less execution time than the PWSR and SSASR methods. Note that parallel computation and C/C++ version code can be used to accelerate the proposed method.

IV. CONCLUSION

This paper proposes a new SSAR method for HSI lossy compression. In order to exploit the spatial and spectral redundance in the HSI, the SSASR method first constructs the superpixels to group neighboring and similar pixels. Then, the spectral signatures of each superpixel are simultaneously represented by the adaptive sparse coding scheme and result in the joint sparsity in the sparse coefficients. Finally, the sparse coefficients are transformed into bits by Huffman coding and constitute the final bitstream with the encoded superpixel map. The proposed SSASR compression method is evaluated by not only the rate–distortion performance but also the spectral fidelity performance. The experimental results show the superiority of the proposed SSASR method over several state-of-the-art compression methods.

In this paper, Huffman coding is used to transform the sparse coefficients into the bitstream without considering the contextual structure in the sparse coefficients. In addition, the number of superpixels is empirically selected, which is not adaptive for different HSIs. Therefore, our future work will focus on using the contextual structure of the sparse coefficients to further compress the superpixels and adaptively determining the number of superpixels for different HSIs.

ACKNOWLEDGMENT

The authors would like to thank the editors and reviewers whose valuable comments and suggestions greatly helped to improve the quality and presentation of this paper.

REFERENCES

[1] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 45–54, Jan. 2014.

[2] T. Lu, S. Li, L. Fang, Y. Ma, and J. A. Benediktsson, "Spectral–spatial adaptive sparse representation for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens*, vol. 54, no. 1, pp. 373–385, Jan. 2016.

[3] T. Lu, S. Li, L. Fang, L. Bruzzone, and J. A. Benediktsson, "Set-to-set distance-based spectral–spatial classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7122–7134, Dec. 2016, doi: 10.1109/TGRS.2016.2596260.

[4] S. Li, T. Lu, L. Fang, X. Jia, and J. A. Benediktsson, "Probabilistic fusion of pixel-level and superpixel-level hyperspectral image classification,"*IEEE Trans. Geosci. Remote Sen.*, vol. 54, no. 12, pp. 7416–7430, Dec. 2016, doi: 10.1109/TGRS.2016.2603190.

[5] J. Li, H. Zhang, L. Zhang, X. Huang, and L. Zhang, "Joint collaborative representation with multitask learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5923–5936, Sep. 2014.

[6] L. Zhang, L. Zhang, D. Tao, and X. Huang, "On combining multiple features for hyperspectral remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 879–893, Mar. 2012.

[7] W. Li and Q. Du, "Collaborative representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1463–1474, Mar. 2015.

[8] L. Zhang, L. Zhang, D. Tao, X. Huang, and B. Du, "Hyperspectral remote sensing image subpixel target detection based on supervised metric learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4955–4965, Aug. 2014.

[9] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[10] S. Mei, Q. Du, and M. He, "Equivalent-sparse unmixing through spatial and spectral constrained endmember selection from an image-derived spectral library," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 8, no. 6, pp. 2665–2675, Jun. 2015.

[11] F. Garcia-Vilchez *et al.* "On the impact of lossy compression on hyperspectral image classification and unmixing," *IEEE Geosci. Remote Sen. Lett.*, vol. 8, no. 2, pp. 253–257, Mar. 2011.

[12] J. Serra-Sagristà and F. Aulí-Llinàs, "Remote sensing data compression," in *Computational Intelligence for Remote Sensing*. Berlin, Germany: Spring-Verlag, Jun. 2008, pp. 27–61.

[13] L. Zhang, L. Zhang, D. Tao, X. Huang, and B. Du, "Compression of hyperspectral remote sensing images by tensor approach," *Neurocomputing*, vol. 147, pp. 358–363, Jan. 2015.

[14] G. K. Wallace, "The JPEG still picture compression standard," *IEEE Trans. Consum. Electron.*, vol. 38, no. 1, pp. xviii–xxxiv, Feb. 1992.

[15] M. Charrier, D. S. Cruz, and M. Larsson, "JPEG2000, the next millennium compression standard for still images," in *Proc. IEEE Int. Conf. Multimedia Comput. Syst. (ICMCS)*, vol. 1, Jun. 1999, pp. 131–132.

[16] W. A. Pearlman, A. Islam, N. Nagaraj, and A. Said, "Efficient, low-complexity image coding with a set-partitioning embedded block coder," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 11, pp. 1219–1235, Nov. 2004.

[17] X. Tang and W. A. Pearlman, "Three-dimensional wavelet-based compression of hyperspectral images," in *Hyperspectral Data Compression*, New York, NY, USA: Springer-Verlag, 2006, ch. 10, pp. 273–278.

[18] J. T. Rucker, J. E. Fowler, and N. H. Younan, "JPEG2000 coding strategies for hyperspectral data," in *Proc. Int. Geosci. Remote Sens. Symp.*, Seoul, South Korea, vol. 1, Jul. 2005, pp. 128–131.

[19] Q. Du, N. Ly, and J. E. Fowler, "An operational approach to PCA+JPEG2000 compression of hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2237–2245, Jun. 2014.

[20] Q. Du and J. E. Fowler, "Hyperspectral image compression using JPEG2000 and principal component analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 2, pp. 201–205, Apr. 2007.

[21] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.

[22] S. Li, H. Yin, and L. Fang, "Remote sensing image fusion via sparse representations over learned dictionaries," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 2677–4779, Sep. 2013.

[23] L. Zhang, Q. Zhang, L. Zhang, D. Tao, X. Huang, and B. Du, "Ensemble manifold regularized sparse low-rank approximation for multiview feature embedding," *Pattern Recognit.*, vol. 48, no. 10, pp. 3102–3112, Oct. 2015.

[24] K. Skretting and K. Engan, "Image compression using learned dictionaries by RLS-DLA and compared with K-SVD," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 1517–1520.

[25] W. Li and Q. Du, "Joint within-class collaborative representation for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2200–2208, Jun. 2014.

[26] W. Fu, S. Li, L. Fang, X. Kang, and J. A. Benediktsson, "Hyperspectral image classification via shape-adaptive joint sparse representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 556–567, Feb. 2016.

[27] L. Fang, S. Li, W. Duan, J. Ren, and J. A. Benediktsson, "Classification of hyperspectral images by exploiting spectral-spatial information of superpixel via multiple kernels," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6663–6674, Dec. 2015.

[28] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1620–1630, Apr. 2013.

[29] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.

[30] G. Tang and A. Nehorai, "Performance analysis for sparse support recovery," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1383–1399, Mar. 2010.

[31] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, Mar. 2013.

[32] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

[33] G. Zhang, X. Jia, and J. Hu, "Superpixel-based graphical model for remote sensing image mapping," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 5861–5871, Nov. 2015.

[34] D. R. Thompson, L. Mandrake, M. S. Gilmore, and R. Castano, "Superpixel endmember detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4023–4033, Nov. 2010.

[35] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral–spatial classification of hyperspectral images with a superpixel-based discriminative sparse model," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4186–4201, Aug. 2015.

[36] J. Li, H. Zhang, and L. Zhang, "Efficient superpixel-level multitask joint sparse representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5338–5351, Oct. 2015.

[37] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Boston, MA, USA: Kluwer, 1992.

[38] Q. Du and C.-I. Chang, "Linear mixture analysis-based compression for hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 4, pp. 875–891, Apr. 2004.

[39] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.

[40] H. Zhang, J. Li, Y. Huang, and L. Zhang, "A nonlocal weighted joint sparse representation classification method for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2056–2065, Jun. 2013.

[41] D. Leviatan and V. N. Temlyakov, "Simultaneous approximation by greedy algorithms," *Adv. Comput. Math.*, vol. 25, no. 1, pp. 73–90, Jul. 2006.

[42] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, no. 6, pp. 948–958, Jun. 2010.

[43] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, Mar. 2006.

[44] J. Huang, T. Zhang, and D. Metaxas, "Learning with structured sparsity," in *Proc. 26th Int. Conf. Mach. Learn.*, 2009, pp. 417–424.

[45] D. A. Huffman, "A method for construction of minimum-redundancy codes," *Proc. IRE*, vol. 40, no. 9, pp. 1098–1101, Sep. 1952.

[46] A. Karami, M. Yazdi, and G. Mercier, "Compression of hyperspectral images using discerete wavelet transform and tucker decomposition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 444–450, Apr. 2012.

[47] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

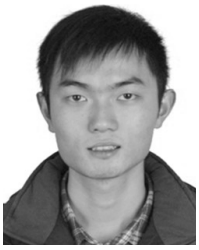
Wei Fu (S'14) received the B.S. degree in electrical engineering from Hunan University, Changsha, China, in 2012, where he is currently pursuing the Ph.D. degree in electrical engineering.

Since 2016, he has been a Visiting Ph.D. Student with the Faculty of Electrical and Computer Engineering, University of Iceland, Reykjavik, Iceland, supported by the China Scholarship Council. His current research interests include hyperspectral image processing and sparse representation.

Shutao Li (M'07–SM'15) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Hunan University, Changsha, China, in 1995, 1997, and 2001, respectively.

In 2001, he joined the College of Electrical and Information Engineering, Hunan University. He was also a Research Associate with the Department of Computer Science, Hong Kong University of Science and Technology (HKUST), Hong Kong, in 2001. From 2002 to 2003, he was a Post-Doctoral Fellow with Royal Holloway College, University of London, Egham, U.K., working with Prof. J. Shawe-Taylor. In 2005, he was a Visiting Professor with the Department of Computer Science, HKUST. He is currently a Full Professor with the College of Electrical and Information Engineering, Hunan University. He is also a Chang-Jiang Scholar Professor appointed by the Ministry of Education of China. He has authored or co-authored over 160 refereed papers. His current research interests include compressive sensing, sparse representation, image processing, and pattern recognition.

Dr. Li was a recipient of two 2nd-Grade National Awards at the Science and Technology Progress of China in 2004 and 2006. He is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and a member of the Editorial Board of *Information Fusion* and *Sensing and Imaging*.



Leyuan Fang (S'10–M'14) received the Ph.D. degree in electrical engineering from Hunan University, Changsha, China, in 2015.

From 2011 to 2012, he was a Visiting Ph.D. Student with the Department of Ophthalmology, Duke University, Durham, NC, USA, supported by the China Scholarship Council. Since 2013, he has been an Assistant Professor with the College of Electrical and Information Engineering, Hunan University. His current research interests include sparse representation and multiresolution analysis in remote sensing and medical image processing.

Dr. Fang was a recipient of the Scholarship Award for Excellent Doctoral Student granted by the Chinese Ministry of Education in 2011.

Jón Atli Benediktsson (S'84–M'90–SM'99–F'04) received the Cand.Sci. degree in electrical engineering from the University of Iceland, Reykjavik, Iceland, in 1984, and the M.S.E.E. and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, USA, in 1987 and 1990, respectively.

From 2009 to 2015, he was the Prorector of Science and Academic Affairs and a Professor of Electrical and Computer Engineering with the University of Iceland, where he became the Rector in 2015. He has authored extensively in his fields of interest. His current research interests include remote sensing, image analysis, pattern recognition, biomedical analysis of signals, and signal processing.

Dr. Benediktsson is a member of the Association of Chartered Engineers at Iceland (VFI), the Societas Scinetiarum Islandica, and the Tau Beta Pi. He is a Fellow of the International Society for Optics and Photonics (SPIE). He was a recipient of the Stevan J. Kristof Award from Purdue University in 1991 as an Outstanding Graduate Student in Remote Sensing. He was also a recipient of the Icelandic Research Council's Outstanding Young Researcher Award in 1997, the IEEE Third Millennium Medal in 2000, the Yearly Research Award from the Engineering Research Institute at the University of Iceland in 2006, the Outstanding Service Award from the IEEE Geoscience and Remote Sensing Society in 2007, and the IEEE/VFI Electrical Engineer of the Year Award in 2013. He was a co-recipient of the University of Iceland's Technology Innovation Award in 2004, the 2012 IEEE Transactions on Geoscience and Remote Sensing Paper Award, the IEEE GRSS Highest Impact Paper Award in 2013, and the International Journal of Image and Data Fusion Best Paper Award in 2014. He was the 2011–2012 President of the IEEE Geoscience and Remote Sensing Society (GRSS) and has been on the GRSS Administrative Committee since 2000. He was the Editor-in-Chief of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS) from 2003 to 2008 and has been an Associate Editor of TGRS since 1999, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS since 2003, and the IEEE ACCESS since 2013. He was the Chairman of the Steering Committee of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING from 2007 to 2010. He serves on the Editorial Board of the IEEE PROCEEDINGS and the International Editorial Board of the *International Journal of Image and Data Fusion*. He is the Co-Founder of a biomedical startup company Oxymap.