A Novel Lossless Compression for Hyperspectral Images by Context-Based Adaptive Classified Arithmetic Coding in Wavelet Domain

Jing Zhang and Guizhong Liu

***Abstract*—A novel hyperspectral-image lossless compression scheme in the wavelet domain is proposed in this letter. This scheme is based on the context-based adaptive classified arithmetic-coding technique. The adaptive classified scheme divides each of the residual images between the two adjacent wavelet images into different classes, resulting in not only skipping the coding of a lot of insignificant zeros but also making the similar coefficients cluster together. Through experiments, we found that, when similar coefficients are clustered together, the arithmetic coding can achieve a higher performance than no clustering. Therefore, we can say that the adaptive classified scheme makes a better use of the characteristics of hyperspectral images and the characteristics of the arithmetic-coding technique. Experiments show that our proposed scheme is capable of providing high compression performance.**

***Index Terms*—Context-based adaptive classified arithmetic coding, hyperspectral images, lifting integer wavelet transforms.**

I. Introduction

Hyperspectral images represent the intensities of energy reflected or emitted by the ground targets, with possibly hundreds of different wavelength bands. As a result, a sequence of hyperspectral images consists of a series of images corresponding to hundreds of continuous spectral bands; such a 3-D representation creates a huge number of data for computer processing and data transmission. Unlike natural images, hyperspectral images have two types of correlations, which are the spatial correlation and the spectral correlation. Making the best use of these two types of correlations is the key to an efficient compression algorithm.

Nowadays, the wavelet transform has been proven successful in many areas. It is employed in most state-of-the-art compression schemes. Many promising 3-D image compression algorithms based on wavelet transforms were proposed recently. The classical 3-D wavelet-image coding algorithm is 3D-SPIHT proposed by Kim and Pearlman [1]. It is an extension of the original 2D-SPIHT [2] and has a 3-D tree structure. 3D-SPIHT has been applied to multispectral image compression by Dragotti *et al.* [3]. Another more efficient 3D-SPIHT named asymmetry-3D-SPIHT for hyperspectral image compression is proposed in [4]. Xu *et al.* [5] extend an embedded block coding with optimized truncation to video coding, which is 3-D embedded subband coding with optimized truncation (3D-ESCOT). Tang *et al.* [6] extend 2DSPECK [7] to 3DSPECK for hyperspectral image compression. Asymmetry-3DSPECK for hyperspectral image compression is proposed in [8]. Some algorithms [9], [10] use JPEG2000 and spectral decorrelation to compress the hyperspectral images.

Lossless coding is a very important feature for hyperspectral image compression. In this letter, we propose a novel lossless compression algorithm based on wavelet transforms. This algorithm is not only simple but also efficient for hyperspectral images. As we all know, discrete wavelet transforms [11] cannot be used to achieve lossless compression. The integer wavelet transform based on the lifting scheme introduced by Sweldens [12] can be carried out to achieve lossless compression while reducing computational complexity. In this letter, we use the integer wavelet transform [13] to compress 3-D hyperspectral images. Each residual image of two adjacent bands in the wavelet domain is divided into different classes according to the significance maps of the wavelet coefficients, and these coefficients are derived from the reference bands. Then, context-based adaptive arithmetic coding is performed for each class independently.

This letter is organized as follows. In Section II, adaptive classified arithmetic coding in the wavelet domain is described in detail. Section III shows some characteristics of the arithmetic-coding technique. In Section IV, the whole compression algorithm is proposed. We discuss the experimental results in Section V. Some conclusions are given in Section VI.

II. Adaptive Classified Arithmetic Coding in the Wavelet Domain

In this section, according to the spectral correlation, each of the residual images between the two adjacent wavelet images is divided into different classes, with each class having similar coefficients.

As is well known, for hyperspectral images, the values of pixels having the same spatial location but in different spectral bands are different energies reflected or emitted by the same ground target. As a result, the spectral correlations are very strong. Therefore, we suppose that for all the bands, after the 2-D spatial wavelet transform, the spectral correlations are still very strong. Fig. 1 describes the two adjacent spectral bands

Manuscript received October 16, 2006; revised March 18, 2007. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Project 60572045, by the Ministry of Education of China Ph.D. Program Foundation under Project 20050698033, and by a Cooperation Project (2005.7–2007.6) with Microsoft Research Asia.

The authors are with the School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: liugz@mail.xjtu.edu.cn).

Digital Object Identifier 10.1109/LGRS.2007.897924

1545-598X/$25.00 © 2007 IEEE

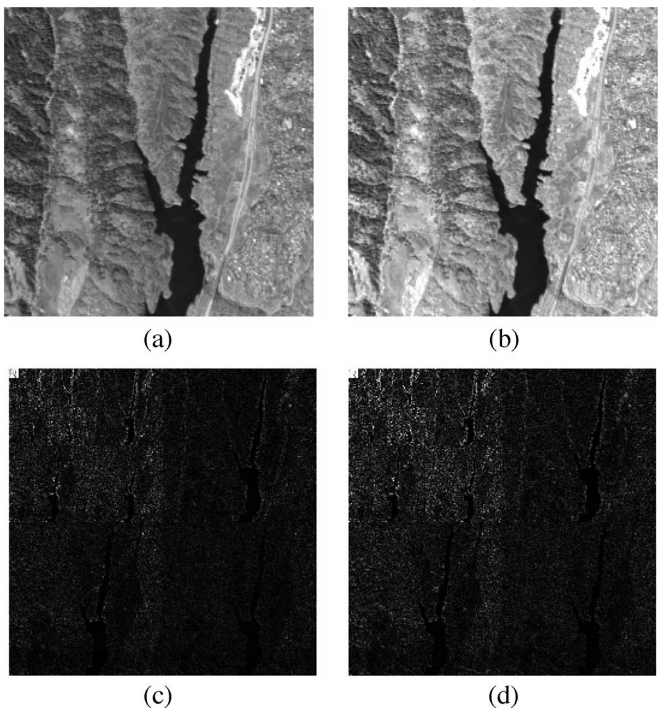

Fig. 1. Images before and after the wavelet transforms of bands 39 and 40 belonging to the first scene of Jasper Ridge AVIRIS image. (a) Band 39 before wavelet transforms. (b) Band 40 before wavelet transforms. (c) Band 39 after wavelet transforms. (d) Band 40 after wavelet transforms.

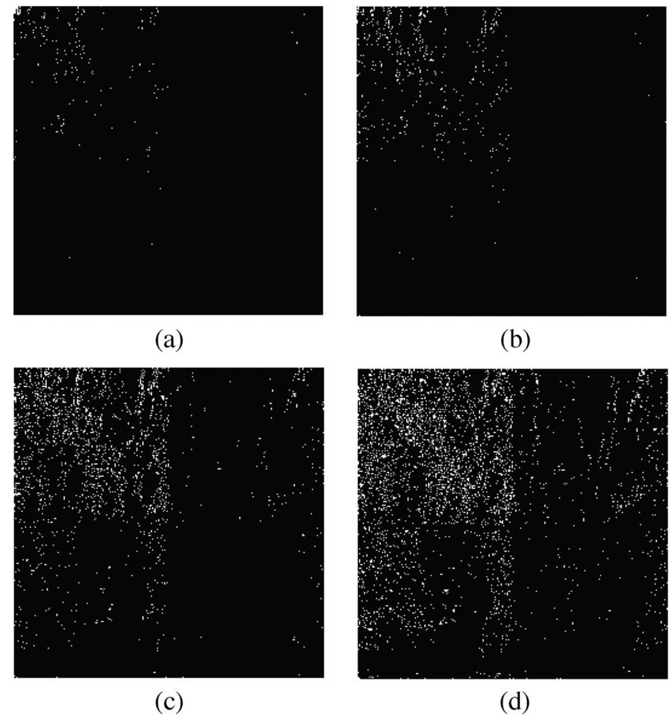

Fig. 2. Significance maps at different thresholds of images in Fig. 1(c) and (d). (a) $T = 2^{10}$, band 39. (b) $T = 2^{10}$, band 40. (c) $T = 2^9$, band 39. (d) $T = 2^9$, band 40.

before and after wavelet transforms [a 256 $\times$ 256 area of bands 39 and 40, which exhibits all key features present in the images, belonging to the first scene of Jasper Ridge Airborne Visible-Infrared Imaging Spectrometer (AVIRIS) image].

The correlation factor of the two images in Fig. 1(a) and (b) is 0.9761, and the correlation factor of the two images in Fig. 1(c) and (d) is 0.9715. Fig. 1 confirms that wavelet transforms just change the spatial structure of the bands; the spectral correlation is still high. In order to make a better use of the spectral correlation in the wavelet domain, now, we analyze the correlation of the two images in Fig. 1(c) and (d). First, we define a significance map [14] of a given threshold T and an image I in (1). Let $|I(x,y)|$ represent the absolute value of the wavelet coefficient at the location (x,y), and $S_T(x,y)$ represent the value at (x,y) of the significance map for the threshold T, i.e.,

$$\begin{aligned} &\text{If} \qquad (T \leq |I(x,y)| < 2^{*}T) \\ &\qquad\qquad S_T(x,y) = 1 \\ &\text{else} \qquad S_T(x,y) = 0. \end{aligned} \tag{1}$$

Fig. 2 shows some significance maps of the wavelet-transformed images in Fig. 1(c) and (d) with two different thresholds. From Fig. 2, it is evident that the significance maps of the two adjacent spectral bands are similar in the wavelet domain. The proportion of the equal values of the two significance maps in Fig. 2(a) and (b) is 99.4%. The proportion of the equal values of the two significance maps in Fig. 2(c) and (d) is 95.9%. Therefore, in the two adjacent spectral bands, in the same position, the coefficients have either big values or small values at the same time. We suppose that the difference of the two big coefficients is potentially big, and the difference of the two small coefficients is potentially small. If we want to divide the residual image of the two adjacent wavelet images into different classes with each class having similar values, we can use these significance maps, one significance map corresponding to one class. Let us compare the number of coding bits in the classification case with the no-classification case for the residual image of Fig. 1(d) from (c). The results shown in Table I confirm that this supposition is sound, as similar residual coefficients are clustered together, saving a lot of bits to be coded. For one residual band, the numbers of coding bitplanes are different for the classes, which are always smaller than the maximum number of coding bitplanes of all the coefficients, as shown in the third column of Table I. If there was no classification, all of the coefficients would be coded with the maximum possible number of coding bitplanes for all the classes. Therefore, the adaptive binary arithmetic coding [15] is performed for each class independently, saving a lot of bits to be coded. Because of the similarity of significance maps for the two adjacent spectral bands, the significance maps of the previous spectral band are used to classify the residual image. In the decoder, before decoding the residual image, the previous spectral band has been decoded, and the significance maps can be obtained. That means that the class map does not need to be transmitted. The introduction of the classification just skips coding many insignificant zeros, does not lose any useful information, and improves the compression performance.

III. Characteristics of Arithmetic Coding

In this section, we restrict our attention to arithmetic coding.

TABLE I
COMPARISON OF CODING BITS OF CLASSIFICATION AND NO CLASSIFICATION FOR THE RESIDUAL IMAGE OF FIG. 1(d) FROM (c)

	Threshold	Number of coding bitplanes	Number of significant symbols	Number of saved bits	Total saved bits
Classification	2^{11}	10	31	31	74269
	2^{10}	11	215	0	
	2^{9}	10	2020	2020	
	2^{8}	10	8490	8490	
	2^{7}	11	15755	0	
	2^{6}	10	15093	15093	
	2^{5}	9	10290	20580	
	2^{4}	9	6312	12624	
	2^{3}	9	3627	7254	
	2^{2}	9	1979	3958	
	2^{1}	9	953	1906	
	2^{0}	8	509	1527	
	0	8	262	786	
No classification		11	65536	0	

TABLE II
EFFECT OF THE INTRODUCTION OF CLASSIFICATION TO ARITHMETIC CODING

Bands of JR1	Bit rate (bpp)		
	No classification	Classification without class map	Classification with class map
1	7.68	5.09	7.63
2	8.41	5.80	8.35
3	8.38	5.77	8.31
4	8.20	5.53	8.08
5	7.94	5.28	7.86
6	7.48	4.86	7.46
7	7.13	4.54	7.15
8	7.17	4.55	7.16
9	7.16	4.52	7.15
10	6.86	4.29	6.88

A. *Effect of Classification*

The performance of the arithmetic coding depends mainly on the estimation of the probability model that the coder will use and the arithmetic coding will approach the entropy of the source. The smaller the entropy of the input data is, the higher the compression ratio is. The adaptive binary arithmetic coding can adaptively explore the local entropy of the image. If the local entropy is smaller, the compression ratio is higher. How to decrease the local entropy is a key of the coding. Through experiments, we have found that, if all the coefficients are divided into different classes, which have similar values, then each class is coded independently, and the local entropy becomes smaller. This is due to the fact that, when coding similar values, lots of “0s” or “1s” appear successively at the local positions; the probability of 0 and the probability of 1 are very unbalanced, and this leads to smaller entropy. The results in Table II confirm this fact. “Classification without class map” means that all the coefficients are divided into different classes with all coefficients in one class having the same number of coding bitplanes, and then, each class is coded independently. It can decrease the bit rate, but, in practice, it cannot be realized. If we want to decode, the class map must be known, but this algorithm does not code the class map. The results in Table II also show that if the class map is coded, the bit rate increases and becomes even higher than in the case of no classification. We assume that, if we can divide all the coefficients into different classes but the class map is not coded, good performance can be achieved. The classification algorithm proposed in Section II can realize this assumption. It classifies all the coefficients into different classes according to the spectral correlation. The classification is based on the significance maps of the previous spectral band, so the class information does not need to be coded.

B. *Effect of Context*

After classification, in each of the classes, there are three types of symbols that should be coded: the first “1” and its previous insignificant “0,” the “1” or “0” after the first “1,” and the sign. The characteristic of the sign is different from other symbols, so when the sign is coded, a fixed context is used for these symbols in all the classes. In the same way, when the “1” or “0” after the first “1” is coded, another fixed context is used for these symbols in all the classes. In each of the classes, the characteristics of the first “1” and its previous insignificant “0” are different, so for each class, we assign a context for these symbols. A lot of experiments show that the introduction of context can improve the performance of arithmetic coding. Therefore, in our proposed algorithm, the context-based adaptive classified arithmetic-coding technique is used.

IV. PROPOSED COMPRESSION ALGORITHM

Our compression scheme is described next.

1) Do the 2-D spatial five-level (5,3) lifting integer wavelet transforms for each of the spectral bands. The residual images are the different images of the two adjacent wavelet images.

TABLE III
EFFECT OF THE INTRODUCTION OF CLASSIFICATION

Image name	Average compression ratio		
	No Classification	Classification + Independent	Classification+ Independent+ Context
JR1	3.02	3.13	3.19
JR2	3.05	3.15	3.21
LL1	3.06	3.12	3.18
LL2	3.02	3.10	3.16
CU1	3.04	3.12	3.18
CU2	2.96	3.06	3.11

2) For the first spectral band, we code all the wavelet coefficients using the binary arithmetic coding in raster order from the most significant bitplane toward the least significant bitplane.
3) According to the spectral correlation, divide each of the residual images into different classes as described in Section II. Then, save the maximum number of coding bitplanes of each class in every residual image as the side information transmitted to the decoder.
4) Perform the context-based adaptive binary arithmetic coding for each class independently.
5) The final code stream is made up of the entire compression code stream and the side information. To decode the code stream, just carry out the reverse process of the encoding.

V. Experimental Results

The performance of the proposed algorithm has been tested on several AVIRIS hyperspectral images. AVIRIS is a Jet Propulsion Laboratory instrument having 224 continuous bands ranging from the visible to the near-infrared regions (400–2500 nm) (http://aviris.jpl.nasa.gov). The spectral components are sampled with a 12-bit precision; after radiometric correction, data are stored as 16-bit signed integers. The unit of the recorded hyperspectral images is the so-called scene, which is a data cube of 512 rows by 614 columns by 224 bands. Typical hyperspectral images consist of three or more consecutive scenes. The hyperspectral images for our test are the first and the second scene of Jasper Ridge, Lunar Lake, and Cuprite (JR1, LL1, CU1, JR2, LL2, CU2). For the sake of simplicity, the bands are 256×256 pixels, starting at the coordinates (200,180) of the original size 512×614.

First we test the performance of the introductions of the classification. Table III shows the average compression ratios (CRs) of 224 bands. "No classification" in the table means the coding of each of the residual images using the adaptive binary arithmetic coding directly. "Classification + independent" means that the classification is introduced, and each class is coded independently. "Classification + independent + context" uses the context-based adaptive binary arithmetic coding. According to Section III, we can see that if the similar coefficients are clustered together, the arithmetic coding has a good performance, but coding the class map will weaken the performance. In our proposed algorithm, we can divide the coefficients into different classes, which have similar values, without coding the class map. The introduction of our proposed classification not only skips the coding of many insignificant zeros but also makes better use of the characteristics of the arithmetic coding, so it shows significant compression performance. From Table III, the introduction of context can further improve the performance, whereas the computational complexity does not increase because the contexts come from the class map, which is known before arithmetic coding.

TABLE IV
COMPARISON OF THE 224-BAND AVERAGE COMPRESSION RATIO

Image name	Average compression ratio				
	3D-SPIHT	A3D-SPIHT	M-CALIC	The proposed algorithm	Reordering+ the proposed algorithm
JR1	2.54	2.81	2.89	3.19	3.29
JR2	2.57	2.85	2.95	3.21	3.31
LL1	2.70	3.00	3.05	3.18	3.29
LL2	2.60	2.91	3.00	3.16	3.26
CU1	2.61	2.91	3.00	3.18	3.33
CU2	2.49	2.79	2.90	3.11	3.24
Average	**2.58**	**2.88**	**2.96**	**3.17**	**3.29**

Next, we compare the proposed lossless compression algorithm with some of the existing algorithms for hyperspectral images. The 224-band average compression ratios are shown in Table IV. "Reordering + the proposed algorithm" is the algorithm; before performing the proposed algorithm, the adaptive spectral-band reordering algorithm proposed in [16] is executed. The adaptive spectral-band reordering algorithm finds out the nearly best reference band for each of the bands. If the reference band is the same as the current band, perform "intraband coding" as in [16] for the current band; otherwise, the residual images are the differences between each of the bands and its nearly best reference band. "3D-SPIHT" is the algorithm using the dyadic 3-D decomposition in [17], and the three-level (5,3) integer wavelet transforms in spatial domain and the three-level Haar integer wavelet transforms in the spectral domain are used in a group-of-picture (GOP) of 16 adjacent spectral bands. "A3D-SPIHT" is the algorithm using an asymmetric tree structure for the 3-D wavelet transform in a GOP of 16 adjacent spectral bands; first, do the five-level (5,3) integer wavelet transforms in the spatial domain, and then do the four-level Haar integer wavelet transforms in the spectral domain. The asymmetric tree structure is the same as in [4]. "M-CALIC" is the algorithm proposed in [18], and all of the parameters here are the same as in [18].

All the results in Table IV are obtained by our own programs. SPIHT is very attractive for lossy compression, because it provides a high peak signal-to-noise ratio at low bit rate, but it is not so good for lossless compression. Because of the high spatial complexity of the hyperspectral images, 3D-SPIHT cannot show its advantage. We can see that A3D-SPIHT is better than 3D-SPIHT; it means that the asymmetric 3-D wavelet transform is more fit for the hyperspectral image. M-CALIC is a complicated context-based algorithm; the introductions of the prediction context and the coding context can achieve a better compression performance. Therefore, it can make a better use of the high spectral correlation of hyperspectral images. Our proposed algorithm is effective and outperforms M-CALIC. It is worth noticing that our proposed algorithm cannot realize

TABLE V
ENCODING TIME OF THE 224 BANDS

Image name	Encoding time of 224 bands (seconds)			
	3D-SPIHT	A3D-SPIHT	M-CALIC	The proposed algorithm
JR1	103.0	264.4	180.0	119.1
LL1	95.1	210.9	179.1	121.1
CU1	98.7	212.3	179.0	118.7
Average	**98.9**	**229.2**	**179.4**	**119.6**

lossless, lossy, and near-lossless compression capabilities in one and the same code stream simultaneously. If the spectral-band reordering is introduced before the proposed algorithm as "reordering + the proposed algorithm," a further improvement can be achieved. In order to evaluate the complexity of the proposed algorithm, we have run 3D-SPIHT, A3D-SPIHT, M-CALIC, and the proposed algorithm programs in lossless mode on a workstation with a Pentium IV 2.4-GHz processor and Windows XP operating system. We have measured the encoding time of 224 bands for each algorithm by using the clock() function. Each of the encoding times is an average over 20 time trials. The experimental results are shown in Table V. It is clear that the encoding time of our proposed algorithm is close to 3D-SPIHT and much less than A3D-SPIHT and M-CALIC. From Tables IV and V, we can see that our proposed algorithm has a moderate computational complexity and achieves a higher compression ratio. Therefore, it is an efficient lossless compression algorithm for hyperspectral images.

VI. CONCLUSION

In this letter, we propose a novel algorithm for hyperspectral-image lossless compression. The introduction of the classification not only makes full use of the spectral correlation and spatial correlation but also makes a better use of the characteristics of the arithmetic coding. The computational complexity of the proposed algorithm is moderate. Therefore, it is a novel and efficient lossless compression algorithm for hyperspectral images.

ACKNOWLEDGMENT

The authors would like to thank the Editor, the Associate Editor, and the anonymous reviewers for their precious comments and suggestions which have been used in the revision of this letter.

REFERENCES

[1] B. J. Kim and W. A. Pearlman, "An embedded wavelet video coder using three-dimensional set partitioning in hierarchical trees (SPIHT)," in *Proc. DCC*, Mar. 1997, pp. 251–260.
[2] A. Said and W. A. Pearlman, "A new, fast and efficient image CODEC based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 3, pp. 243–250, Jun. 1996.
[3] P. L. Dragotti, G. Poggi, and A. R. P. Ragozini, "Compression of multispectral images by three-dimensional SPIHT algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 1, pp. 416–428, Jan. 2000.
[4] X. Tang, S. Cho, and W. A. Pearlman, "3D set partitioning coding methods in hyperspectral image compression," in *Proc. ICIP*, Sep. 2003, vol. 2, pp. 239–242.
[5] J. Xu, Z. Xiong, S. Li, and Y. Zhang, "Three dimensional embedded subband coding with optimized truncation (3-D ESCOT)," *J. Appl. Comput. Harmonic Anal.*, vol. 10, no. 3, pp. 290–315, May 2001.
[6] X. Tang, W. A. Pearlman, and J. W. Modestino, "Hyperspectral image compression using three-dimensional wavelet coding," in *Proc. SPIE/IS&T Electron. Imaging*, 2003, pp. 1037–1047.
[7] A. Islam and W. A. Pearlman, "An embedded and efficient low-complexity hierarchical image coder," in *Proc. SPIE—Visual Comm. Image Processing*, 1999, vol. 3653, pp. 294–305.
[8] J. Wu, Z. Wu, and C. Wu, "Lossy to lossless compressions of hyperspectral images using three-dimensional set partitioning algorithm," in *Proc. SPIE—Electronic Imaging and Multimedia Technology*, 2005, pp. 484–493.
[9] H. S. Lee, N. H. Younan, and R. L. King, "Hyperspectral image cube compression combining JPEG-2000 and spectral decorrelation," in *Proc. IGARSS*, Jun. 2002, pp. 3317–3319.
[10] B. Penna, T. Tillo, E. Magli, and G. Olmo, "Embedded lossy to lossless compression of hyperspectral images using JPEG 2000," in *Proc. IGARSS*, Jul. 2005, pp. 140–143.
[11] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.
[12] W. Sweldens, "The lifting scheme: A new philosophy in biorthogonal wavelet constructions," in *Proc. SPIE Conf.*, 1995, vol. 2569, pp. 68–79.
[13] A. R. Calderbank, I. Daubechies, W. Sweldens, and B. L. Yeo, "Lossless image compression using integer to integer wavelet transforms," in *Proc. ICIP*, Sep. 1997, pp. 596–599.
[14] J. Zhang and G. Liu, "A novel lossless compression for hyperspectral images by adaptive classified arithmetic coding in wavelet domain," in *Proc. ICIP*, Oct. 2006, pp. 2269–2272.
[15] I. H. Witten, R. M. Neal, and J. G. Cleary, "Arithmetic coding for data compression," *Commun. ACM*, vol. 30, no. 6, pp. 520–540, Jun. 1987.
[16] J. Zhang and G. Liu, "An efficient reordering prediction based lossless compression algorithm for hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 2, pp. 283–287, Apr. 2007.
[17] B. Kim, Z. Xiong, and A. Pearlman, "Low bit rate, scalable video coding with 3-D set partitioning in hierarchical trees (3-D SPIHT)," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 8, pp. 1374–1387, Dec. 2000.
[18] E. Magli, G. Olmo, and E. Quacchio, "Optimized onboard lossless and near-lossless compression of hyperspectral data using CALIC," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 1, pp. 21–25, Jan. 2004.