

# Hybrid Compression of Hyperspectral Images Based on PCA with Pre-encoding Discriminant Information

Chulhee Lee, Sungwook Youn, Taek Jeong, Eunjae Lee, and Joan Serra Sagristà

**Abstract**— It has been shown that image compression based on principal component analysis (PCA) provides good compression efficiency for hyperspectral images. However, PCA might fail to capture all the discriminant information of hyperspectral images since features that are important for classification tasks may not be high in signal energy. To deal with this problem, we propose a hybrid compression method for hyperspectral images with pre-encoding discriminant information. A feature extraction method is first applied to the original images, producing a set of feature vectors that are used to generate feature images and then residual images by subtracting the feature-reconstructed images from the original ones. Both feature images and residual images are compressed and transmitted. Experiments on data from airborne AVIRIS sensor indicate that the proposed method provides better compression efficiency with improved classification accuracy than conventional compression methods.

**Index Terms**— compression, hyperspectral images, principal component analysis, feature images, residual images, discriminant information.

## I. INTRODUCTION

THE advancement of sensor technology produces remotely-sensed data that have a large number of spectral bands [1]. There is an increasing need for efficient compression techniques for these hyperspectral images [2].

Many researchers have studied lossless and lossy compression techniques, where the latter can achieve higher compression ratios than the former. Since adjacent bands of hyperspectral data are highly correlated, most compression techniques use this property to remove spectral redundancy: most lossless techniques resort to prediction, while most lossy techniques resort to transform-based approaches. In particular, in transform-based methods, principal component analysis (PCA) (also known as Karhunen-Loève transform (KLT)) [3] has been commonly used, often followed by two-dimensional transforms such as the discrete wavelet transform (DWT) [4] or the discrete cosine transform (DCT) [5]. Wavelet transform-based methods have drawn great interest too and a number of 2D wavelet-based techniques have been extended to

3D applications, including SPIHT [6], SPECK [7], and tarp coding [8]. Also, JPEG2000 standard (Annex I, Part 2) allows compressing hyperspectral images with arbitrary spectral decorrelation. Region based coding schemes have been studied too; often yielding improved SNR performance [9-10].

Most lossy compression methods have been developed to minimize mean squared errors between the original and the reconstructed pixels. However, discriminant information required to distinguish between the various classes is also vital for classification purposes applications [11]. As an example, JPEG2000 coders coupled with spectral PCA produce good performance in terms of SNR [12-13], but their classification accuracy may not be satisfactory [14] since they may not effectively preserve the discriminant features for classification, mostly because these features may not be large in terms of energy.

In this paper, we propose a hybrid compression method that takes into account the discriminating information of hyperspectral images. First, we apply a feature extraction method to obtain feature images, which are then used to generate feature-reconstructed images. These feature-reconstructed images are subtracted from the original images to produce residual images. The feature images and some eigenimages of the residual images are compressed using conventional compression techniques.

The rest of this paper is organized as follows: in Section II, the feature extraction and the feature images are introduced, along with the proposed PCA-based compression method with pre-encoding discriminant information. In Section III, experimental results and discussions are provided. Finally, some conclusions are drawn in Section IV.

## II. HYBRID COMPRESSION METHOD

### A. Feature extraction and feature images

Linear feature extraction can be viewed as a linear transform. The feature extraction method produces a set of feature vectors,  $\{\beta_i\}$ , and an extracted-feature is computed as follows:

$$y_i = \beta_i^T X, \quad (1)$$

where  $X$  represents an observation in the  $N$ -dimensional

This work was supported by grant No. (R01-2006-000-11223-0) from the Basic Research Program of the Korea Science & Engineering Foundation, and by FEDER, the Spanish Government (MINECO), and the Catalan Government under grants TIN2012-38102-C03-03 and 2014SGR-691.

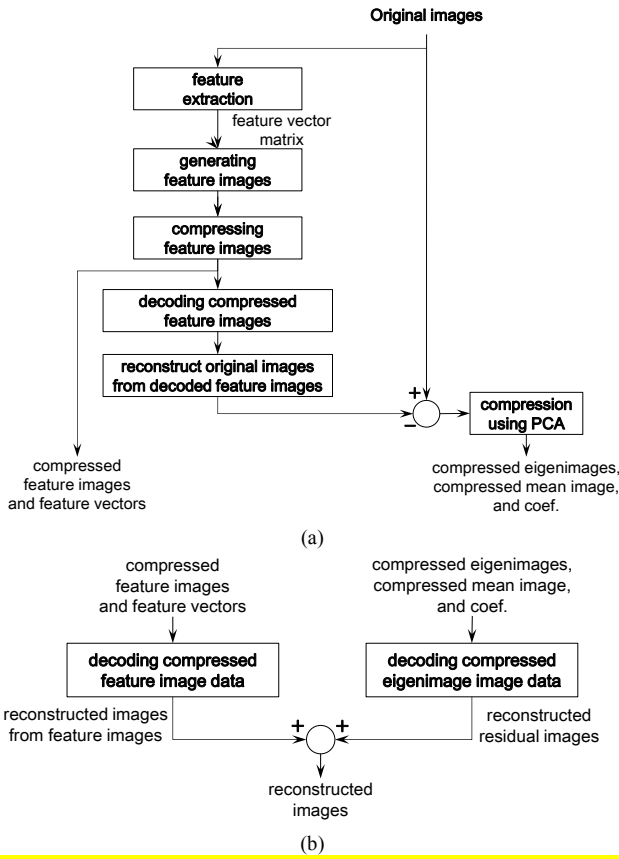


Fig. 1. Block diagram of the proposed compression method. (a) Encoding. (b) Decoding.

space. In most cases, the set of feature vectors  $\{\beta_i\}$  can be considered orthonormal. A number of feature extraction methods have been proposed for pattern classification in the past, including canonical analysis [15] and decision boundary feature extraction (DBFE) [16]. In this paper, we select DBFE because it can utilize both the mean and covariance differences; however, any other feature extraction method can be used for the proposed compression method.

Most feature extraction methods, including canonical analysis and DBFE, use covariance matrices, which should be invertible. Due to high correlations between adjacent bands, the covariance matrix of hyperspectral images may not be invertible even with a large number of training samples. To deal with this problem, we used the band combination procedure [17] and the band expansion method [18].

### B. Proposed encoding method with discriminant information preserved

Fig. 1(a) shows a block diagram of the proposed compression method. We assumed that the original images contained  $N$  spectral bands and  $K$  pixels in each band. We let  $J_1, J_2, \dots, J_N$  be the  $N$  spectral bands where  $J_i$  was a  $K \times 1$  column vector. For notational convenience, these vectors were presented in a  $K \times N$  matrix, i.e.,  $J = [J_1, J_2, \dots, J_N]$ . First, we apply feature extraction to the original images based on a set of given classes. This process will produce a set of feature vectors.

It is assumed that the set of feature vectors forms an orthonormal basis. In this case, the following feature vector matrix can be viewed as a unitary transform:

$$B = [\beta_1, \beta_2, \dots, \beta_N] \quad (2)$$

where  $\beta_i$  is a  $N \times 1$  column vector.

Let  $X$  be a pixel vector of the original image, which corresponds to each column vector of  $J^T$ . Then, we can represent  $X$  as a linear combination of  $\{\beta_i\}$ :

$$X = \sum_{j=1}^N y_j \beta_j \quad (3)$$

where  $y_j = \beta_j^T X$  represents an extracted-feature found by a feature extraction method. In most cases, it is possible to retain most of the discriminant information with a small number of extracted features.

Using these extracted features, a subset of extracted feature images is produced as follows:

$$\begin{aligned} F_i &= J\beta_i = [J_1, J_2, \dots, J_N]\beta_i \\ &= \beta_{i,1}J_1 + \beta_{i,2}J_2 + \dots + \beta_{i,N}J_N \quad (i = 1, \dots, S) \end{aligned} \quad (4)$$

where  $\beta_{i,j}$  represents the  $j^{\text{th}}$  component of  $\beta_i$ ,  $F_i$  is the  $i^{\text{th}}$  extracted feature image and  $S$  is the number of extracted features. From now on, we will refer to these extracted-feature images as feature images. These  $S$  feature images are encoded and the corresponding feature vectors are quantized. Any 2D or 3D image compression method can be used to compress the feature images, but we used the *1D+2D JPEG2000* approach due to its superior compression performance. From the encoded feature images and feature vectors, the original images can be reconstructed. In other words, a pixel vector of the original image can be reconstructed as follows:

$$\hat{Z} \approx \sum_{j=1}^S \hat{y}_j \hat{\beta}_j \quad (5)$$

where  $\hat{y}_j$  represents the pixel value of the  $j$ -th feature image. Thus, we reconstructed the original images from the decoded feature images, which are called “feature-reconstructed” images and denoted by  $\hat{J}^{\text{feature}}$ . By subtracting the feature-reconstructed images from the original images, we generated residual images:

$$J_i^{\text{res}} = J_i - \hat{J}_i^{\text{feature}} \quad (i = 1, \dots, N) \quad (6)$$

PCA was applied to these residual images and a subset of eigenimages was compressed with *1D+2D JPEG2000* in this paper. In other words, each band of a residual image is

TABLE I  
CLASS DESCRIPTION

- Sub-region1		
Class Species	No. samples	No. training samples
Buildings	375	188
Corn	819	410
Corn-CleanTill	966	483
Corn-CleanTill-EW	1320	660
Grass	172	86
Hay	231	116
NotCropped	480	240
Pasture	483	242
Soybeans	836	418
Soybeans-CleanTill	1050	525
Soybeans-CleanTill-EW	1722	861
Water	189	95
Trees	330	165
Wheat	940	470
Woods	252	126

represented as follows:

$$J_i^{res} \approx \sum_{j=1}^n c_{j,i} \phi_j + \hat{\mu} \quad (i = 1, \dots, N) \quad (7)$$

where  $\{\phi_i\}$  is an eigenvector ( $K \times 1$ ) of the covariance matrix of the residual images as calculated in [19],  $\hat{\mu}$  represents a reconstructed mean image ( $K \times 1$ ) and  $c_{j,i} = \phi_i^T (J_j^{res} - \hat{\mu})$ .  $\{\phi_i\}$  represents the eigenimages of the residual images and the coefficients  $\{c_{j,i}\}$  are quantized. From the compressed eigenimages and quantized coefficients, we can produce reconstructed images:

$$\hat{J}_i^{res} = \sum_{j=1}^n \hat{c}_{j,i} \hat{\phi}_j + \hat{\mu} \quad (i = 1, \dots, N) \quad (8)$$

where  $\hat{J}_i^{res}$  represents the reconstructed  $i$ -th residual image,  $\hat{c}_{j,i}$  a quantized coefficient and  $\hat{\phi}_i$  the reconstructed  $i$ -th eigenimage. Finally, the hyperspectral images are reconstructed by adding the reconstructed feature images to the reconstructed residual images (Fig. 1(b)). In the following experiments, 32 bits were used for quantizing each coefficient since the memory requirements for coefficient quantization were negligible compared to the memory requirements for image data and the mean image ( $\hat{\mu}$ ) was compressed at 3.46 bits per pixel.

### III. EXPERIMENTAL RESULTS

The Indian Pine AVIRIS data set was used in the experiments [20]. This data set contains 220 spectral bands and the spatial size of the bands is  $2166 \times 614$  pixels. From the data set, a sub-region of  $256 \times 256$  pixels (sub-region1) has been widely used by remote-sensing researchers. The selected sub-region is shown in Fig. 2 (50<sup>th</sup> band). We chose 15 classes in sub-region1; Table I reports the class information.

The Gaussian ML classifier was used, assuming a Gaussian

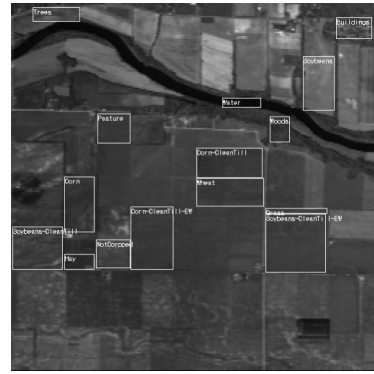


Fig. 2. AVIRIS image with selected classes. Sub-region1 ( $256 \times 256$ , 50<sup>th</sup> band).

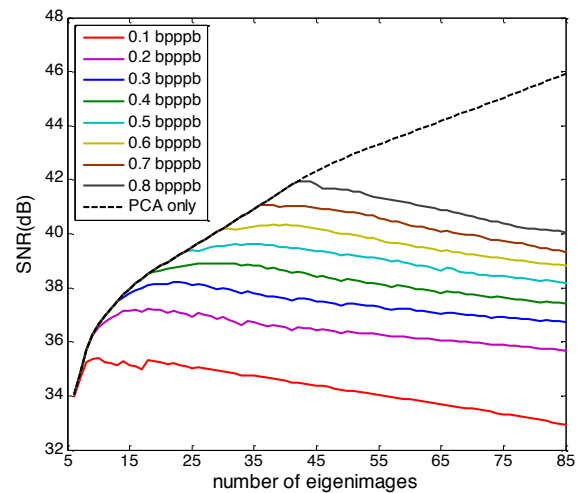


Fig. 3. SNR performance as the number of eigenimages varies for subPCA/1D+2D JPEG2000. Results for AVIRIS sub-region1 image.

TABLE II  
NUMBER OF EIGENIMAGES FOR EACH BIT-RATE

bpppb	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
Number of Eigenimages	7	17	21	29	33	39	35	41

distribution for each class [15]. To avoid the singularity problem of covariance matrices and the Hughes phenomenon [21], 220 bands were reduced to 20 features by combining adjacent bands for classification purposes [17].

To evaluate the performance of our proposed technique, a comparison with several other coding techniques had to be performed, namely we investigated the 1D+2D JPEG2000 approach when the spectral transform is carried out by a DWT 3 levels (Kakadu v7 implementation with 4 levels in the spatial domain has been used). The 1D+2D JPEG2000 method was applied to compress both the original image and a subset of eigenimages. Fig. 3 shows the SNR performance for various bitrates and numbers of encoded eigenimages. It can be seen that increasing the number of eigenimages did not always improve SNR performance. In this paper, the number of eigenimages at each bitrate was selected to have maximum SNRs. Table II shows the optimal number of eigenimages for each bit-rate.

We used SNR and classification accuracy to evaluate the

**TABLE III**  
**VARIOUS RATIO OF BIT ALLOCATION TO FEATURE IMAGES AND EIGENIMAGES**

			Proposed method with 20 feature images (ratio, X(feature):Y(eigenimage))									
bpppb	1D+ 2D JPEG 2000	subPCA/ 1D+ 2D JPEG 2000	9:1	7:1	5:1	3:1	1:1	1:3	1:5	1:7	1:9	
SNR (dB)	0.1	31.48	<b>34.92</b>	24.53	24.57	24.63	26.51	30.33	32.71	33.29	33.60	<b>33.69</b>
	0.2	34.52	<b>36.95</b>	29.02	29.92	31.11	32.60	34.77	35.85	36.10	36.16	<b>36.20</b>
	0.3	36.30	<b>37.98</b>	31.57	32.32	33.22	34.43	36.16	37.13	37.24	37.32	<b>37.35</b>
	0.4	37.51	<b>38.68</b>	33.35	34.04	34.80	35.85	37.13	37.85	38.06	38.09	<b>38.11</b>
	0.5	38.64	<b>39.36</b>	34.33	34.97	35.69	36.48	37.83	38.66	38.77	<b>38.78</b>	38.73
	0.6	39.55	<b>40.06</b>	35.11	35.72	36.23	36.98	38.43	39.26	39.40	<b>39.43</b>	39.40
	0.7	40.46	<b>40.74</b>	35.52	36.02	36.48	37.33	39.01	39.92	40.05	<b>40.08</b>	40.05
	0.8	41.22	<b>41.58</b>	36.08	36.45	37.06	37.83	39.57	40.56	40.73	40.73	<b>40.73</b>
Acc (%)	0.1	<b>94.48</b>	82.14	97.09	<b>97.22</b>	97.14	97.08	96.68	96.37	96.55	96.75	97.22
	0.2	<b>91.60</b>	87.98	97.67	<b>97.95</b>	97.45	96.53	94.80	95.16	95.63	95.47	95.60
	0.3	<b>90.69</b>	88.91	<b>97.51</b>	96.30	96.18	94.87	93.71	94.25	94.73	95.36	95.66
	0.4	<b>90.89</b>	89.28	<b>96.31</b>	96.04	94.70	93.60	93.12	94.28	94.58	95.13	95.42
	0.5	<b>91.50</b>	89.68	<b>95.34</b>	94.90	93.85	92.41	92.51	93.78	94.73	94.99	95.21
	0.6	<b>90.96</b>	90.21	94.05	93.95	93.14	92.41	92.08	93.92	94.44	94.99	<b>95.08</b>
	0.7	91.14	<b>91.38</b>	94.40	93.61	92.66	92.45	92.16	93.39	94.55	94.93	<b>95.37</b>
	0.8	91.26	<b>91.97</b>	93.98	92.84	92.93	92.24	92.16	93.35	94.29	94.80	<b>95.33</b>

compression methods. In the experiments, the bit-rate was defined on a per-band basis, i.e., 1 bpppb indicates 1 bit per pixel per band. The SNR was computed as follows:

$$SNR = 10 \log_{10} \frac{\sigma^2}{MSE}, \quad (9)$$

where  $MSE = E[(x - \hat{x})^2]$  and  $\sigma^2$  represents the variance of the selected image. The value of  $\sigma^2$  was approximately  $1.778 \times 10^6$  for AVIRIS sub-region1. The classification accuracy was computed using test data that were not used to estimate parameters. In the experiments, half of the available samples were used for training and the other half were used for test purposes (see Table I).

Table III presents the SNR and classification accuracy when the bit distribution ratio varied from 9:1 (feature image: eigenimage) to 1:9. Results reported in Table III were obtained with 20 feature images, although a similar performance was observed for different numbers of feature images. At low bit rates (0.1~0.4 bpppb), the classification accuracy was highest when most bits were allocated to feature images (e.g., 9:1) while SNR was poor. On the other hand, the 1:9 ratio produced good performance in SNR and acceptable classification accuracy. At high bit rates (0.5~0.8 bpppb), the classification accuracy was highest when most bits were allocated to eigenimages (e.g., 1:9). The 1:9 ratio also produced the best SNR performance. Thus, we used the 1:9 ratio for both low and high bit rates.

Fig. 4 shows a performance comparison of the SNRs (Fig. 4(a)) and classification accuracy (Fig. 4(b)) when different numbers of feature images (10, 15, 20 and 25) were used and the 1:9 ratio was applied. As the number of feature images increased, the SNRs of the proposed method decreased slightly while the classification accuracy increased. The subPCA/1D+2D JPEG2000 showed the best SNR performance. The SNRs of the proposed method showed the second best

performance. On the other hand, the classification accuracy of the proposed method noticeably improved compared to 1D+2D JPEG2000 and subPCA/1D+2D JPEG2000 for all bit rates. Fig. 5 shows the classification results for each class of the proposed method (20 feature images), the subPCA/1D+2D JPEG2000 method and the 1D+2D JPEG2000 method at 0.4 bpppb. For some classes, the improvements were substantial. For example, for Soybeans-CleanTill, the classification accuracy of the proposed method was 92.1% while the classification accuracy values of the subPCA/1D+2D JPEG2000 and 1D+2D JPEG2000 were 72.7% and 77.3%, respectively. Also, the classification accuracy of the Corn-CleanTill improved from 82.1% (subPCA/1D+2D JPEG2000) and 83% (1D+2D JPEG2000) to 93.5%.

#### IV. CONCLUSIONS

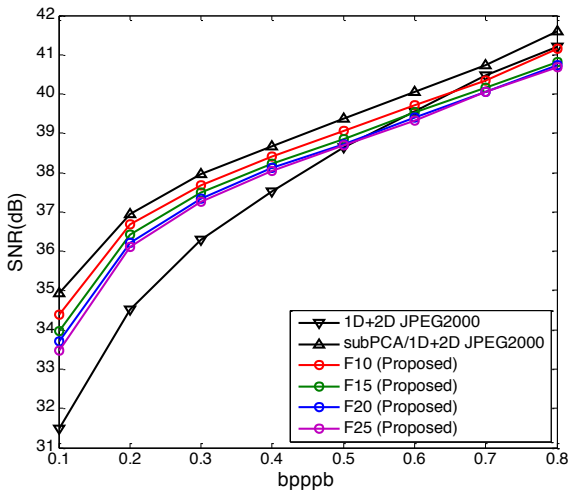
In this paper, we proposed a hybrid compression method for hyperspectral images based on PCA along with encoding residual discriminant information. We first applied a feature extraction method to obtain the feature images and then encoded the hyperspectral images using the dominant feature images. Then we generated residual images by subtracting the reconstructed images from the original images. By applying PCA to the residual images, we generated the eigenimages, which were also encoded. The dominant eigenimages and discriminant feature images were compressed using conventional image compression techniques to achieve better performance. Experiments with AVIRIS showed that the proposed method produces better compression efficiency with improved classification accuracy than existing compression methods such as 1D+2D JPEG2000 and subPCA/1D+2D JPEG2000.

#### REFERENCES

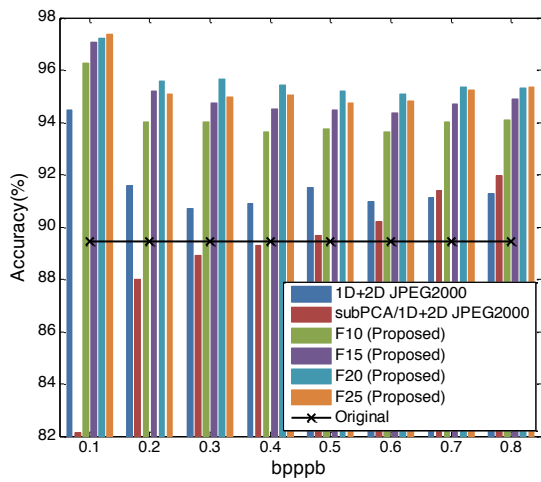
- [1] P. M. Mather, *Computer Processing of Remotely-Sensed Images*, England: John Wiley & Sons Ltd, 2nd edition 1999.

[2] S.-E. Qian, Optical Satellite Data Compression and Implementation. Bellingham, WA: SPIE—Int. Soc. Opt. Eng., 2013.  
 [3] I. Blanes and J. Serra-Sagrìstà, “Cost and scalability improvements to the Karhunen-Loève transform for remote-sensing image coding,” *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 7, pp. 2854–2863, Jul. 2010.

[4] T. Markas and J. Reif, “Multispectral image compression algorithms,” *Data Compression Conference DCC '93*, pp. 391-400, 1993.  
 [5] S. Lim, K. Sohn and C. Lee, “Compression for hyperspectral images using three dimensional wavelet transform,” *IGARSS 2001*, vol. 1, pp. 109-111, 2001.  
 [6] 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, 2000.  
 [7] X. Tang and W. A. Pearlman, “Three-dimensional wavelet-based compression of hyperspectral images,” in *Hyperspectral Data Compression*, G. Motta, F. Rizzo, and J. A. Storer, Eds. Norwell, MA: Kluwer Academic Publishers, 2006, ch. 10, pp. 273–308.  
 [8] Y. Wang, J. T. Rucker and J. E. Fowler, “Three-dimensional tarp coding for the compression of hyperspectral images,” *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 2, pp. 136–140, 2004.  
 [9] M. Cagnazzo, G. Poggi and L. Verdoliva, “Region-based transform coding of multispectral images,” *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 2916–2926, 2007.  
 [10] L. Chang, Y.-L. Chang, Z. S. Tang and B. Huang, “Group and region based parallel compression method using signal subspace projection and band clustering for hyperspectral imagery,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 3, pp.565 -578, 2011.  
 [11] S. Qian, Optical Satellite Signal Processing and Enhancement. SPIE, Nov. 2013.  
 [12] 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, 2007.  
 [13] B. Penna, T. Tillo, E. Magli, and G. Olmo, “Transform coding techniques for lossy hyperspectral data compression,” *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1408–1421, May 2007.  
 [14] F. García-Vilchez, J. Muñoz-Mari, M. Zortea, I. Blanes, V. G. Ruiz, G. Camps-Valls, A. Plaza, and J. Serra-Sagrìstà, “On the impact of lossy compression on hyperspectral image classification and unmixing,” *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 2, pp. 253–257, 2011.  
 [15] J. A. Richards, Remote Sensing Digital Image Analysis. Berlin, Germany: Springer-Verlag, 1993.  
 [16] C. Lee and D. A. Landgrebe, “Feature extraction based on the decision boundaries,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, no. 4, pp. 388-400, 1993.  
 [17] C. Lee and D. A. Landgrebe, “Analyzing high dimensional multispectral data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 31, no. 4, pp. 792-800, 1993.  
 [18] G. Strang, Linear Algebra and Its Applications. Tokyo, Japan: Third Edition. Harcourt Brace Jovanovich, Inc., 1988.  
 [19] S. Lim, K. Sohn and C. Lee, “Principal component analysis for compression of hyperspectral images,” *IGARSS 2001*, vol. 1, pp. 97-99, 2001.  
 [20] G. Vane, R. O. Green, T. G. Chrien, H. T. Enmark, E. G. Hansen and W. M. Porter, “The airborne visible/infrared imaging spectrometer (AVIRIS),” *Remote Sensing of Environment*, vol. 44, pp. 127-143, 1993.  
 [21] G. F. Hughes, “On the mean accuracy of statistical pattern recognizers,” *IEEE Trans. Inform. Theory*, vol. IT-14, no. 1, pp. 55-63, 1968.



(a)



(b)

Fig. 4. Performance comparison using different numbers of feature images in sub-region1. (a) SNR. (b) Accuracy.

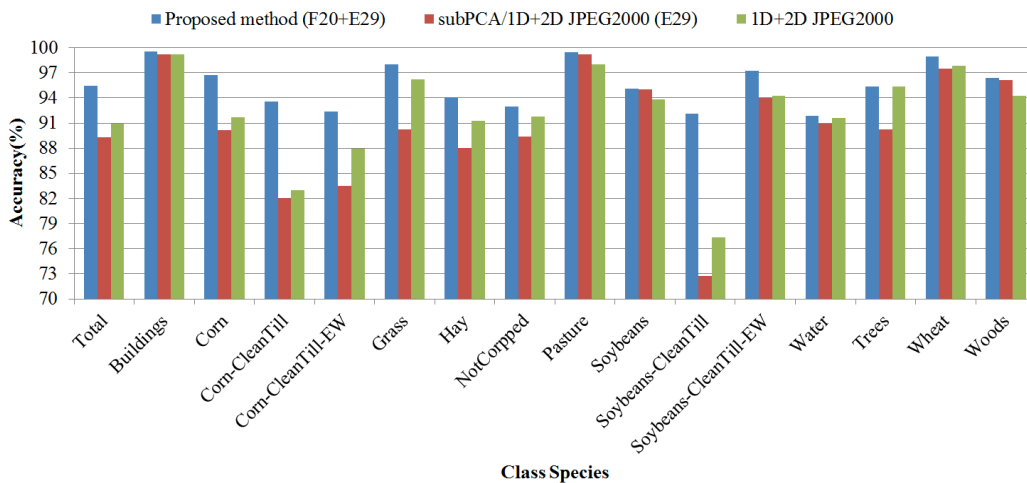


Fig. 5. Performance comparison of accuracy results in each class.