# Improved Class-Based Coding of Multispectral Images With Shape-Adaptive Wavelet Transform

M. Cagnazzo, S. Parrilli, G. Poggi, and L. Verdoliva

***Abstract*—In this letter, we improve the class-based transform-coding scheme proposed by Gelli and Poggi for the compression of multispectral images. The original spatial-coding tools, 1-D discrete cosine transform and scalar quantization, are replaced by shape-adaptive wavelet transform and set partitioning in hierarchical trees. Numerical experiments show that the improved technique outperforms the original one for medium- to high-quality compression and is consistently superior to all reference techniques.**

***Index Terms*—Compression, multispectral, segmentation, shape-adaptive wavelet transform (SAWT).**

## I. Introduction

MULTISPECTRAL images are characterized by better and better spatial, spectral, and radiometric resolution and, hence, by ever-increasing demands of communication and storage resources. Often, such demands exceed the system capacity like, for example, in the downlink from satellite to Earth stations, where the channel bandwidth is often much inferior to the intrinsic data rate of the images, some of which must be discarded altogether. In this situation, as well as in many others, high-fidelity compression of the images represents a very appealing alternative. As a matter of fact, there has been intense research activity on this topic in the last few years [1]–[7], focusing, particularly, on transform-coding techniques, due to their good performance and limited computational complexity. Linear transform coding, however, does not take into account the nonlinear dependences existing among different bands, due to the fact that multiple land covers, each with its own interband statistics, are present in a single image. Based on this observation, a class-based coder (referred to, from now on, as CBC) was proposed in [1] that addresses the problem of interband dependences by segmenting the image into several classes, corresponding as much as possible to the different land covers of the scene. As a consequence, within each class, pixels share the same statistics and exhibit only linear interband dependences, which can be efficiently exploited by conventional transform coding.

The coding scheme is summarized in Fig. 1. The segmentation is carried out by means of tree-structured vector quantization (TSVQ), and the resulting map is encoded without loss of information. TSVQ segmentation also provides a rough encoding of the image through the class means, which are subtracted from the data. All residuals of the same class are then pooled together and are subject to a Karhunen–Loeve transform (KLT) along the spectral dimension, discrete cosine transform (DCT) in the spatial domain, and, finally, scalar quantization of the coefficients with optimal bit allocation. To take into account the different spectral statistics of the classes, a different KLT matrix is used for each class in order to better compact energy in the first few transform bands.

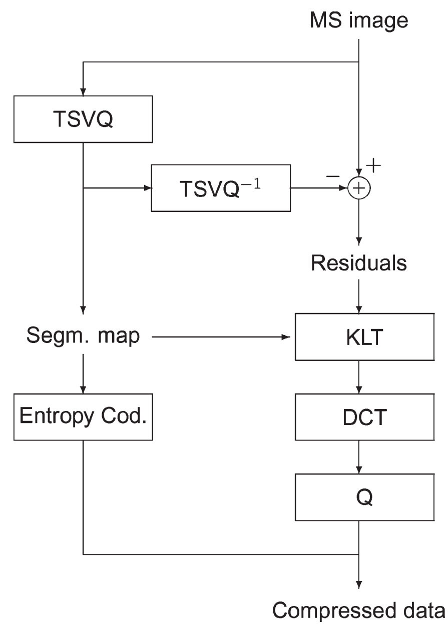



Fig. 1. Original class-based coding scheme.

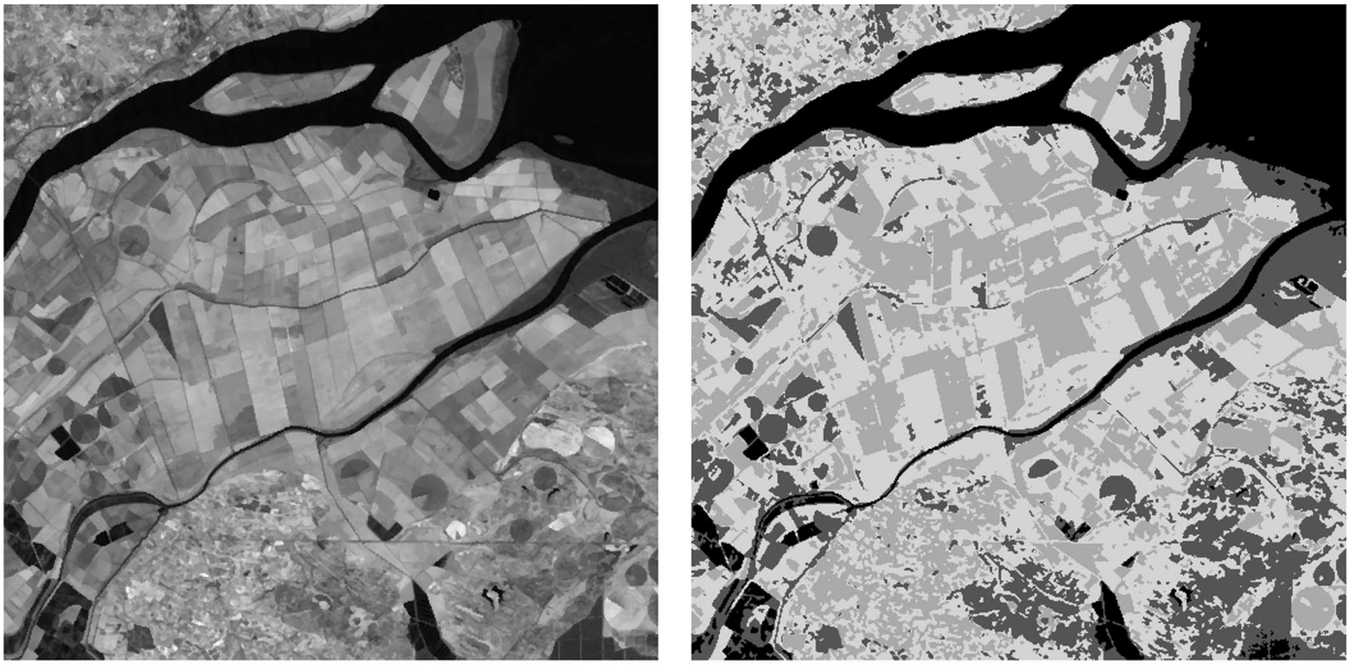
Experiments on several multispectral test images proved this technique to have an excellent rate-distortion performance. As a running example, we consider a Landsat Thematic Mapper (TM) image (six bands, 512 × 512 pixels, 8 bits/sample) of a region near Lisbon, a band of which is shown in Fig. 2(a). Fig. 3 reports the rate-distortion (RD) curves (SNR versus coding rate in bits per sample) obtained by CBC when $C = 2, 4$, and 8 classes are used in the segmentation and subsequent phases. For comparison, the curve obtained without segmentation is also reported. We see that the CBC significantly outperforms the reference "flat" coder, particularly when a relatively large number of classes is used. CBC curves start from an unfavorable rate-distortion point due to the cost of the segmentation

Manuscript received August 4, 2006; revised April 23, 2007.

The authors are with the Dipartimento di Ingegneria Elettronica e delle Telecomunicazioni, Università Federico II, 80125 Napoli, Italy (e-mail: poggi@unina.it).

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/LGRS.2007.900696



(a) (b)

Fig. 2. Landsat TM test image. (a) Band five. (b) Four-class segmentation.

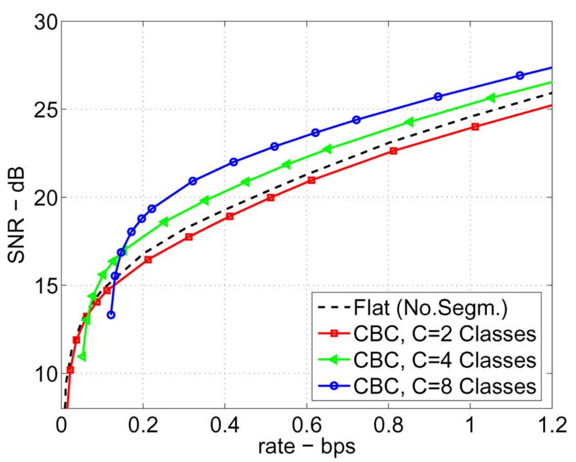



Fig. 3. RD performance of the original DCT-based coder.

map, but then, the SNR grows rapidly because the classified data are much more homogeneous and easier to encode. It is worth pointing out that the segmentation map is, by itself, a valuable piece of information for the end user and is included automatically in the coded stream.

The weakest point of this technique is in the spatial-domain processing, i.e., the DCT of KLT bands, and the subsequent encoding of the transform coefficients. In fact, since TSVQ produces segmentation maps with arbitrary geometry, classes are typically composed by a large number of unconnected regions with irregular shape, as shown, for example, in Fig. 2(b) for the case of four classes, and hence, one cannot encode them by simple bidimensional DCT. After testing the various viable encoding strategies, including shape-adaptive DCT, it resulted that collecting all pixels of a class in a vector and encoding it as a 1-D source provided the best performance.[1] Nonetheless, this linearization tends to destroy all residual spatial dependences in the data and, hence, to impair the overall performance.

In recent years, however, the encoding of regions with arbitrary shape has been the object of significant research efforts driven by the quest for efficient object-based video-coding techniques in the context of the MPEG-4 standard. In 2000, Li and Li [8] proposed a simple algorithm for shape-adaptive wavelet transform (SAWT), which works seamlessly with objects of arbitrary geometry. At about the same time, various algorithms were developed [9], [10] to efficiently encode the coefficients produced by SAWT, with very promising results. We used some of these tools for the compression of multispectral images in [6] and [7], where, unlike in [1], the focus was on the encoding of compact individual regions rather than classes and only global spectral transforms were used. Given the encouraging results, we now want to use these new tools to improve the spatial-domain processing phase of our class-based coding technique so as to take full advantage of the potential of this approach. Of course, our class-based segmentation maps, with a very large number of nonconnected segments, can severely challenge the efficiency of SAWT and coding, so we have to carefully assess this new technique in real-world situations. In the next section, we describe briefly the new coding scheme, highlighting the key points to be analyzed, and, in Section III, report the results of some numerical experiments and draw conclusions.

## II. Wavelet-Based CBC

We want to improve the class-oriented transform-coding technique proposed in [1] by acting only on the spatial-domain

[1]Similar results have been later found independently in [4].

processing. Therefore, the structure of the coding scheme is the same as that of Fig. 1 except for the fact that some blocks are updated with the introduction of new technical solutions. We keep using TSVQ to carry out segmentation because of its very low complexity and good rate-distortion performance. As for the map coding, we will resort to the algorithm proposed in [11] for palette images, which proves very efficient, even though this step has a limited impact anyway on the overall performance. Class-adaptive KLT will not be changed as it stands at the core of the class-oriented approach. The major innovations take place in the two spatial-domain coding blocks, where DCT (i.e., data linearization followed by 1-D-DCT) is replaced by SAWT and scalar quantization by shape-adaptive set partitioning in hierarchical trees (SPIHT) coding (SA-SPIHT) [9], which is a version of the well-known bit-plane coder SPIHT [12], adapted to work on arbitrary-geometry support. Although both techniques are already described in the literature, there are some points that merit discussion.

SAWT presents a number of desirable properties. The number of transform coefficients equals the number of pixels in the object; all spatial relationships are preserved; and it falls back to ordinary WT when the object is rectangular. However, it was developed having in mind the encoding of video objects and, hence, has been intensely tested on simple foreground/background images where only a few objects are present, possibly just two, rather large, and with relatively smooth contours. In our application, things are definitely different: in a typical segmentation map like the one shown in Fig. 2(b), an object (i.e., a class) appears as the collection of a large number of nonconnected regions, some of which are very small, even isolated points. The SAWT algorithm is extremely robust and keeps working correctly with this kind of objects, which means that one can recover the image from the coefficients with no errors. However, its energy-compaction ability might suffer significantly in these conditions,[2] with severe consequences on the overall performance. Similar considerations apply to the SA-SPIHT coder, which relies on the zero-tree concept: when an object is dispersed over most of the image, a large number of nodes must be created and information must be sent about their significance, to encode relatively small groups of pixels.

In summary, while the use of wavelet-based coding tools seems to promise some performance gain, the actual results depend on the fragmentation of the map and, eventually, on the number of classes used in segmentation. Using more classes, one obtains more homogeneous sources for subsequent transform coding but also a more fragmented map, which could lead to coding inefficiencies. This also suggests us to leave open the option of modifying the segmentation strategy, once experimental results are available, in order to produce more regular maps.

To complete the description of the coding scheme, we describe, briefly, the rate allocation block, not shown in the block diagram. After all objects are coded at high rate, their rate-distortion curves are used to carry out a Lagrangian minimization which, for each value of the Lagrange parameter, gives the number of bits required by each object. This process, inspired by the postcompression-rate-allocation algorithm of embedded block coding with optimized truncation (EBCOT) [14], allows one to optimize performance and also, if desired, to perform embedded encoding by the insertion of suitable markers in the data stream.

[2]A preliminary analysis of the efficiency of SAWT and SA-SPIHT in shape-adaptive coding appears in [13].

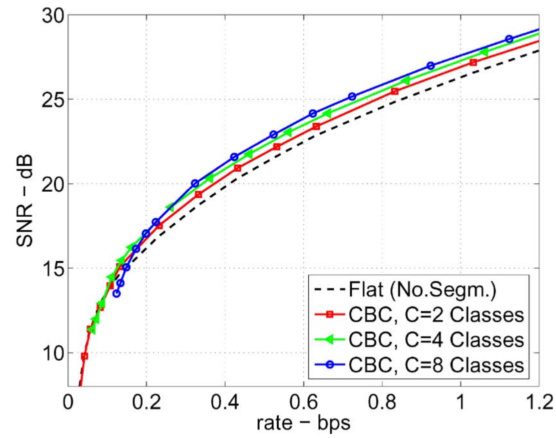



Fig. 4. RD performance of the new wavelet-based coder.

## III. Experimental Results

We carried out experiments on several multispectral images and observed always a similar behavior, but results are reported only for the TM image described in the previous section and, later on, for the “Low-Altitude” AVIRIS hyperspectral image (224 bands, 512 × 512 pixels, 16 bits/sample) available in [15].

In Fig. 4, we show the rate-distortion curves obtained with the wavelet-based version of CBC for $C = 2, 4,$ and $8$ classes and, for comparison, the curve obtained with the wavelet-based coder without segmentation. It is worth noting that this “flat” coding scheme, originally proposed in [7], can be regarded as a refinement of the 3-D-SPIHT coder that is proposed in [5], improved by using KLT instead of WT in the spectral domain (with better energy compaction) and by resorting to 2-D-SPIHT on all spectral bands with explicit rate allocation. It appears that the use of segmentation keeps providing a performance gain with respect to the flat coder, although the gap is now more limited, about 1.5 dB at best, instead of the 2–3 dB observed with the original coder.

More important, we want to compare the performance of the new wavelet-based coder with that of the original coder and, hence, report in Fig. 5 the best rate-distortion curves for both techniques $(C = 8)$, together with those of the two flat reference schemes. First of all, the wavelet-based flat coder consistently outperforms the DCT-based one, as was to be expected given the well-known superior performance of wavelet-based coding for natural images. When we add the segmentation step, i.e., considering the class-oriented coders, the comparison is not so clear-cut anymore. In particular, at low bit rates, the DCT-based coder outperforms the wavelet-based one, with an advantage of up to 2 dB at 0.2 bits/sample. The opposite happens at high bit rates, beyond 0.5 bits/sample, where the wavelet-based coder works better and better with a gain of

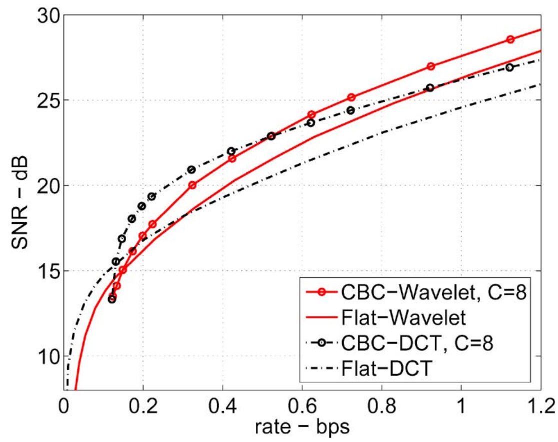



Fig. 5. Performance comparison between wavelet- and DCT-based CBC.

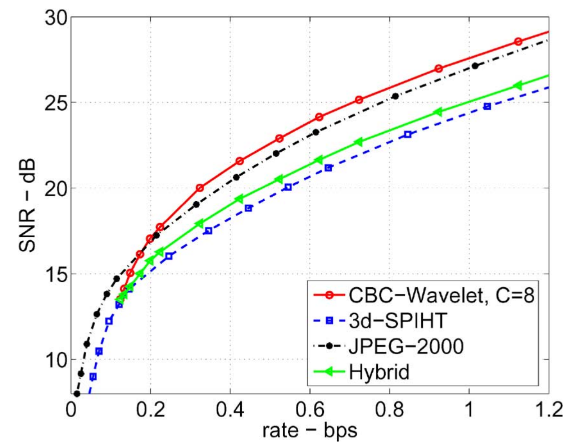



Fig. 6. Performance comparison between wavelet-based CBC and various reference coders.

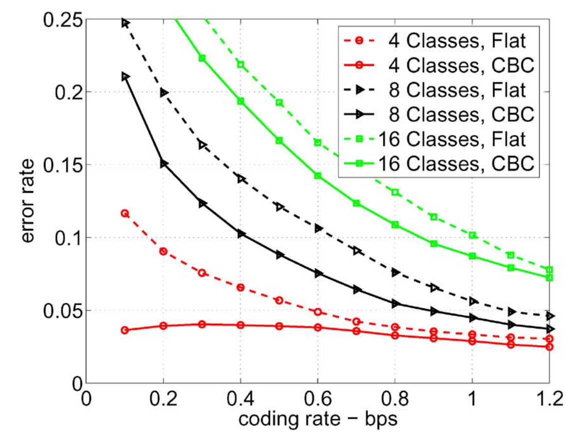



Fig. 7. Segmentation error after compression with the CBC and reference coders.

2 dB at 1.2 bits/sample. We can explain this behavior by recalling that SA-SPIHT is rather inefficient at the beginning because it deals with a very fragmented map and must spend many bits to describe significance trees for coefficients scattered over the whole image.[3] In "steady state," however, it becomes more efficient than 1-D-DCT, because spatial relationships among coefficients have been retained and can now be exploited. As a consequence, the slope of its rate-distortion curve is steeper, which makes the new scheme more and more convenient as the available coding resources grow. In summary, the new wavelet-based scheme is certainly preferable when high-quality images are desired (which happens quite often with remote-sensing imagery, where "near-lossless" compression is typically required), while the original scheme is better when resources are very limited.

We complete this analysis by comparing the performance of the proposed CBC with those of several significant reference schemes, i.e., 3-D-SPIHT [5] and JPEG-2000 [14], both implemented using the KLT as a spectral transform and, finally, a hybrid coding scheme in which the class-based approach is used to perform the spectral KLT, but then, the transformed eigenimages are coded by flat JPEG-2000 with proper rate allocation. Results are reported in Fig. 6, and show that the proposed wavelet-based CBC outperforms all reference schemes by less than 1 dB in the case of JPEG-2000 and up to 3 dB for 3-D-SPIHT. The hybrid scheme remains well under CBC and JPEG-2000, suggesting that, once segmentation is carried out, one should try to exploit the class information in all instances to make up for its initial cost.

A major concern when compressing remote-sensing images is the preservation of their diagnostic value for subsequent analysis, such as segmentation, classification, detection, etc. To provide some insight about CBC's ability to preserve data integrity, we first compressed the test image at various rates by CBC with four-class segmentation, then segmented independently each compressed image by means of a simple unsupervised clustering algorithm, the well-known K-means, and, finally, computed the error rate, i.e., the fraction of erroneously classified pixels, with respect to a reference segmentation. Lacking ground-truth data, the reference was taken as the segmentation of the original uncompressed image, even though this introduces a small negative bias. Results are reported in Fig. 7, together with those obtained with the flat-reference scheme. With four-class segmentation, CBC provides excellent results with an accuracy over 96% at all rates, while the flat coder approaches this performance only at coding rates above 1 bits/sample. It must be underlined, however, that this result is strongly biased by the nature of our coding scheme, based itself on segmentation, which allows for such a good performance even when the image quality is poor. On the other hand, the presence of a segmentation map embedded in the data is actually a major strength of our approach, and it could even be used explicitly (unlike in this experiment) to improve accuracy. With 8- and 16-class segmentation, instead, compression and segmentation are clearly decoupled, but CBC keeps providing a performance significantly better than flat coding, with an accuracy of about 96% and 93% for 8 and 16 classes, respectively, at the higher rates.

Finally, we present compression results for a different source, the AVIRIS hyperspectral image "Low Altitude." We use only 192 out of the total 224 bands, since some bands are clearly

[3]To improve performance at low rates, we also experimented with smoother segmentation maps, obtained through morphological filtering of the original maps, but overall results, not reported here for brevity, were disappointing, so we are not going to consider this option anymore.

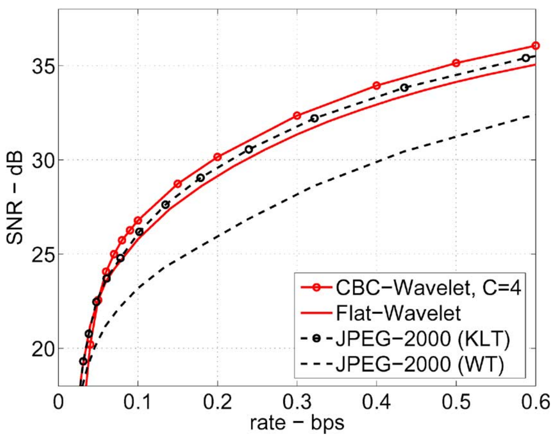



Fig. 8. RD curves for the AVIRIS image.

useless and must be discarded altogether. For this image, it is also convenient to normalize bands to zero mean and unit variance before encoding, since the rate-allocation procedure would otherwise give little or no resources to low-power, but informative, bands, with a detrimental effect on the image quality.[4] In Fig. 8, we report results for the CBC coder, the flat reference scheme, and JPEG-2000 multicomponent and observe a behavior quite similar to that observed for the Landsat TM image. Note that there are just a few distinct information classes in this image, and therefore, the wavelet-based CBC provides the best performance with four classes rather than eight. JPEG-2000 is quite competitive, if KLT is used to decorrelate bands prior to encoding, with a performance very close to that of wavelet-based CBC, losing just 0.6 dB at high rate. On the contrary, JPEG-2000 with WT in the spectral domain exhibits a performance gap of about 4 dB with respect to the other techniques.

In conclusion, the class-based coding approach always guarantees a very good performance, because the KLT is much more effective on homogeneous sources than on the whole image. Its weak point is the spatial transform and coding, since land covers and, hence, classes are usually scattered over the whole image. The use of SAWT and SA-SPIHT in place of 1-D-DCT, partially overcomes this problem and allows one to improve performance at medium and high bit rates, i.e., for the high-quality compression typically required by the end users.

[4]Note that the SNR obtained with this preprocessing is necessarily lower than that obtained without normalization.

## REFERENCES


[1] G. Gelli and G. Poggi, "Compression of multispectral images by spectral classification and transform coding," *IEEE Trans. Image Process.*, vol. 8, no. 4, pp. 476–489, Apr. 1999.

[2] 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.

[3] J. E. Fowler and D. N. Fox, "Embedded wavelet-based coding of three-dimensional oceanographic images with land masses," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 284–290, Feb. 2001.

[4] M. Petrou, P. Hou, S. Kamata, and C. I. Underwood, "Region-based image coding with multiple algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 3, pp. 562–570, Mar. 2001.

[5] X. Tang, W. A. Pearlman, and J. W. Modestino, "Hyperspectral image compression using three-dimensional wavelet coding," *Proc. SPIE*, vol. 5022, pp. 1037–1047, 2003.

[6] M. Cagnazzo, G. Poggi, L. Verdoliva, and A. Zinicola, "Region-oriented compression of multispectral images by shape-adaptive wavelet transform and SPIHT," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2004, pp. 2459–2462.

[7] M. Cagnazzo, G. Poggi, and L. Verdoliva, "A comparison of flat and object-based transform coding techniques for the compression of multispectral images," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2005, pp. 657–660.

[8] S. Li and W. Li, "Shape-adaptive discrete wavelet transforms for arbitrarily shaped visual object coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 5, pp. 725–743, Aug. 2000.

[9] A. Kawanaka and V. R. Algazi, "Zerotree coding of wavelet coefficients for image data on arbitrarily shaped support," in *Proc. Data Compression Conf.*, Mar. 1999, p. 534.

[10] J. E. Fowler, "Shape-adaptive coding using binary set splitting with k–d trees," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2004, pp. 1301–1304.

[11] V. Ratnakar, "RAPP: Lossless image compression with runs of adaptive pixel patterns," in *Proc. 32nd Asilomar Conf. Signals, Syst., Comput.*, Nov. 1998, pp. 1251–1255.

[12] 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.

[13] M. Cagnazzo, S. Parrilli, G. Poggi, and L. Verdoliva, "Costs and advantages of object-based image coding with shape-adaptive wavelet transform," *EURASIP J. Image Video Process.*, vol. 2007, 78 323, 2007. DOI:10.1155/2007/78323.

[14] D. Taubman, "High performance scalable image compression with EBCOT," *IEEE Trans. Image Process.*, vol. 9, no. 7, pp. 1158–1170, Jul. 2000.

[15] *Free AVIRIS Standard Data Products Web Page of the Jet Propulsion Laboratory*. [Online]. Available: http://aviris.jpl.nasa.gov/html/aviris.freedata.html