AN EFFICIENT AND HIGHLY PARALLEL HYPERSPECTRAL IMAGERY COMPRESSION SCHEME BASED ON DISTRIBUTED SOURCE CODING

Ngai-Man Cheung and Antonio Ortega

Signal and Image Processing Institute, Ming Hsieh Department of Electrical Engineering
University of Southern California, Los Angeles, CA

ABSTRACT

This paper extends our previous work on hyperspectral imagery compression based on distributed source coding (DSC). We apply DSC principles to facilitate efficient parallel encoder implementations with moderate memory requirement. Based on our previously proposed wavelet-based DSC framework, we propose a novel adaptive coding scheme that judiciously combines DSC and intra coding tools, taking into account the source statistics and inter-band correlation, as well as the coding gains and limitations imposed by tools. Bits extracted from wavelet coefficients tend to have different statistics and inter-band correlation at different significance levels and in different wavelet subbands. Therefore, it is non-trivial to determine the optimal coding strategy. Toward this we propose modeling techniques to estimate the performance of DSC/intra coding under different bits extraction scenarios. This model is used to define adaptive coding strategies that can optimally incorporate different bit-extraction techniques with DSC/intra coding tools according to the bit significance levels and wavelet subbands. Experimental results demonstrate that adaptive coding can achieve up to 4dB improvement over a non-adaptive system, and the improved DSC-based system is comparable to some 3D wavelet system in terms of coding performance. While we focus on hyperspectral images in this paper, many of the proposed techniques are applicable to other wavelet-based DSC image and video applications.

1. INTRODUCTION

Hyperspectral images consist of hundreds of spectral bands capturing the ground objects at different wavelengths. The images usually have very large raw data size, so efficient compression is necessary for practical applications. Moreover, hyperspectral images are usually captured by satellites that use embedded processors with limited resources. Therefore, encoding complexity is critical.

In a hyperspectral data-set many spectral bands are highly correlated, and various techniques have been proposed to exploit this inter-band correlation. In *inter-band prediction approaches* [1], a band is predicted using previously encoded bands and the resulting prediction residuals are encoded using standard image coding techniques. One disadvantage of inter-band prediction approaches is that they are inherently serial, since each band is encoded based on a predictor obtained from previously decoded bands. Moreover, it is difficult to achieve rate-scalability with the "closed-looped" inter-band prediction approaches. In *3D wavelet approaches*, including 3D-SPECK and 3D-SPIHT [2], inter-band correlation is exploited by performing filtering across spectral bands so that most of the signal energy is concentrated in low pass subbands (corresponding to low spatial and "cross-band" frequencies). A drawback of these methods is that they lead to complex memory management issues, and a naive implementation would require significant memory storage due to cross-band filtering.

We first proposed hyperspectral image compression algorithms based on distributed source coding (DSC) in [3, 4]. DSC [5, 6, 7, 8] allows separate and independent encoding of multiple correlated sources, and can theoretically achieve a rate as low as that of centralized encoding. Therefore, DSC tools can enable parallel and efficient hyperspectral imagery encoding. Our earlier wavelet-based system [3, 4] combines DSC techniques with the popular SPIHT algorithm [9] such that the *sign* and *refinement* bits of wavelet coefficients are DSC-coded, whereas the *significance maps* are intra-coded with *zerotree* coding. We demonstrated DSC-based hyperspectral image compression can achieve encouraging coding performance [3, 4]. In addition to our proposed wavelet-based algorithms, a pixel domain DSC system was proposed for lossless hyperspectral image compression in [10].

Based on our previously proposed wavelet-based DSC framework, this paper investigates improved algorithms to encode wavelet coefficients. Specifically, we propose a novel adaptive coding scheme that judiciously chooses whether to apply intra coding *or* DSC to encode the coefficients bits. The scheme takes into account the source statistics and inter-band correlation level of different sets of bits to encode (e.g., different subbands) and can substantially improve the coding performance of the DSC-based systems. The problem of determining the optimal combination of intra coding/DSC tools is non-trivial. Firstly, coefficients bits belonging to different wavelet subbands tend to have different statistics, and the bits at different significance levels tend to exhibit different levels of inter-band correlation. Therefore, the relative coding efficiencies of intra coding/DSC tools will depend on the bit significance level and wavelet subband. Moreover, wavelet coefficients bits can be encoded as "raw" bit-planes (e.g., each bitplane is directly encoded using a line by line scan) or the bitplane information can be split into significance/sign/refinement maps before encoding, as in SPIHT [9] or JPEG2000 [11]. While the latter approach exploits differences in statistics between refinement bits and significance maps in order to improve coding gain, it leads to some limitations on how we can apply DSC. This is because a correctly decoded significance map is needed in order to determine the position of the refinement bits, and thus if this were to be coded in DSC mode, correct decoding would need to be guaranteed in order for the refinement information to be useful. Moreover, the significance map is often encoded so that single codewords convey significance at multiple locations (e.g., zerotrees within SPIHT) leading to variable length representations. This makes it difficult to apply DSC to the encoded stream as both this stream and the corresponding side information will be of possibly different lengths. Thus, in practice, either significance is rep-

This work was supported in part by NASA-JPL.

resented without coding (and DSC is used) or significance is coded (and no DSC is used, so the significance map is in effect transmitted as intra information).

Toward determining the optimal coding strategy, we propose modeling techniques to estimate the coding gains achievable by DSC and intra coding under different bit-extraction scenarios. Based on the modeling results, the adaptive coding scheme optimally incorporates different bit-extraction techniques with DSC/intra coding tools depending on the bit significance level and wavelet subband. Experimental results demonstrate that up to 4dB improvement can be achieved by adaptive coding when compressing the NASA AVIRIS image data-sets [12]. The improved DSC-based system is comparable to some existing 3D wavelet system in terms of coding performance.

This paper is organized as follows. In Section 2 we outline the DSC-based hyperspectral image compression with the improved coding strategy. In Section 3 we examine the coding gains of intra coding/DSC tools under different bits extraction scenarios, and discuss the proposed adaptive coding scheme. Section 4 presents the experimental results, and Section 5 concludes the paper. Although we focus on hyperspectral images in this paper, many of the discussions are applicable to other wavelet-based DSC video and image systems such as [13, 14].

2. DSC-BASED HYPERSPECTRAL IMAGE COMPRESSION

Figure 1 depicts the encoding algorithm of the DSC-based hyperspectral image compression. To compress the current spectral band B_i, we apply wavelet transform and iteratively extract sign and magnitude bits from the wavelet coefficients. The extracted bits can be encoded by intra-coding (zerotree coding in our case) or DSC. In the latter case, the information shall be decoded using as side-information (SI) the same type of bits of the same significance extracted from $a\hat{B}_{i-1} + b$, where $\hat{B}_{i-1}$ is the previous adjacent reconstructed band available only at the decoder, and a and b are the linear prediction coefficients. The exact algorithms for coefficient bits extraction and compression, which are the focus of this paper, will be discussed in the next section. To determine the coding rate in the case of DSC, we need to estimate the crossover probability between the source bit-planes and their corresponding side-information. This is accomplished by a *model-based estimation*. Assuming an encoding system where each band is assigned to a different processor, the model-based estimation requires only limited inter-processor communication, which facilitates parallel encoding. Note that it is acceptable to use the original, i.e., B_{i-1} to approximate $\hat{B}_{i-1}$ in determining the coding rate because we are focusing on a high fidelity application. More details on the general framework and correlation estimation can be found in [3, 4, 15].

3. CODING STRATEGY

In this section, we examine the coding gains of intra coding/DSC tools under different bit extraction scenarios and discuss the proposed adaptive coding scheme.

Figure 2(a) illustrates a typical bit-plane by bit-plane bit extraction for wavelet coefficients. As is usually done in wavelet image compression, we extract a sign bit only when the corresponding coefficient becomes significant. The extracted sign bits can be encoded by DSC or intra coding (e.g., arithmetic coding). As will be discussed in detail in the next section, it is more efficient to encode sign

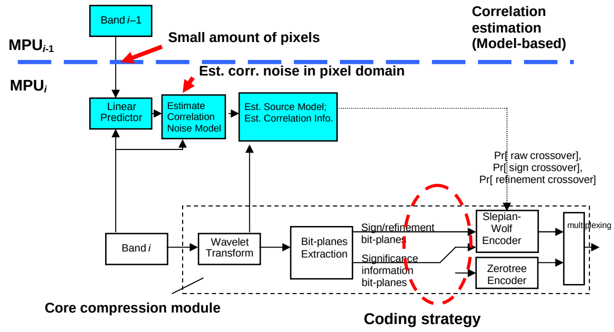
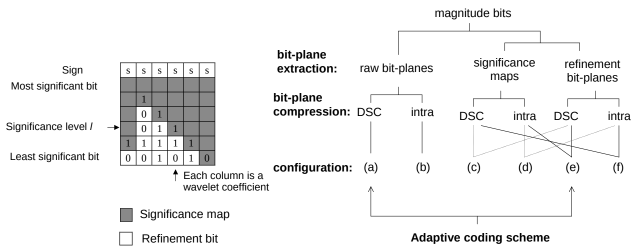


Fig. 1. The DSC-based hyperspectral image compression.

bits using DSC.



(a) Bit extraction. (b) Coding configurations.

Fig. 2. Bit extraction and compression configurations.

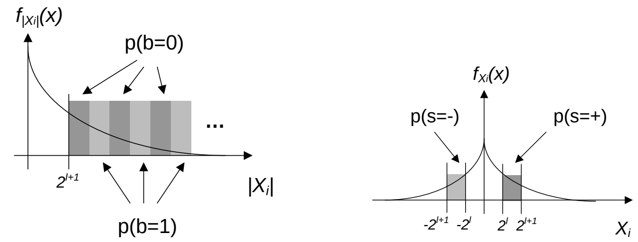
The magnitude bits can be extracted in two different ways. We can extract and encode all the magnitude bits of the same significance level (i.e., a raw bit-plane) in one single pass. Alternatively, we can partition the magnitude bits into significance maps and refinement bits, and encode them separately using DSC or intra coding (zerotree coding in our case). Therefore, it is possible to compress the magnitude bits using several different *coding configurations*, each represents a possible combination of DSC and intra coding under different bit extraction scenarios (Figure 2(b)). Our goal is to select and appropriately combine some of these configurations so that the optimal overall coding performance can be achieved. In what follows we first quickly describe why some configurations are not of practical interest and then examine the rest in more detail. As will be discussed, the optimal coding strategy involves judicious application of configurations (a) and (e) depending on bit significance levels and wavelet subbands, leading to the proposed adaptive coding scheme.

In Figure 2, both configurations (b) and (d) lead to independent (intra) encoding of each band. Generally configuration (d) is more efficient, as it exploits the differences in the zero-th order statistics of the significance maps and refinement bits (e.g., via set-partitioning in SPIHT). Thus, we will not consider raw bit-plane encoding (configuration (b)). We also eliminate (c) and (f) as both utilize DSC to encode significance maps, which could potentially result in a vulnerable system. This is because significance maps carry important structural information about the positions of significance and refinement bits. While a single error in the significance maps could lead to incorrect decoding of all the remaining bits, DSC usually has a small but non-zero probability of decoding failure. This is true in particular in our application, where, due to long delay, it may be impractical to use feedback as has been proposed in the literature [8].

In the following sections, we discuss sign bits compression by DSC or intra coding, and magnitude bits compression by configurations (a), (d) and (e). Note that the difference between (d) and (e) is in refinement bits compression and will be discussed in Section 3.1, while the differences between (a) and (e) are in significance bits extraction/compression and will be discussed in Section 3.2.

3.1. Refinement/sign bits compression

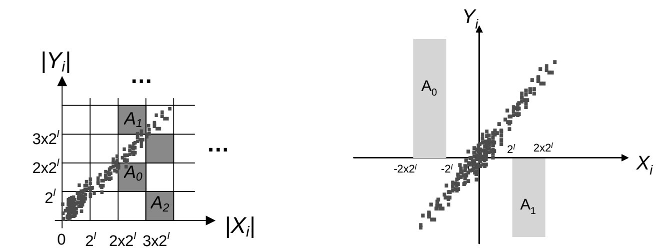
It is well known that both sign and refinement bits of typical wavelet coefficients have entropy close to one. Figures 3(a) and (b) illustrate how these refinement/sign bit probabilities can be obtained from the p.d.f. of X_i (coefficients in the ith subband). Laplacian p.d.f.'s could be chosen as typical models for the wavelet coefficients within a band. In practice this means that methods for entropy coding of refinement and sign bits achieve only relatively modest gains.



(a) Refinement bit. (b) Sign bit.

Fig. 3. Probability that a refinement/sign bit being zero or one.

On the other hand, source X_i and SI Y_i are highly correlated in hyperspectral imagery (the Y_i are ith wavelet subband coefficients of the previous spectral band after linear prediction), and samples of (X_i, Y_i) concentrate mostly near the diagonal in a scatter plot. Therefore, it is possible to compress refinement/sign bits by exploiting inter-band correlation. For refinement bits, crossover events correspond to regions denoted A_j in the sample space of X_i and Y_i in Figure 4(a), which shows that the crossover probability will tend to be small, and substantial compression of refinements bits can be achieved by using DSC [1]. Similarly, for the sign bits, crossover



(a) Refinement bit crossover. (b) Sign bit crossover.

Fig. 4. Events of refinement/sign bits crossover.

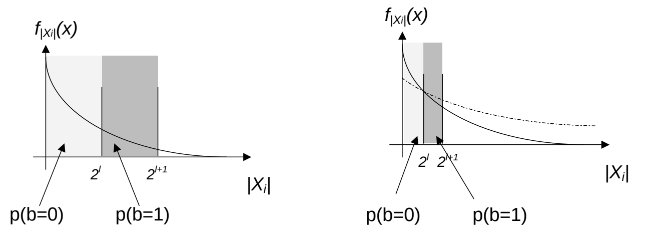
events (corresponding to regions A_j in Figure 4(b)) are also rare, except at the least significance levels (small l), when the crossover regions reside near to the origin (A_j starts at $|X_i| = 2^l$).

[1]Under some assumptions, the theoretical compression rate of DSC for a binary source is given by $H(p)$, where $p \leq 0.5$ is the crossover probability between the source and SI [16]. Therefore, the compression efficiency of DSC increases as the crossover probability becomes smaller.

Therefore, in our system, we compress sign and refinement bits with DSC to exploit inter-band correlation. This eliminates configuration (d) for magnitude bits compression, and only (a) and (e) are left for further evaluation.

3.2. Compression of significance maps

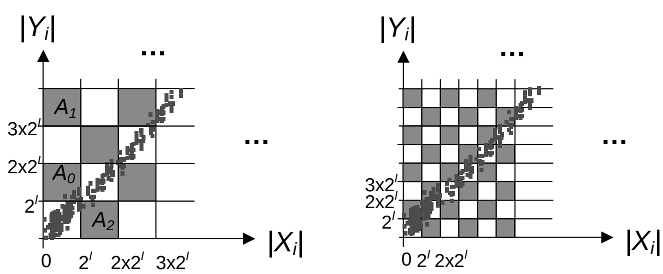
Significance maps are biased toward zero in general (i.e., insignificant coefficients are more numerous). This can be verified from the distribution of the coefficients (Figure 5(a)). Therefore, intra coding can achieve compression for significance maps. However, this bias will tend to decrease at the least significance levels (when l is small) as shown in Figure 5(b). Accordingly, intra coding of significance maps tends to become less efficient when coding the least significant bit-planes. In addition, the bias tends to decrease as the variance of the coefficients increases (see Figure 5(b); a more rigorous mathematical justification will be given later). Therefore, for low-pass subbands and high decomposition level subbands, intra coding may not be very efficient.



(a) Most signif. levels. (b) Least signif. levels/large var.

Fig. 5. Bias in significance maps.

Significance maps bits can also be compressed using DSC. In this case, since both the refinement bits and significance maps bits are encoded by DSC, we compress them together (i.e., as a raw bit-plane) in one pass. For a raw bit-plane, the bit crossover events correspond to the regions A_j in Figure 6(a), leading to relatively small crossover probability and thus making DSC a good choice to encode this data. On the other hand, when the significance level l is small, the area of each crossover region decreases (A_j are square regions with length 2^l), and they become more evenly distributed over the sample space (Figure 6(b)). As a result, more samples fall within the crossover regions, the crossover probability increases and DSC becomes less efficient.



(a) Most significance levels. (b) Least significance levels.

Fig. 6. Raw bit-planes crossover.

To summarize, both intra coding or DSC could achieve considerable compression of significance maps, and configurations (a) and

(e) are promising candidates. However, it is unclear which one we should use in different situations to achieve the optimal coding performance. Therefore, we propose modeling techniques to precisely analyze their performance for different source characteristics and correlation levels.

3.3. Modeling

The modeling techniques estimate the number of *coded* bits for configurations (a) and (e). Recall that in configuration (a) we encode the entire raw bit-plane using DSC, whereas in (e) we partition the bits into refinement bits and significance maps and encode them separately using DSC and intra coding respectively (Figure 2). To determine the optimal coding strategy, we need to estimate and compare the number of coded bits at each significance level l and for each wavelet subband i. The estimated number of coded bits for configuration (a) is given by:

$$N_i \times (H(p_{raw}(l,i)) + m), \tag{1}$$

where N_i is the number of uncoded raw bits at a given significance level in wavelet subband i (hence N_i equals to the number of coefficients in subband i), and $p_{raw}(l,i)$ is the estimated crossover probability of the raw bit-plane at significance level l for coefficients in wavelet subband i. $H(p_{raw}(l,i))$ is the theoretical compression rate using DSC, and we add a margin m to account for the performance of practical systems. The estimates for $p_{raw}(l,i)$ can be derived by integrating the joint p.d.f. $f_{X_iY_i}(x,y)$ over the crossover regions A_j as shown in Figure 6. Details can be found in [15], where we discuss how to estimate crossover probabilities for the purpose of determining the encoding rate.

The estimated number of coded bits for configuration (e) is given by:

$$N_{ref}(l,i) \times (H(p_{ref}(l,i)) + m) + N_{signif}(l,i) \times \gamma(l,i), \tag{2}$$

where $N_{ref}(l,i)$ and $N_{signif}(l,i)$ are the number of uncoded refinement bits and significance map bits at significance level l in subband i respectively, $p_{ref}(l,i)$ is the estimate for refinement bits crossover probability, and $\gamma(l,i)$ is the compression ratio achieved by intra coding. $H(p_{ref}(l,i))+m$ is the compression ratio achieved by practical DSC scheme, and we can estimate $p_{ref}(l,i)$ as in [15]. We model the significance map bits as zeroth order binary source and estimate $\gamma(l,i)$ by $H(p_0(l,i))$, where $p_0(l,i)$ is the probability that significance map bits being zero. Assume the wavelet coefficients in subband i is Laplacian distributed with parameter β_i, i.e, $f_{X_i}(x) = \frac{1}{2}\beta_i e^{-\beta_i|x|}$, following from the definitions of refinement bits and significance maps, $N_{ref}(l,i)$, $N_{signif}(l,i)$ and $p_0(l,i)$ can be estimated by

$$N_{ref}(l,i) = N_i \exp(-\beta_i 2^{l+1}) \tag{3}$$

$$N_{signif}(l,i) = N_i(1 - \exp(-\beta_i 2^{l+1})) \tag{4}$$

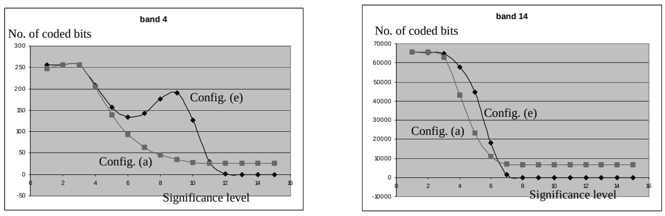
$$p_0(l,i) = \frac{1 - \exp(-\beta_i 2^l)}{1 - \exp(-\beta_i 2^{l+1})} \tag{5}$$

We found that $H(p_0(l,i))$ is a good estimate for the compression efficiency of zerotree coding.

3.4. Adaptive coding scheme

We compare (1) and (2) to determine the optimal coding configuration at each significance level and for each wavelet subband. Figure 7 plots coded bits estimated by (1) and (2) against significance levels for coding configurations (a) and (e) in two wavelet subbands. It can be seen that at the higher significance levels both schemes can achieve substantial compression, but better performance is achieved by compressing the significance map using intra coding. On the other hand, in the middle significance levels, coding the entire raw bit-plane with DSC can achieve better results. As for the least significant bit-planes, neither scheme performs well, as bits have entropy near one and the correlation with corresponding bits in the SI is minimal.

Based on these modeling results, we propose an *adaptive coding scheme*: when coding the most significant bit-planes at the beginning, we partition the magnitude bits into refinement bits and significance maps, and apply DSC and intra coding (zerotree coding) respectively (i.e., configuration (e)). Later, in the middle significance levels, we switch to compress the entire raw bit-planes using DSC (i.e., configuration (a)) (Figure 8). We use (1) and (2) to determine the significance level at which configurations switching should occur. Note that for different subbands, switching could occur at different significance levels. Switching would occur earlier (at a higher significance level) for high decomposition level subbands as intra coding is less efficient there[2]. Intuitively, in high decomposition level subbands, coefficients would become significant earlier, and zerotree coding would become inefficient if the number of significant coefficients to be signaled increases.



(a) The 4th subband. (b) The 14th subband.

Fig. 7. Modeling results.

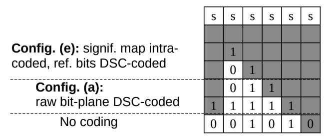


Fig. 8. Adaptive coding scheme.

4. EXPERIMENTAL RESULTS

We have applied our adaptive coding scheme to DSC-based hyperspectral image compression, using (1) and (2) to determine the significance level at which a switch between configurations occurs for each subband. We compared with the non-adaptive scheme in [3], which uses zerotree coding for significance maps for all significance levels (i.e. configuration (e) only). We used the NASA AVIRIS image data-sets in the experiment [12] . The original image consists

[2]Recall $p_0(l,i)$ is the probability of significance map bits being zero. We can rewrite (5) as $p_0(l,i) = \frac{1-\tau}{1-\tau^2}$, with $\tau = \exp(-\beta_i 2^l)$, hence $0 \le \tau \le 1$. It can be shown that $p_0(l,i)$ decreases monotonically with increasing τ, with $p_0(l,i) = 1$ when $\tau = 0$ and $\lim_{\tau\to 1} p_0(l,i) = 0.5$. Accordingly, intra coding would become less efficient for significance maps when τ is large, i.e., when l is small (at the least significance levels) and when β_i is small (when coefficients distributions have large variances, in low-pass subbands and high decomposition level subbands).

of 224 spectral bands, and each spectral band consists of 614×512 16-bits pixels. In the experiment, we compressed 512×512 pixels in each band. Figures 9 and 10 show some of the results in compressing images *Cuprite* (radiance data) and *Lunar* (reflectance data). Here MPSNR $= 10 \log_{10}(65535^2/\text{MSE})$, where MSE is the mean squared error between all the original and reconstructed bands. As shown in the figures, the adaptive coding scheme can provide considerable and consistent improvements in all cases, with up to 4dB gain at some bit-rates.

We have also compared the DSC-based systems with several 3D wavelet systems (3D ICER) developed in NASA-JPL [12]. As shown in Figure 9, the DSC-based system is comparable to a simple 3D wavelet system (FY04 3D ICER) in terms of coding efficiency. The simple 3D wavelet system uses the standard dyadic wavelet decomposition and a context-adaptive entropy coding scheme to compress coefficients bits. However, there is still performance gap when comparing the DSC-based systems to a more recent and sophisticated version of 3D wavelet (Latest 3D ICER). The more recent 3D wavelet developed in NASA-JPL exploited the spatial correlation remained in the correlation noise [12]. This could be one direction to improve the DSC-based systems, which currently use simple i.i.d. model for correlation noise and ignore the dependency between correlation noise symbols. We have also compared the DSC-based systems with 2D SPIHT, and the DSC-based systems can achieve 8dB gains at some bit-rates as shown in the figures.

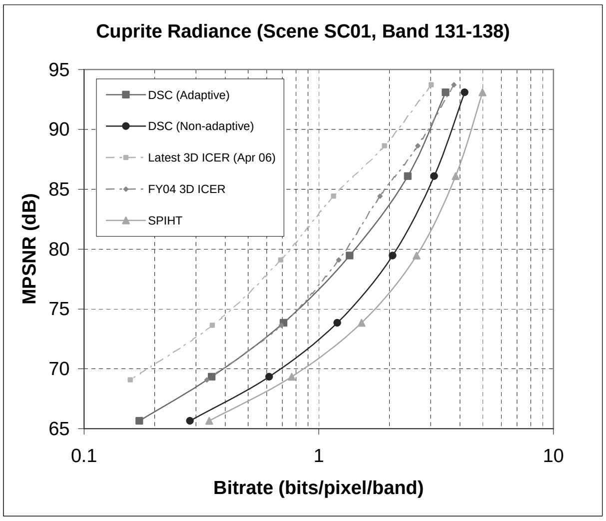


Fig. 9. Coding performance: Cuprite.

5. CONCLUSIONS

We have investigated the optimal combination of DSC/intra coding to encode wavelet coefficients bits in hyperspectral image compression. We have proposed modeling techniques to estimate the coding gains of different configurations, and the adaptive coding scheme to optimally combine different bits extraction techniques with DSC/intra coding. Experimental results have demonstrated that up to 4dB improvements can be achieved by the adaptive coding scheme, and the improved DSC-based system is comparable to some 3D wavelet system. For future work we plan to investigate better methods to model the correlation noise to improve coding efficiency.

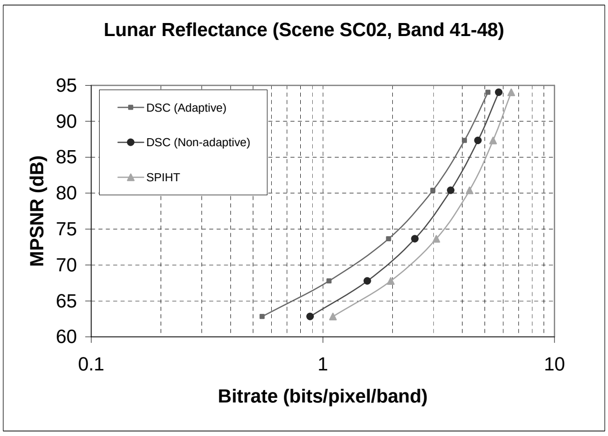


Fig. 10. Coding performance: Lunar.

6. REFERENCES

[1] A. C. Miguel, A. R. Askew, A. Chang, S. Hauck, and R. E. Ladner, "Reduced complexity wavelet-based predictive coding of hyperspectral images for FPGA implementation," in *Proc. Data Compression Conference (DCC)*, 2004.

[2] X. Tang, W. A. Pearlman, and J. W. Modestino, "Hyperspectral image compression using three-dimensional wavelet coding: A lossy-to-lossless solution," submitted to *IEEE Transactions on Geoscience and Remote Sensing*.

[3] C. Tang, N.-M. Cheung, A. Ortega, and C. S. Raghavendra, "Efficient inter-band prediction and wavelet based compression for hyperspectral imagery: A distributed source coding approach," in *Proc. Data Compression Conference (DCC)*, 2005.

[4] N.-M. Cheung, C. Tang, A. Ortega, and C. S. Raghavendra, "Efficient wavelet-based predictive Slepian-Wolf coding for hyperspectral imagery," *EURASIP Journal on Signal Processing - Special Issue on Distributed Source Coding*, Nov. 2006.

[5] D. Slepian and J.K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Information Theory*, vol. 19, pp. 471–480, July 1973.

[6] A.D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Information Theory*, vol. 22, pp. 1–10, Jan. 1976.

[7] S. Pradhan and K. Ramchandran, "Distributed souce coding using syndromes (DISCUS): Design and construction," in *Proc. Data Compression Conference (DCC)*, 1999.

[8] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proceedings of the IEEE, Special Issue on Advances in Video Coding and Delivery*, vol. 93, no. 1, pp. 71–83, Jan. 2005.

[9] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec using set partitioning in hierarchical trees," *IEEE Trans. Circuits and Systems for Video Technology*, pp. 243–250, June 1996.

[10] M. Barni, D. Papini, A. Abrardo, and E. Magli, "Distributed source coding of hyperspectral images," in *Proc. Int'l Symp. Geoscience and Remote Sensing (IGARSS)*, 2005.

[11] D. Taubman and M. Marcellin, *JPEG2000: Image compression fundamentals, standards and practice*, Kluwer, 2002.

[12] A. Kiely, M. Klimesh, H. Xie, and N. Aranki, "ICER-3D: A progressive wavelet-based compressor for hyperspectral images," Tech. Rep., NASA-JPL IPN Progress Report, 2006, http://ipnpr.jpl.nasa.gov/progress_report/42-164/164A.pdf.

[13] Caimu Tang and Cauligi S. Raghavendra, "Bitplane coding for correlations exploitation in wireless sensor networks," in *Proc. Int'l Conf. Communications (ICC)*, Seoul, May 2005.

[14] Xun Guo, Yan Lu, Feng Wu, Wen Gao, and Shipeng Li, "Wyner-Ziv video coding based on set partitioning in hierarchical tree," in *Proc. Int'l Conf. Image Processing (ICIP)*, 2006.

[15] N.-M. Cheung and A. Ortega, "A model-based approach to correlation estimation in wavelet-based distributed source coding with application to hyperspectral imagery," in *Proc. Int'l Conf. Image Processing (ICIP)*, 2006.

[16] A.D. Liveris, Z. Xiong, and C.N. Georghiades, "Compression of binary sources with side information at the decoder using LDPC codes," *IEEE Communications Letters*, Oct. 2002.