Predictive Lossless Compression of Regions of Interest in Hyperspectral Images With No-Data Regions

Hongda Shen, W. David Pan, *Senior Member, IEEE*, and Dongsheng Wu

***Abstract*—This paper addresses the problem of efficient predictive lossless compression on the regions of interest (ROIs) in the hyperspectral images with no-data regions. We propose a two-stage prediction scheme, where a context-similarity based weighted average prediction is followed by recursive least-square filtering to de-correlate the hyperspectral images for compression. We then propose to apply separate Golomb-Rice codes for coding the prediction residuals of the full context pixels and boundary pixels, respectively. To study the coding gains of this separate coding scheme, we introduce a mixture geometric model to represent the residuals associated with various combinations of the full context pixels and boundary pixels. Both information-theoretic analysis and simulations on synthetic data confirms the advantage of the separate coding scheme over the conventional coding method based on a single underlying geometric distribution. We apply the above prediction and coding methods to four publicly available hyperspectral image datasets, attaining significant improvements over several other state-of-the-art methods including the shape-adaptive JPEG 2000 method.**

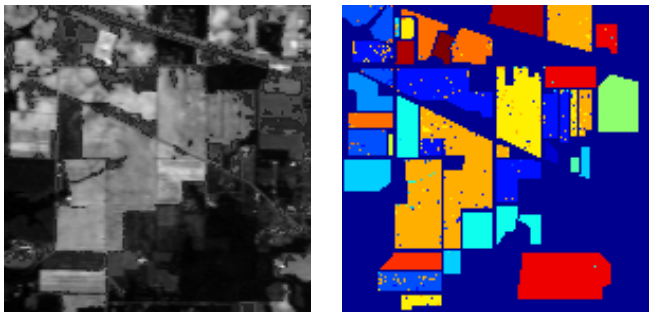
***Index Terms*—Region of interest, lossless compression, hyperspectral images, no-data regions, predictive coding.**

I. Introduction

Hyperspectral imaging technique has been widely used in many applications of remote sensing (RS) nowadays owing to its high spatial and spectral resolutions [1]. Hyperspectral data sets are normally of very large size which makes data acquisition, storage and transmission tasks very difficult and even problematic when network downlink bandwidth is limited. Also, as the number of hyperspectral imaging sensors grows, it is clear that data compression technique will play a crucial role in the development of hyperspectral imaging technique [2]. Lossy compression has been an effective technology because it significantly improves the compression efficiency at the cost of selective information loss. The fact that human visual and hearing system are not sensitive to certain types and levels of distortion caused by information loss enables the lossy compression. While lossy compression methods typically provide much larger data reduction than lossless methods, they might not be suitable for hyperspectral images used in many accuracy demanding applications, where the images are intended to be analyzed automatically by computers. For these applications, lossless compression methods can guarantee no loss in the reconstructed data. Thus lossless compression is more suitable choice for accuracy demanding RS applications than lossy compression [1].

H. Shen and W. D. Pan are with the Department of Electrical and Computer Engineering, University of Alabama in Huntsville, Huntsville, AL, 35899 USA. Emails: hs0017@uah.edu, pand@uah.edu.

D. Wu is with the Dept. of Mathematical Sciences, University of Alabama in Huntsville, Huntsville, AL 35899, USA. Email: dw0001@uah.edu.



(a) Spectral band 30. (b) ROI map.

Fig. 1: Sample dataset ("indian_pines_corrected") and ROI map.

Hyperspectral images generated by the hyperspectral imaging sensors have multiple components. For a specific application, it is very likely that only some regions of the entire image carry information of interest. We call these important parts of image, regions of interest (ROIs). Rather than compressing the entire image, sometimes we need to compress only ROIs in the image. Recently, a new concept known as "remote sensing image with *no-data* regions compression" has been proposed for the first time in [3]. The regions of an image where the data are considered of less interest are defined as *no-data* regions. Fig. 1 shows a hyperspectral image with numerous ROIs identified (Fig. 1a) and the map of these different ROIs (Fig. 1b). For each ROI, it is marked with a different color in Fig. 1b. If we focus our attention of a single ROI, then all the other ROIs can be considered as no-data regions. Three main situations where no-data regions arise in remote sensing applications have also been listed in [3]: 1) geometric and radiometric correction; 2) atmospheric events cover; and 3) ROIs are determined by the user/application. In remote sensing, it is very common that only certain image regions are important and informative rather than the entire images, especially for those object tracking or surveillance applications. Since the no-data regions do not provide useful information to some applications, compression methods will benefit significantly from simply not compressing those no-data regions for the sake of a higher compression ratio.

Many methods have been proposed to locate the ROIs

in the hyperspectral images. ROI classifiers are trained by applying supervised learning methods, such as support vector machine (SVM) [4] and deep belief network [5], to the labeled hyperspectral images first. Then we can predict if one pixel belongs to the predefined ROI or not. Moreover, human factors also play an important role in identifying the ROI. Experienced and trained specialists can manually select the ROI. However, different from the ROI of fixed size or regular shape, flexibility is usually required in the ROI identification of the real-world applications. Thus, ROIs of arbitrary shape and size are most likely to be expected for most of the RS applications. It is fairly easy to see those ROIs of arbitrary shape or size because of the existence of no-data regions in Fig. 1b: almost all of those ROIs have irregular shapes while some of ROIs even have dots and holes. In fact, the shape and size of ROIs vary a lot with different applications or environment.

Different from traditional ROI compression, all the non-ROI pixels are assumed not available in the ROIs with *no-data* regions compression situation. In other words, traditional solutions to ROI compression like JPEG 2000 with *Maxshift* are no longer applicable to this problem. Shape-adaptive (SA) [6][7] coding technique was designed to handle regions of arbitrary shape within the image, which makes itself a good tool to deal with no-data regions. A hyperspectral image can be viewed as a three-dimensional dataset with two spatial dimensions and one spectral dimension [1]. ROIs within the hyperspectral image have arbitrary shape and size only in the 2-D spatial configuration rather than the spectral dimension. Thus, SA 2-D methods can be extended easily to ROI with no-data region in the hyperspectral image by applying 2-D SA methods for each spectral band separately. Currently, most of the proposed SA ROI compression methods [3][8][9] for hyperspectral images can be generalized into a two-stage framework: 1) apply 2-D SA transform like SA-DWT to the ROIs of each spectral band, and then apply 1-D ordinary transform (DWT or PCA) along the spectral dimension for each pixel in the ROI; 2) bitplane encoding appproaches are applied to encode the transform coefficients obtained from the previous stage for only ROI pixels, while non-ROI pixels are ignored. As reported in the paper [3], three different 3-D SA coding have been designed: 3-D-Object Based-Set Partitioning in Hierachical Tress (OB-SPHIT) [10], 3-D-Object Based-Set Partitioned Embedded Block Coder (OB-SPECK)[8], and 3-D-Binary Set Splitting with K-D Tress (BISK) [9]. Also, as it is pointed out in the paper [3], TARP [11] and WDR [12] can be generalized to a 3-D SA version. JPEG 2000, a well-known compression standard based on wavelet transforms, provides the capability of encoding ROIs in an image by using the *Maxshift* technique [13]. A 3-D SA encoding scheme has been developed in [3] specifically for multispectral images containing no-data regions through the JPEG 2000 framework with delicate modifications. In [3], experimental results showed this scheme can yield better coding performance than SA-SPHIT and BISK in a lossy way.

These aforementioned SA ROI compression methods are all transform-based methods. While transform-based schemes can generally yield excellent lossy coding performance, their lossless coding performance is not as good as those specialized prediction-based methods [2]. Given the fact that lossless compression is required for those important ROIs, how to adapt the prediction-based compression methods to ROIs with no-data regions becomes a more crucial problem. To achieve high compression ratio without information loss, the prediction-based methods need be modified to be "shape-adaptive" to ROIs of arbitrary shape or size.

Prediction-based lossless compression approaches take advantage of the strong correlation of image signals. The intensity value of pixel of interest is estimated based on its context and the prediction error (residual) is encoded by an entropy coder. Since this prediction scheme decorrelates the image signal, the entropy of the residual is reduced, leading to smaller bit-rate. Given the fact that spectral bands in the hyperspectral image are highly correlated, linear model is usually utilized to decorrelate the colocated pixels in different spectral bands [14]. [15] presents an optimal linear predictor in the minimum mean-square sense followed by an entropy code on the residual signals. A linear predictor named as Spectral-oriented Least SQuares (SLSQ) is proposed in the paper [16] to exploit the spectral correlation of the hyperspectral image. Wiener filtering has also been adopted to hyperspectral image lossless compression [17]. [18] proposed to use two predictors: one interband linear predictor, and the other is interband least square predictor. With the usage of more memory, the compression ratio of this method is optimized. Recently, adaptive filtering techniques such as least mean square (LMS) [19] and Kalman filtering [14] have been used as predictors in a online learning fashion for lossless compression of the hyperspectral image, yielding good compression performance. It is worth mentioning that LMS method [19] developed by the NASA Jet Propulsion Lab (JPL) has been optimized and renamed as "Fast Lossless (FL)" algorithm and was selected as the core predictor in the CCSDS new standard for Multispectral and Hyperspectral Data Compression [20]. As well as linear prediction, nonlinear predictors have gained some success. [21] proposed a 3-D CALIC (M-CALIC) for lossless and near-lossless compression of the hyperspectral image, by extending classical 2-D CALIC to 3-D scenarios. In [2], a new non-linear predictor known as Context-based Conditional Average (CCAP) was proposed for the spatial decorrelation of hyperspectral image. In addition, some researchers intend to cluster pixels with similar statistics together and apply prediction within each cluster. For example, [22], [23] applied K-means to pre-process the hyperspectral image and achieve the best compression performance by far. Fuzzy K-means has also been used as a switch among a set of linear regression predictors in [24]. Although clustering techniques help improve the prediction accuracy, it relies on the availability of data information from all the spectral bands, while are normally not available or only partially available in many real-world or real-time applications. Thus, clustering techniques has its limit for the lossless compression of the hyperspectral image.

Fig. 2 gives an example of ROI with no-data region in one hyperspectral image. Non-ROI pixels in this hyperspectral image are all shown in white whereas these colored pixels are all from one ROI. While the ROI has an arbitrary shape

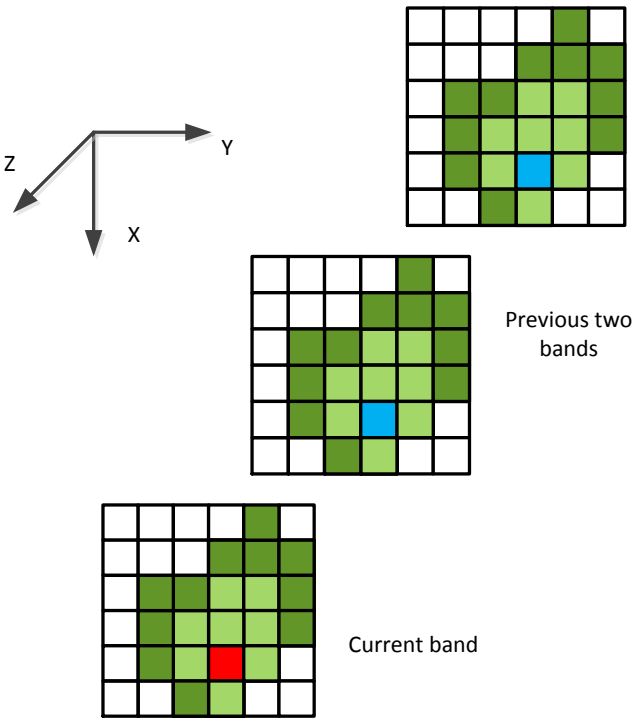


Fig. 2: A sample ROI with no-data region in an hyperspectral image. *Boundary pixels* are colored in dark green while *full-context pixels* are colored in light green. The red pixel is a randomly selected ROI pixel at the current band with two co-located pixels from previous bands colored in blue.

spatially (x and y dimensions), it has the same shape in each spectral band (the z dimension). Among those ROI pixels, some of them have non-ROI pixels in their causal context. We call these pixels *boundary* pixels (colored in dark green in Fig. 2). Note that, for any ROI pixel, its co-located pixels from previous bands are always considered available at predictor because they will be decoded prior to the current pixel at decoder side. In contrast, those ROI pixels whose causal context pixels belong to the ROI are called *full-context* pixels (colored in light green). The existence of ROIs in the hyperspectral images creates the boundaries separating the ROI and non-ROI pixels in the spatial dimension. This issue is often addressed by extending this boundary [3][6] in SA transform-based methods. Furthermore, non-ROI pixels can be skipped during the bitplane coding in those transform-based methods. In contrast to transform-based methods, boundary pixel issue is not trivial for prediction-based methods because of lack of full context support, which makes prediction accuracy inconsistent for boundary pixels. The prediction becomes less accurate for those cases in which a ROI pixel has only one ROI pixel in its context. The simplest solution to this problem is to pad certain preselected constants to those boundary pixels' contexts to make up for those missing context pixels [1]. But this substitution could cause a new problem: it may degrade the prediction performance and produce relatively larger residuals since the padded constants do not reflect the actual statistics of the local image signal. As a result, the overall compression ratio may suffer from this potential degradation. It is worth noting that some ROI identification methods may create some random dots-or-holes type ROI pixels out of false detection. Thus, those dots-or-holes like ROI pixel do not have any ROI pixel in their context. Naturally, it is impossible to apply any existing prediction scheme for this special ROIs of arbitrary shape and size in the hyperspectral image. Hence, adaptive modifications to those aforementioned prediction-based methods are required in order to apply to ROIs with no-data regions. In order to losslessly compress the ROIs with no-data regions in the hyperspectral image, we propose a new method based on a complete analysis of this special problem in this paper.

Our contributions are: 1) we first conducted an information-theoretic analysis to justify the proposed scheme of separating boundary pixels from full context pixels in prediction and coding; 2) we then proposed a two-stage predictor, including a 2-D spatial predictor based on context similarities, and a 1-D spectral predictor using recursive least square (RLS) filtering to achieve good lossless compression performance on ROIs with no-data regions in hyperspectral images. The prediction residuals of boundary and full-context pixels are encoded by two Golomb-Rice (GR) encoders separately. Golomb Codes were first introduced in [25]. In the Golomb Coder family, the Golomb-Rice code is widely used for lossless compression of non-negative integers thanks to its easy implementation and low-complexity. GR coding can be considered as a special case of the general Golomb coding with parameters being limited to powers of two. In contrast to the Huffman coding, GR coding has the advantage of not having to transmit the Huffman table. Hence, two GR coders with different parameters are applied to encode residuals of boundary pixels and full-context pixels independently in this implementation of MGD golomb coding designed for ROIs of arbitrary shape and size.

The paper is organized as follows. In Section II, we formalize the problems generated from lossless compression for ROIs of arbitrary shape and size and provide a thorough information-theorectic analysis of this special problem. In Section III, we discuss the proposed algorithm in details. In Section IV, we provide the results on lossless compression performance for ROIs of arbitrary shape and size in four hyperspectral image datasets, and compare the proposed algorithm with other state-or-the-art methods. Finally, the conclusions are drawn in the Section V.

II. Problem Analysis

In each hyperspectral image, some regions convey very crucial information to a specific RS application while other regions do not. We call the former Regions of Interest (ROIs) and the latter No-Data Regions. ROIs can be identified by some machine learning algorithms as discussed in Section I or manually selected by a human viewer. Once the ROIs have been identified, a binary ROI map which shows the locations of ROI pixels is also generated. Generally, prediction-based lossless compression utilizes the causal context pixels to

estimate the value for each one of the pixels in the 2-D image. The values of those context pixels will be available at the decoder during the decoding to allow for lossless reconstruction. However, for some of ROI pixels, part of its causal context belong to no-data regions and are not available to the predictor. As a result, prediction scheme yields much more inaccurate performance.

As discussed in Section I, boundary pixels inevitably exist in each band of one hyperspectral image, especially when ROIs are of irregular shape and large size. Fortunately, there is no such problem along the spectral dimension of the hyperspectral image, given the fact that membership to ROI or no-data regions only affect pixels in the 2-D spatial scenario. Naturally, it is reasonable to separate prediction scheme into two stages and apply different predictors accordingly.

A. Prediction Scheme

We evaluate the potential compression gain that can be achieved using an information-theoretic analysis similar to that in [14]. We investigate how the general information change with different prediction schemes for *boundary* and *full-context* pixels and estimate the potential improvement.

1) 2-D Spatial Prediction: We consider each band of one hyperspectral image as a 2-D image. In other words, the entire hyperspectral image can be treated as a pile of such 2-D images. So each band of the hyperspectral image will be analyzed independently from the other bands. Moreover, we only work on one band at a time so that only spatial correlation will be exploited. We treat a 2-D image as a random variable X so that the average amount of information contained in this image is given as the entropy $H(X)$. Thus, X takes on the values in the set $A = \{0, 1, ..., 2^{N_B} - 1\}$ whose cardinality equals 2^{N_B} where N_B is the bpp data format for this hyperspectral image. So the entropy $H(X)$ can be further computed as $H(X) = -\sum_{i \in A} p_i \log_2 p_i$ where $p_i = P(X = i),\ i = 0, 1, ..., 2^{N_B} - 1$.

As we have already discussed, causal context pixels shown in the Fig. 2 are commonly used as inputs to the predictor to make the estimation. Therefore, the prediction scheme can be mathematically modelled by the conditional entropy because conditional entropy quantifies the uncertainty associated with random variable when conditional random variables are known. We denote the context of each pixel as C_j and the conditional probability is actually a contextual probability, $p_{i|C_j}$. The conditional entropy can be redefined as $H(X|C_j) = -\sum_{i \in A} p_{i|C_j} \ \log_2 p_{i|C_j}$. Then the conditional entropy of the entire image can be computed as: $H(X|C) = \sum_j p(C_j) H(X|C_j)$, where $p(C_j)$ represents the probability of j^{th} causal context. *Boundary* pixels, as we defined earlier, only have part of their full causal contexts so the context C_b of each boundary pixel is just a subset of the full context C, i.e., $C_b \subset C$. It follows $H(X|C_b) \geq H(X|C)$ by the convexity of H, which indicates that the minimum achievable bit rate for boundary pixels is usually larger than that of full-context pixels. So we propose to apply different 2-D spatial predictors to boundary pixels and full-context pixels, respectively, in order to minimize $H(X|C_b)$ and $H(X|C)$ at the same time to eventually reduce the conditional entropy of each spectral band.

2) 1-D Spectral Prediction: Spectral prediction is relatively easy because for each ROI pixel in each band, the previous multiple bands of co-located pixel can be used as its context for prediction. In other words, each ROI pixel has a full support from its context in spectral dimension so that the boundary pixel problem in the 2-D spatial domain does not exist in the spectral dimension. Moreover, [14] proved mathematically the effectiveness of decorrelating the dependency between the current band and previous band using a similar conditional entropy analysis model. Here only ROI pixels will be considered instead of the entire hyperspectral image given the fact that no-data regions do not carry any useful information and their intensity values are not accessible either. Thus, a linear multiband prediction scheme can be easily applied to those ROI pixels with minimal modification regardless of boundary or full-context pixels. For example, in Fig. 2, an arbitrary ROI pixel (colored in red) can be estimated by a linear combination of its colocated pixels from previous bands (colored in blue).

B. Entropy Coding

It has been widely accepted that the global statistics of residuals from a predictor in one image can be well-modelled by a Geometric Distribution (GD) starting from zero [26]. Therefore, in this section, we analyze the residuals of ROIs and propose a new model based on GD to analyze the statistics of residuals of ROIs in the hyperspectral image.

After the aforementioned two-stage prediction, two different prediction methods are applied to boundary and full-context pixels separately. Residuals produced for boundary and full-context pixels tend to have different statistics as pointed out: the conditional entropy for boundary pixels are usually larger than full-context pixels. Therefore, we propose that GD with different parameters, named as Mixture Geometric Model (MGD), depicts the statistical characteristics of the residual data for boundary and full-context pixels more accurately. Before we present the details of this MGD model, we define two ratios for each band:

ROI Ratio:

$$R_1 = \frac{\#\ of\ ROI\ pixels}{Total\ \#\ of\ pixels}, \tag{1}$$

Boundary Ratio:

$$R_2 = \frac{\#\ of\ Boundary\ pixels}{\#\ of\ ROI\ pixels}, \tag{2}$$

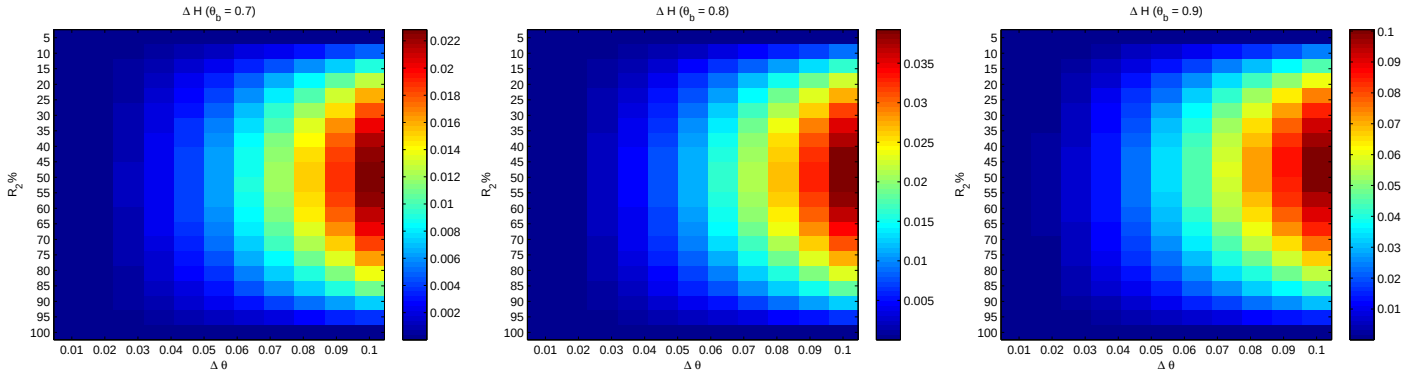
According to our assumption, the distribution of residuals of boundary pixels is modelled by one GD:

$$P_b(i) = (1 - \theta_b){\theta_b}^i, \quad i \in \mathbf{Z}^+, \quad \theta_b \in (0, 1). \tag{3}$$

While the distribution of residuals of full-context pixels is similarly modelled by another GD:

$$P_f(i) = (1 - \theta_f){\theta_f}^i, \quad i \in \mathbf{Z}^+, \quad \theta_f \in (0, 1). \tag{4}$$

To our best knowledge, all the existing ROI lossless compression methods apply only a single coder to the entire ROI. Thus, a single estimated GD will be computed to model the



(a) $\theta_b = 0.7$ (b) $\theta_b = 0.8$ (c) $\theta_b = 0.9$

Fig. 3: Simulation results.

residuals of entire ROI. We call this single geometric model (SGD) in order to differentiate it from our proposed MGD. Let's denote this estimated GD as:

$$P_e(i) = (1-\theta_e)\theta_e{}^i, \quad i \in \mathbf{Z}^+, \quad \theta_e \in (0,1). \tag{5}$$

where parameter θ_e can be estimated from the sample residuals of all the ROI pixels including boundary pixels and full-context pixels. Hence, the total bit rate saving by encoding all the pixels using MGD based separate coding can be computed as the weighted KL distances where the weights are determined by the ratio R_2:

$$\Delta H = R_2 \cdot KL(P_b||P_e) + (1-R_2) \cdot KL(P_f||P_e) \tag{6}$$

Our goal is to evaluate the potential gain that can be achieved if MGD-based coding is employed. To make a fair comparison, we use this ΔH as a cost function. So the optimal estimated P_e can be obtained by minimizing this cost function. After solving $d\Delta H/d\theta_e = 0$, we have

$$\frac{1}{1-\theta_e} = \frac{1}{\theta_e}\left[\frac{R2 \cdot \theta_b}{1-\theta_b} + \frac{(1-R2)\theta_f}{1-\theta_f}\right]. \tag{7}$$

Let $K = \frac{R_2 \cdot \theta_b}{1-\theta_b} + \frac{(1-R_2)\theta_f}{1-\theta_f}$, then the optimal θ_e is computed as:

$$\theta_e^{opt} = \frac{K}{K+1}. \tag{8}$$

Furthermore, we can rewrite the equations above in terms of sample mean μ_b, μ_f and μ_e for each geometric distribution, P_b, P_f and P_e given the fact that $\mu = \theta/(1-\theta)$ for the geometric distributions. It can be further shown from Eq. (8), we can further derive that $\mu_e^{opt} = R_2 \times \mu_b + (1-R_2)\mu_f$, which is exactly the sample mean of this MGD. So this μ_e^{opt} can be replaced with μ_b and μ_f to express ΔH as a lower bound of performance gain using MGD, since this optimal estimated P_e is already the solution to the minimizer of ΔH. One of those important properties of the KL distance, $KL(P||Q) \geq 0$, known as Gibbs' inequality, guarantees that ΔH, the total saved bit rate, is always non-negative. In other words, we can always benefit from using MGD under our assumption.

To show the effectiveness of this MGD model numerically, a simple simulation has been conducted: a total number of 5×10^5 data points are first generated using a mixture of two geometric distributions $R_2 \times Geom(\theta_b) + (1-R_2) \times Geom(\theta_f)$ where R_2, the boundary pixel ratio, determines which GD source is the dominant one. We define $\Delta\theta = \theta_b - \theta_f$ as a measurement of the deviation of θ_f from θ_b and we fix θ_b at $\{0.7, 0.8, 0.9\}$. Owing to a large pixel intensity range of the hyperspectral images, it is very likely the prediction residuals are still considerably large. And prediction of boundary pixels in the ROI suffers insufficient context, which produces typically larger residuals compared to the full-context pixels as shown in our prior analysis. So it is reasonable to assume θ_b is relatively larger than θ_f in this experiment for a better approximation of the real case. Therefore, $\Delta\theta$ is set to vary from 0 to 0.1 at step size of 0.01 in our simulation. For each pair of R_2 and $\Delta\theta$, the corresponding ΔH values were computed using Eq. (6) and reported in the Fig. 3. It can be seen that ΔH increases with $\Delta\theta$ generally. This means that the advantage of the MGD model becomes most pronounced when θ_b and θ_f deviates the most from each other. Also, we can see that the ΔH tends to maximize as R_2 approaches 50%. As mentioned earlier, R_2 determines which GD dominates this simulated data from two mixed sources. In the case of $R_2 = 50\%$, two GDs play an equal role in this mixed data, therefore the separate coding based on the MGD model is expected to provide the largest possible coding gain than the SGD model. In addition, the range of each colorbar in Fig. 3 shows the theoretical performance gain of the MGD model over SGD model corresponding to different θ_b. Obviously, this gain increases as θ_b goes higher. As we have explained, high θ_b value can actually better approximate the real case.

The analysis and simulation show that our separate coding based on the MGD model can reduce the entropy of the residuals of ROI pixels to achieve better compression performance compared to the SGD model. We have also conducted an experiment on real hyperspectral image data with results provided in Section IV.

III. The Proposed Algorithm

Fig. 4 shows the whole procedure of our proposed method, with a two-stage prediction followed by a MGD Golomb-Rice encoder.

A. Two-stage Prediction

1) Context-Similarity based Conditional Average Prediction: At the first stage, a Context-Similarity based Weighted

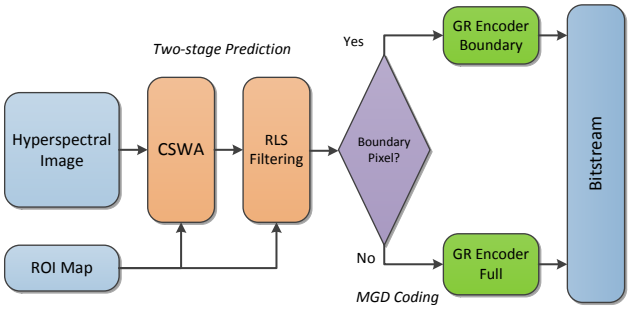


Fig. 4: Block diagram of the algorithm.

Average Prediction (CSWA) approach is presented to remove the redundancy for ROIs of the 2-D spatial image of each band. More specifically, this approach is designed to reduce the 1st-order entropy of the residuals. Within the each band, boundary pixels and full-context pixels are processed differently due to their inherent different statistical characteristics.

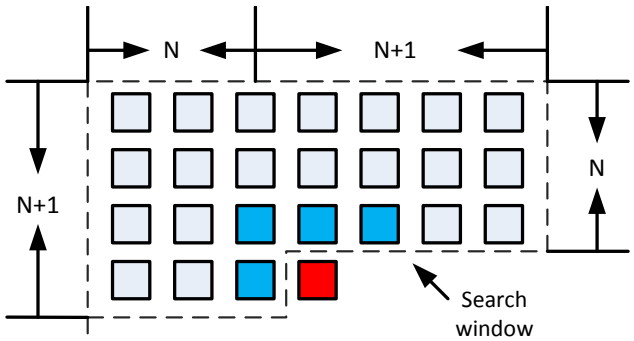


Fig. 5: 2-D CSWA.

Context-based Conditional Average (CCAP) was proposed in [2] to work as 2-D spatial predictor for hyperspectral image. Also, Median Edge Detector (MED) proposed in JPEG-LS [26] was initially designed as a 2-D natural image predictor and has been proved effective as 2-D spatial predictor for hyperspectral image in [19]. However, CCAP method neglects strong nonlinearity in the common structured regions in the hyperspectral image. As a remedy to this nonlinearity problem, we propose a 2-D spatial predictor based on context-similarity between the current context and the contexts of ROI pixels in the support of current pixel. In contrast to CCAP, instead of simple average, we take into account the context distance/similarity needs to be taken into account to make more accurate prediction. This was motivated by an image denoising method, non-local mean (NLM) filtering [27], which is essentially a weighted average filter with high weights assigned to highly similar sample context and low weights assigned to less similar sample context. Specifically, we modify the CCAP method by assigning large weights to those ROI pixels whose causal contexts are similar to the current context and vice versa. In our CSWA method, multiple distances metrics can be used to measure the context similarity, sum of absolute difference (SAD), sum of L_2 norm or context correlation. In our work, SAD is selected for its computational simplicity.

Fig. 5 illustrates this CSWA method by showing a small patch of pixels of one arbitrary band. Given a full-context ROI pixel of one band in the hyperspectral image $x_{i,j}$ (colored in red in Fig. 5, its causal context is denoted as $C_{i,j}$ (colored in blue in Fig. 5) which represents all the ROI pixels in its context. Besides, the search window of current ROI pixel is denoted as $S(i,j)$ (dashlined region in Fig. 5 and its size equals $N(N+1)$), which only includes all the available full-context pixels in the causal neighborhood for local estimation whereas boundary pixels are excluded. Therefore, our CSWA method can be described by the following equation:

$$y_{i,j} = \frac{\sum_{(m,n)\in S(i,j)} w_{m,n} I(m,n) x_{m,n}}{\sum_{(m,n)\in S(i,j)} w_{m,n} I(m,n)} \tag{9}$$

where $y_{i,j}$ is the estimate of $x_{i,j}$ at the location (i,j) using its support $S(i,j)$ and the weight of each ROI pixel in the $S(i,j)$ can be computed by:

$$w_{m,n} = \frac{1}{D(m,n,i,j)} \tag{10}$$

where $D(m,n,i,j)$ is the distance measure between the $C_{m,n}$ and $C_{i,j}$. And the selection factor $I(m,n)$ is defined as:

$$I(m,n) = \begin{cases} 1 & \text{if } w_{m,n} > T, \\ 0 & \text{if } else. \end{cases} \tag{11}$$

where T is a empirically chosen threshold to filter out those small weights which normally mean high dissimilarity. It is worth noting that using this selection factor will eliminate certain amount of pixels from contributing to the prediction. So when the number of pixels after the thresholding selection is not sufficient to make a good estimate, a simple arithmetic average will be used to replace this CSWA. Given the fact that boundary pixels do not have a complete context, it is not practical to find pixels with similar context. Hence, similarly, simple arithmetic average of available ROI pixels in the context will replace this CSWA for those boundary pixels.

CSWA predicts the ROI pixel values according to the similarities between its context and the contexts of pixels in the support. Those pixels with highly similar contexts will contribute more to the prediction than those with less similarity. This approach works very well in those common structured regions because of the fact that pixels with similar contexts almost always exist in those structured regions of image.

2) Recursive Least Square Filtering: Linear adaptive filtering techniques, such as Least Mean Square (LMS) [19], Least Square (LS) [16], [17] and Kalman Filtering (KF) [14], have been proposed to decorrelate the hyperspectral image for further compression. With some modifications, these adaptive filtering approaches can be used for hyperspectral image ROIs as well [1]. But LMS suffers greatly from its slow convergence speed, which makes the prediction much less accurate especially when the data size is not large enough to guarantee convergence, which is often the case for ROIs within an image. LS and KF perform better than LMS regarding the convergence speed. But they both require matrix inversion operations which normally cost too much computation. Given

the fact that most RS applications are time-sensitive, LS and KF may not be the best choice for lossless compression of the hyperspectral image. Recursive Least Square (RLS) [28] filtering is known for its fast convergence speed and its relatively higher but tolerable computational complexity. Thus, we choose RLS over other linear optimization methods for spectral prediction.

For each data point in the training sequence, RLS filtering recursively finds the optimal weights that minimize the weighted linear least squares cost function defined as $|d - \mathbf{u}^T\mathbf{w}|^2$, where d, $\mathbf{u}$ and $\mathbf{w}$ are the desired data, filtering input vector and weight vector respectively. So the input vector $\mathbf{u}_{M\times 1}$ of each pixel is defined, in this application, as the previous M bands at the same spatial location and d is the pixel value at the current band. A brief description of this adapted RLS applied in our ROI spectral prediction is given as follows:

1) Initialize $\mathbf{w}(0) = 0$ and the auxiliary matrix $\mathbf{P}(0) = \epsilon\mathbf{I}$ where $\mathbf{I}$ is the identity matrix of size $M \times M$ and ϵ is a small constant; The ROI pixel index $i = 1$.

2) For ROI pixel i compute

$$r(i) = 1 + \mathbf{u}(i)^T\mathbf{P}(i-1)\mathbf{u}(i), \tag{12}$$

$$\mathbf{k}(i) = \mathbf{P}(i-1)\mathbf{u}(i)/r(i), \tag{13}$$

$$e(i) = d(i) - \mathbf{u}(i)^T\mathbf{w}(i-1), \tag{14}$$

$$\mathbf{w}(i) = \mathbf{w}(i-1) + \mathbf{k}(i)e(i), \tag{15}$$

$$\mathbf{P}(i) = \mathbf{P}(i-1) - \mathbf{k}(i)\mathbf{k}^T r(i). \tag{16}$$

3) Update ROI pixel index to the next ROI pixel $i = i + 1$ and repeat 2) till i reaches N_{ROI}, where N_{ROI} represent the total number of ROI pixels for each band.

4) Reset $\mathbf{w}(0)$, $\mathbf{P}(0)$ and i for each band.

Apparently, this RLS filtering for spectral prediction is one online learning optimization method, which is very suitbale for this on-the-fly application. Instead of using only instantaneous values as in LMS, RLS filtering uses M most recent input vectors and desired data. With better appoximation, it is reasonable for RLS to converge faster than LMS. This feature facilitates RLS to be a better tool for spectral prediction when the number of ROI pixels are insufficent. Note that RLS distributes the computation load into each iteration as pointed out in [28]. Thus, though RLS converges faster than LMS, its complexity is typically higher than LMS and actual run-time cost increases with the number of pixels involved. However, for the ROI lossless compression, the number of pixels in each ROI is generally much less than the number of pixels in the entire image. So RLS seems to be a good fit for ROI spectral prediction.

B. MGD Golomb Coding Implementation

In Section II, we have already analyzed the statistical characteristics of ROI pixels and proposed a MGD model to guide the coding: apply two different GR coders on prediction residuals of boundary pixels and full-context pixels respectively. It is well-known that GR Coder are usually applied to encode data whose distribution is GD. Given a nonnegative integer y, mth order GR scheme encodes it in two steps: a *unary* representation of $\lfloor y/m \rfloor$, and a binary representation of y mod m using m bits.

After two-stage prediction, residuals of the all the pixels including boundary and full-context pixels are obtained and mapped to non-negative numbers for GR coding as follows:

$$f(n) = \begin{cases} 2n & \text{if } n \geq 0, \\ -2n-1 & \text{if } n < 0. \end{cases} \tag{17}$$

where n refers to the residual of ROI pixels. Note that it is necessary to round all the residuals to the nearest integer before this mapping and the decoder will duplicate this operation to ensure lossless decoding and reconstruction.

Since the actual bitsream length is controlled directly by the GR coder parameter, we implement these two GR coders by following the entropy coding configuration specified in CCSDS 123 standard [20] for our MGD model. With the assistance of this MGD model and two GR coders, all the residuals of ROI pixels of each band will be encoded into two separate bitstreams for further processing.

IV. Test Results

A total of four hyperspectral image datasets were tested: *Indian_Pines*, *Kennedy_Space_Center (KSC)*, *Salinas* and *Pavia_University* as listed in Table I. Detailed information of these datasets can be found in [29]. Based on their classification ground truth maps provided in [29], Support Vector Machine (SVM) based classification method proposed in [4] has been implemented to produce their ROI maps with high accuracy. Due to the limited space, we show the ROI map of only one dataset, *Indian_Pines*, in Fig. 6.

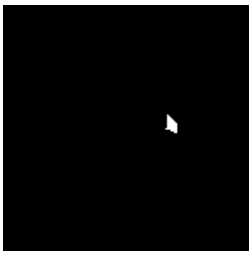
TABLE I: DATASETS USED

Dataset	Size	# of ROIs	Sensor
Indian_Pines	$145 \times 145 \times 200$	17	AVIRIS
Pavia_University	$610 \times 340 \times 103$	10	ROSIS
Salinas	$512 \times 217 \times 204$	17	AVIRIS
Kennedy_Space_Center	$462 \times 464 \times 176$	14	AVIRIS

As we introduced in Section I, much effort has been made to compress the hyperspectral image losslessly. However,to our best knowledge, there is not much work focusing on ROIs in the hyperspectral image so far. Most of the existing hyperspectral image lossless compression approaches require various types of modifications at different levels to adapt to ROI compression. So we only select to modify two most representative methods: 1) FL method from CCSDS standard 2) shape-adaptive JPEG 2000 method. Combining the approach [1] particularly designed for this problem for the first time, we use these three state-of-the-art methods as a benchmark set to evaluate our algorithm:

1) The ROI-LMS method was proposed to target this lossless compression of ROIs in the hyperspectral image [1]. A simple arithmetic average filtering is applied to the *boundary* pixels while conventional LMS is applied on *full-context* pixels followed by a separate GR coding scheme. However, LMS suffers from slow convergence on ROIs of small size. In addition, even though separate

(a) ROI 0 (Background).



(b) ROI 1.

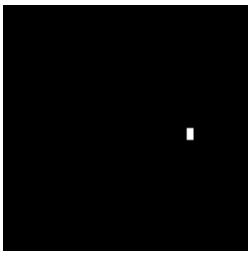
(c) ROI 2.

(d) ROI 3.

(e) ROI 4.

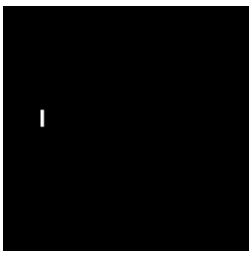
(f) ROI 5.

(g) ROI 6.

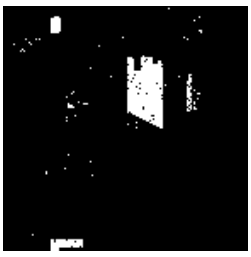


(h) ROI 7.

(i) ROI 8.



(j) ROI 9.

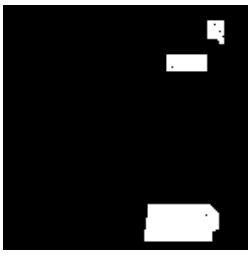


(k) ROI 10.

(l) ROI 11.

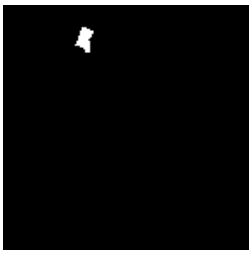
(m) ROI 12.

(n) ROI 13.



(o) ROI 14.

(p) ROI 15.



(q) ROI 16.

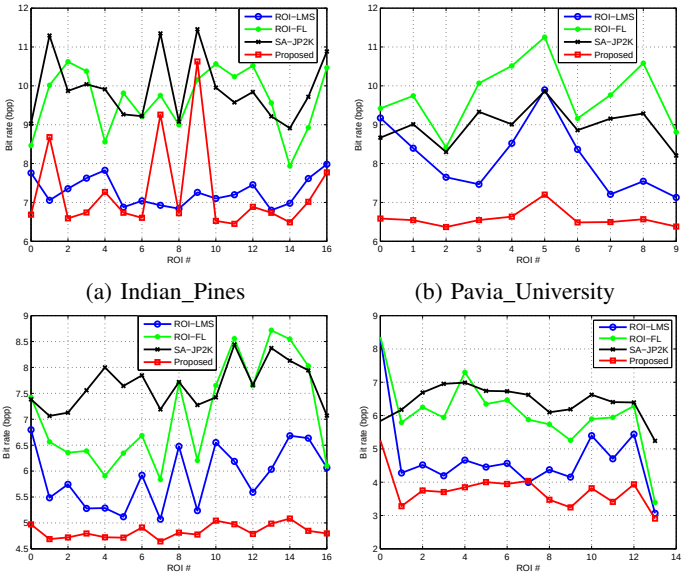
Fig. 6: Individual ROI maps (ROI pixels are shown in white while non-ROI pixels in black). Note: ROI 0 (Fig. 6a) can be viewed as the background, which contains all the pixels outside the other 16 ROI's.

coding was utilized in that work, a thorough mathematical analysis was missing. The new method we present in this paper aims to address these issues.

2) In CCSDS standard, a variable learning rate for LMS was proposed for the entire hyperspectral image [19]. However, it is almost impossible to directly apply this FL method to ROIs becuase of the typically irregular shape given the fact its variable learning rate is designed and optimized based on the entire hyperspectral image rather than ROIs. Hence, we modified this method by applying an arithmetic averaging operation on boundary pixels, while LMS with fixed learning rate will be applied to full-context pixels only. It is worth noting that separate (boundary/full-context pixels) coding scheme is not applied in this method. This method is denoted as ROI-FL.

TABLE II: COMPARISON OF BIT RATES (BITS/PIXEL) WITH OTHER SCHEMES ON "INDIAN_PINES"

ROI#	NEW	ROI-LMS	ROI-FL	SA-JP2K	R_1	R_2
0	**6.68**	7.76	8.46	9.03	51.25	21.27
1	8.68	**7.05**	10.02	11.29	0.22	56.52
2	**6.59**	7.35	10.62	9.87	6.84	36.09
3	**6.74**	7.62	10.38	10.04	3.76	38.81
4	**7.26**	7.83	8.56	9.91	1.13	23.21
5	**6.73**	6.87	9.81	9.26	2.30	26.65
6	**6.60**	7.04	9.21	9.22	3.55	27.75
7	9.25	**6.93**	9.75	11.35	0.13	57.14
8	**6.72**	6.84	9.00	9.08	2.28	15.42
9	10.62	**7.26**	10.16	11.46	0.10	100
10	**6.52**	7.10	10.56	9.96	4.59	36.75
11	**6.44**	7.20	10.24	9.57	11.73	28.83
12	**6.88**	7.46	10.52	9.84	2.87	32.28
13	**6.73**	6.80	9.56	9.21	0.98	20.98
14	**6.48**	6.98	7.94	8.91	5.99	14.14
15	**7.01**	7.61	8.93	9.72	1.84	21.50
16	**7.77**	7.98	10.47	10.88	0.44	39.78
Total	**6.67**	7.49	9.12	9.30	100	24.96



(a) Indian_Pines (b) Pavia_University

(c) Salinas (d) Kennedy_Space_Center

Fig. 7: Compression bit rates on four datasets.

3) SA JPEG 2000 (SA-JP2K) was proposed in [3] to handle ROIs of hyperspectral image and achieved excellent lossy compression performance over other transform-based methods. We removed the truncation and quantization process so that it supports lossless compression.

Table II shows the bit rates achieved by four different lossless compression methods on each ROI of the test dataset, *Indian_Pines*, shown in Fig. 6. As shown, it is obvious that our proposed new method outperforms the ROI-FL and SA-JP2K [3] in all 17 ROIs. Furthermore, this proposed method yields lower bit rate than ROI-LMS method except for some small ROIs (ROI 1, 7 and 9 in Fig. 7a). ROIs of extreme small size (low R_1 ratio) can not provide sufficient pixels for adaptive online learning algorithm to converge though RLS is known for its fast convergence speed. Thus, the multi-band prediction using RLS produces less accurate estimate, which increases the bit rate of the compressed bitstream.

Boundary ratio (R_2) indicates the shape information of ROI: more boundary pixels in one ROI, more irregular shape this ROI has in general. Table II shows that our method performs the best when R_1 and R_2 are both relatively high. Fig. 7 shows the bit rates comparisons for all four datasets. Apparently, our proposed method (red curves) generally performs better than the other three methods. Note that all methods do not do well on ROI 0 of dataset *KSC*. It turned out that there is a large amount of impulsive noises in the last 100 bands of this dataset. Given the fact that linear prediction technique cannot handle impulsive noise well and ROI 0 takes about over 97% pixels of this dataset, we believe adaptive learning based method needs some restraints on sudden change of weights in the learning algorithm for this special case. But our method still outperforms all other three methods on this impulsive noise corrupted dataset.

Similar to our simulation in Section II, we compared the MGD model and estimated SGD model on the residuals of real hyperspectral image produced by our proposed method. Their bit rates are shown in the third and fourth column of the Table III respectively. Empirical shannon entropy of residuals of the MGD model, H, estimates the information lower bound of bit rate conveyed in each ROI of the test dataset. As we can see, the MGD model has significantly lower bit rate than the estimated single GD model for each ROI in the dataset. Note that those ROIs of extreme small size do not provide sufficient pixels to produce statistically meaningful estimates of probabilities, thereby making H less useful. All these results verify our information-theoretic analysis of MGD model. Particularly, for those ROIs whose R_2 values are close to 50%, the corresponding performance gain of MGD against SGD are significantly higher than other ROIs. In the case where the true underlying distributions of either the boundary pixels or the full-context pixels are not exactly following the geometric distributions, the MGD model can provide a better approximate to the underlying distribution, thereby leading to higher compression ratios than the SGD model. Furthermore, implementation of separate coding of boundary pixels and full-context pixels are expected to produce shorter bitstreams.

TABLE III: BIT RATE GAIN ACHIEVED BY MGD.

ROI#	H	MGD	SGD	R_1	R_2
0	5.0543	6.6862	6.6889	51.25	21.27
1	4.5285	8.6804	9.2471	0.22	56.52
2	4.9763	6.5908	6.5964	6.84	36.09
3	5.0034	6.7403	6.7571	3.76	38.81
4	5.1309	7.1684	7.4701	1.13	23.21
5	4.8514	6.7348	6.7413	2.30	26.65
6	4.8986	6.6037	6.6048	3.55	27.75
7	3.8963	9.2579	9.9911	0.13	57.14
8	4.7801	6.7241	6.7878	2.28	15.42
9	3.9219	10.6273	10.6273	0.10	100
10	5.0189	6.5256	6.6345	4.59	36.75
11	5.0128	6.4491	6.4552	11.73	28.83
12	5.0296	6.8869	6.8869	2.87	32.28
13	4.7900	6.8338	7.1186	0.98	20.98
14	4.7307	6.4883	6.4949	5.99	14.14
15	4.9384	7.0157	7.0484	1.84	21.50
16	4.9362	8.0427	8.4159	0.44	39.78

To further validate our algorithm lossless compression per-

TABLE IV: BIT RATES IN BITS PER PIXEL FOR LOSSLESS COMPRESSION OF FOUR TEST DATASETS

Dataset	FL	The proposed Algorithm
Indian_Pines	6.60	6.54
Pavia_University	4.89	4.85
Salinas	6.54	6.52
Kennedy_Space_Center	4.97	5.21

formance, we have conducted another experiment, where we treated the entire image as on single ROI and applied the proposed method on this ROI. We used the C implementation of the FL method in [20], which was downloaded online [30]. We set all the parameters in FL method by following the work [31]. The results are summarized in Table IV. It can be seen that our algorithm slightly outperforms FL method on all test datasets except for $Kenneday\_Space\_Center$. This dataset contains substantial impulse noise, which might be better handled by an optimized implementation of the FL method with a restraint on sudden change of weights in LMS learning. Although the proposed method performs very closely to the standard approach on lossless compression of the entire hyperspectral image, it was designed for lossless compression of ROIs with no-data regions.

V. Conclusion

We have proposed a new lossless compression algorithm on arbitrarily shaped ROIs in hyperspectral images. The algorithm has several innovations: First, a two-stage prediction was introduced to effectively de-correlate the hyperspectral image spatially and spectrally. Then, based on the assumption that the prediction residuals would be better modeled by a mixed geometric distribution rather than a single geometric distribution, a separate Golomb-Rice coding method was employed to encode the prediction residuals. Simulation results on the four hyperspectral image datasets with predefined ROIs with no data regions showed that the proposed algorithm provided much higher compression ratios than several other methods, including the state-of-the-art shape-adaptive JPEG 2000 method.

References

[1] H. Shen, W. D. Pan, and Y. Wang, "A novel method for lossless compression of arbitrarily shaped regions of interest in hyperspectral imagery," in *Proc. 2015 IEEE SoutheastCon*, April 2015.

[2] H. Wang, S. D. Babacan, and K. Sayood, "Lossless hyperspectral-image compression using context-based conditional average," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 4187–4193, Dec. 2007.

[3] J. Gonzalez-Conejero, J. Bartrina-Rapesta, and J. Serra-Sagrista, "JPEG 2000 encoding of remote sensing multispectral images with no-data regions," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 2, pp. 251–255, April 2010.

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

[5] Y. Chen, X. Zhao, and X. Jia, "Spectral spatial classification of hyperspectral data based on deep belief network," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens*, , no. 99, pp. 1–12, 2015.

[6] 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.

[7] M. Cagnazzo, S. Parrilli, G. Poggi, and L. Verdoliva, "Cost and advantage of object-based image coding with shape-adaptive wavelet transform," *J. Image Video Process.*, , no. 1, pp. 19, Jan. 2007.

[8] X. Tangi, W. A. Pearlman, and J. W. Modestino, "Hyperspectral image compression using three-dimensional wavelet coding," *IEEE Trans. Geosci. Remote Sens.*, vol. 3653, pp. 294–350, 2002.

[9] J. T. Rucker and J. E. Fowler, "Coding of ocean-temperature volumes using binary set splitting with k-d trees," in *Proc. Int. Geosci. Remote Sens. Symp.*, Sept. 2004, pp. 289–292.

[10] G. Minami, Z. Xiong, A. Wang, and S. Mehrotra, "3-d wavelet coding of video with arbitrary regions of support," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 9, pp. 1063–1068, Sep 2001.

[11] Y. Wangi, J. T. Rucker, and J. E. Fowler, "Embedded wavelet-based compression of hyperspectral imagery using tarp coding," in *Proc. Int. Geosci. Remote Sens. Symp.*, July 2003, vol. 3, pp. 2027–2029.

[12] J. Tian and R. O. Wells, "Embedded image coding using wavelet difference reduction," *Wavelet Image and Video Compression*, vol. 50, pp. 289–301, 2002.

[13] M. D. Adams, "The JPEG-2000 still image compression standard," 2001.

[14] E. Magli, "Multiband lossless compression of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 4, pp. 1168–1178, April 2009.

[15] R. E. Roger and M. Cavenor, "Lossless compression of AVIRIS images," *IEEE Trans. Image Process.*, vol. 5, no. 5, pp. 713–719, May 1996.

[16] F. Rizzo, B. Carpentieri, G. Motta, and J. A. Storer, "High performance compression of hyperspectral imagery with reduced search complexity in the compressed domain," in *Proc. Data Compression Conf.*, March 2004, pp. 479–488.

[17] C. C. Lin and Y. T. Hwang, "An efficient lossless compression scheme for hyperspectral images using two-stage prediction," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 3, pp. 558–562, July 2010.

[18] F. Rizzo, B. Carpentieri, G. Motta, and J. A. Storer, "Low-complexity lossless compression of hyperspectral imagery via linear prediction," *IEEE Signal Process. Lett.*, vol. 12, no. 2, pp. 138–141, Feb. 2005.

[19] M. Klimesh, "Low-complexity lossless compression of hyperspectral imagery via adaptive filtering," in *The Interplanetary Network Progress Report*, Jet Propulsion Laboratory, Pasadena, California, Nov. 2005, pp. 1–10.

[20] "Lossless multispectral & hyperspectral image compression CCSDS 123.0-B-1, Blue Book, May 2012," http://public.ccsds.org/publications/archive/123x0b1ec1.pdf, 2015 (accessed December 10, 2015).

[21] E. Magli, G. Olmo, and E. Quacchio, "Optimized onboard lossless and near-lossless compression of hyperspectral data using CALIC," *IEEE Trans. Geosci. Remote Sens.*, vol. 1, no. 1, pp. 21–25, Jan 2004.

[22] J. Mielikainen and B. Huang, "Lossless compression of hyperspectral images using clustered linear prediction with adaptive prediction length," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 6, pp. 1118–1121, Nov. 2012.

[23] J. Wu, W. Kong, J. Mielikainen, and B. Huang, "Lossless compression of hyperspectral imagery via clustered differential pulse code modulation with removal of local spectral outliers," *IEEE Signal Processing Letters*, vol. 22, no. 12, pp. 2194–2198, Dec 2015.

[24] B. Aiazzi, P. Alba, L. Alparone, and S. Baronti, "Lossless compression of multi/hyper-spectral imagery based on a 3-d fuzzy prediction," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 5, pp. 2287–2294, Sep 1999.

[25] S. Golomb, "Run-length encodings (corresp.)," *IEEE Trans. Inf. Theory*, vol. 12, no. 3, pp. 399–401, Jul. 1966.

[26] M. J. Weinberger, G. Seroussi, and G. Sapiro, "The LOCO-I lossless image compression algorithm: principles and standardization into JPEG-LS," *IEEE Trans. Image Process.*, vol. 9, no. 8, pp. 1309–1324, Aug. 2000.

[27] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, June 2005, vol. 2, pp. 60–65 vol. 2.

[28] W. Liu, Jose. C. Principe, and S. Haykin, *Kernel Adaptive Filtering: A Comprehensive Introduction*, Wiley Publishing, 1st edition, 2010.

[29] "Hyperspectral remote sensing scenes data," http://www.ehu.es/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes, 2014 (accessed August 30, 2014).

[30] "The software implementation of CCSDS 123.0-B-1, Blue Book, May 2012," https://amstel.estec.esa.int/tecedm/misc/ESA_OSS_license.html, 2016 (accessed April 10, 2016).

[31] E. Aug, J. E. Snchez, A. Kiely, I. Blanes, and J. Serra-Sagrist, "Performance impact of parameter tuning on the CCSDS-123 lossless multi- and hyperspectral image compression standard," *Journal of Applied Remote Sensing*, vol. 7, no. 1, pp. 074594–074594, 2013.