

Article

Efficient Lossless Compression of Multitemporal Hyperspectral Image Data

Hongda Shen ^{1,†}, Zhuocheng Jiang ² and W. David Pan ^{2,*}

¹ Bank of America Corporation, New York, NY 10020, USA; hongdadeelearning@gmail.com

² Department of Electrical and Computer Engineering, University of Alabama in Huntsville, Huntsville, AL 35899, USA; zj0004@uah.edu

* Correspondence: pand@uah.edu

† Note: This paper is continuation of the first author's PhD work, which is not associated with his current affiliation.

Received: 14 November 2018; Accepted: 27 November 2018; Published: 2 December 2018



Abstract: Hyperspectral imaging (HSI) technology has been used for various remote sensing applications due to its excellent capability of monitoring regions-of-interest over a period of time. However, the large data volume of four-dimensional multitemporal hyperspectral imagery demands massive data compression techniques. While conventional 3D hyperspectral data compression methods exploit only spatial and spectral correlations, we propose a simple yet effective predictive lossless compression algorithm that can achieve significant gains on compression efficiency, by also taking into account temporal correlations inherent in the multitemporal data. We present an information theoretic analysis to estimate potential compression performance gain with varying configurations of context vectors. Extensive simulation results demonstrate the effectiveness of the proposed algorithm. We also provide in-depth discussions on how to construct the context vectors in the prediction model for both multitemporal HSI and conventional 3D HSI data.

Keywords: lossless compression; multitemporal hyperspectral images; information theoretic analysis; predictive coding

1. Introduction

Hyperspectral imaging (HSI) technologies have been widely used in many applications of remote sensing (RS) owing to the high spatial and spectral resolutions of hyperspectral images [1]. In some applications (e.g., hyperspectral imaging change detection [2–5]), we need to collect a sequence of hyperspectral images over the same spatial area at different times. The set of hyperspectral images collected over one location at varying time points is called multitemporal hyperspectral images [6,7]. From these multitemporal images, changes of observed locations over time can be detected and analyzed. Figure 1 illustrates a typical multitemporal hyperspectral image dataset. Each stack represents one 3D HSI. A sequence of 3D HSI stacks are captured by the HSI sensor over time.

Hyperspectral datasets tend to be of very large sizes. In the case of 4D multitemporal HSI datasets, the accumulated data volume increases very rapidly (to the Gigabyte or even Terabyte level), thereby making data acquisition, storage and transmission very challenging, especially when network bandwidth is severely constrained. As the number of hyperspectral images grows, it is clear that data compression techniques play a crucial role in the development of hyperspectral imaging techniques [8,9]. Lossy compression can significantly improve the compression efficiency, albeit at the cost of selective information loss. However, the fact that human visual systems are not sensitive to certain types and levels of distortions caused by information loss makes lossy compression useful. While lossy compression methods typically provide much larger data reduction than lossless methods,

they might not be suitable for many accuracy-demanding hyperspectral imaging applications, where the images are intended to be analyzed automatically by computers. Since lossless compression methods can strictly guarantee no loss in the reconstructed data, lossless compression would be more desirable in these applications.

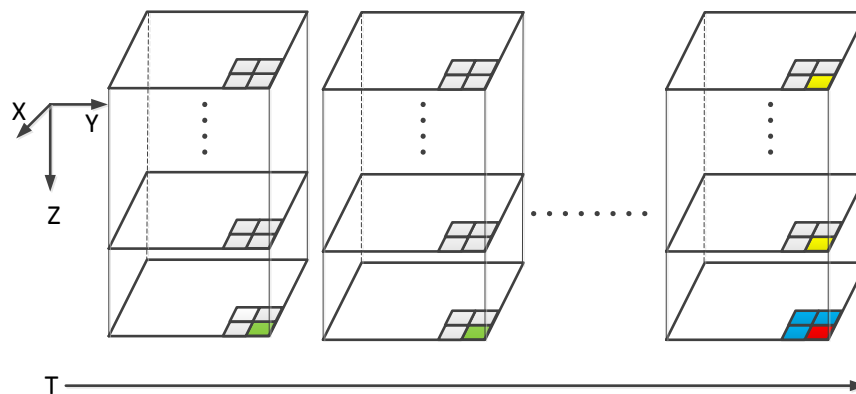


Figure 1. A multitemporal hyperspectral image dataset, where X and Y are the spatial directions, Z is the spectral direction, and T is the temporal direction.

Many efforts have been made to develop efficient lossless compression algorithms for 3D HSI data. LOCO-I [10] and 2D-CALIC [11] utilize spatial redundancy to reduce the entropy of prediction residuals. To take advantage of strong spectral correlations in HSI data, 3D compression methods have been proposed, which includes 3D-CALIC [12], M-CALIC [13], LUT [14] and its variants, SLSQ [15] and CCAP [16]. Also, some transform-based methods, such as SPIHT [17], SPECK [18], etc., can be easily extended to lossless compression even though they were designed for lossy compression.

Recently, clustering techniques have been introduced into 3D HSI data lossless compression and produced state-of-the-art performance over publicly available datasets. In [19], B. Aiazzi et al. proposed a predictive method leveraging crisp or fuzzy clustering to produce state-of-the-art results. Later, authors in both [20,21] again utilized the K-means clustering algorithm to improve the compression efficiency. Although these methods can yield higher compression, their computational costs are significantly higher than regular linear predictive methods. Plus, it is very difficult to parallel the process to leverage hardware acceleration if clustering technique is required as a preprocessing step in those approaches. In addition to the goal of reducing the entropy of either prediction residuals or transform coefficients, low computational complexity is another influential factor because many sensing platforms have very limited computing resources. Therefore, a low-complexity method called the “Fast Lossless” (FL) method, proposed by the NASA Jet Propulsion Lab (JPL) in [22], was selected as the core predictor in the Consultative Committee for Space Data Systems (CCSDS) new standard for multispectral and hyperspectral data compression [23], to provide efficient compression on 3D HSI data. This low-complexity merit also enables efficient multitemporal HSI data compression.

Multitemporal HSI data has an additional temporal dimension compared to 3D HSI data. Therefore, we can take advantage of temporal correlations to improve the overall compression efficiency of 4D HSI data. Nonetheless, there is very sparse work on lossless compression of multitemporal HSI data in the literature. Mamun et al. proposed a 4D lossless compression algorithm in [24], albeit lacking details on the prediction algorithms. In [25], a combination of Karhunen-Loève Transform (KLT), Discrete Wavelet Transform (DWT), and JPEG 2000 was applied to reduce the spectral and temporal redundancy of 4D remote sensing image data. However, the method can only achieve lossy compression. Additionally, Zhu et al. proposed another lossy compression approach for multitemporal HSI data in [7], based on a combination of linear prediction and a spectral concatenation of images. For the first time, we addressed lossless compression of multitemporal HSI data in [6], by introducing a *correntropy* based Least Mean Square filter for the Fast Lossless (FL) predictor. While the benefit

of exploiting temporal correlations in compression has been demonstrated by some papers such as [26,27], in this work, we conduct an in-depth information-theoretic analysis on the amount of compression achievable on multitemporal HSI data, by taking into account both the spectral and temporal correlations. On the other hand, this additional temporal decorrelation definitely poses a greater challenge to data processing speed especially for those powerful but computationally expensive algorithms, e.g., [19–21]. Therefore, we propose a low-complexity linear prediction algorithm, which extends the well-known FL method into a 4D version to achieve higher data compression, by better adapting to the underlying statistics of multitemporal HSI data. Note that most existing 3D HSI compression methods can be extended into 4D versions with proper modifications. However, this is beyond the scope of this paper.

The remainder of this paper is organized as follows. First, in Section 2, we give an overview of the multitemporal HSI datasets used in the study, which include three publicly available datasets, as well as two multitemporal HSI datasets we generated by using hyperspectral cameras. In Section 3, we present the information-theoretic analysis, followed by the introduction of a new algorithm for multitemporal HSI data lossless compression in Section 4. Finally, we present simulation results in Section 5 and make some concluding remarks in Section 6.

2. Datasets

Since there is little prior work on multi-temporal hyperspectral image compression, publicly available multi-temporal HSI datasets are very rare. Currently, the time-lapse hyperspectral radiance images of natural scenes [28] are the only available datasets to our best knowledge. Therefore, we created another two datasets capturing two scenes of Alabama A&M University campus using the portable Surface Optics Corporation (SOC) 700 hyperspectral camera [29] to enrich the relevant resources and facilitate further research. Hence, we introduce both data sources especially our datasets in detail before the actual analysis and algorithm development.

2.1. Time-Lapse Hyperspectral Imagery

Time-lapse hyperspectral imaging technology has been used for various remote sensing applications due to its excellent capability of monitoring regions-of-interest over a period of time. Time-lapse hyperspectral imagery is a sequence of 3D HSIs captured over the same scene but at different time stamps (often at a fixed time interval). Therefore, time-lapse hyperspectral imagery can be considered as a 4D dataset, whose size increases significantly with the total number of time stamps.

In [28], the authors made public several sequences of hyperspectral radiance images of scenes undergoing natural illumination changes. In each scene, hyperspectral images were acquired at about one-hour intervals. We randomly selected three 4D time-lapse HSI datasets, Levada, Gualtar and Nogueiro. Basic information of these three datasets are listed in the Table 1. Detailed information of these datasets can be found in [30]. Each single HSI has the same spatial size, 1024×1344 , with 33 spectral bands. Both Gualtar and Nogueiro have nine time stamps while Levada has seven. Note that the original data for these datasets was linearly mapped into $[0, 1]$ and stored using “double” floating point format (64 bits) [28]. In order to evaluate the prediction-based lossless compression performance of algorithms, we pre-process the datasets by re-mapping the data samples back to their original precision of 12 bits. The resulting sizes of the datasets range from 454.78 MB (for seven frames) to 584.71 MB (for nine frames).

Figure 2 shows the Levada, Nogueiro and Gualtar sequences from top to bottom. Detailed information about the Levada sequence can be found in [28]. Note that only 2D color-rendered RGB (Red, Green and Blue) images are shown in Figure 2 instead of the actual HSI data for display purpose. Since time-lapse HSIs are captured over the same scene at different time instants with gradually changing natural illumination, we can see that images at different time instants are very similar in Figure 2. These temporal correlations can be exploited to improve the overall compression efficiency.

Table 1. Multitemporal hyperspectral image datasets.

Dataset	Size	The Number of Time Frames	Precision (bits)
Levada	1024 × 1344 × 33	7	12
Gualtar	1024 × 1344 × 33	9	12
Nogueiro	1024 × 1344 × 33	9	12
Scene-1	640 × 640 × 120	21	12
Scene-2	640 × 640 × 120	16	12



Figure 2. Some sample images at different time instants from the time-lapse hyperspectral image datasets (from top to bottom: Levada, Nogueiro and Gualtar).

2.2. AAMU Datasets

Due to very few 4D HSI datasets available in the public domain, we created some new datasets to increase the data diversity of our study. To this end, we used a SOC 700 hyperspectral camera (manufactured by Surface Optics Corporation, CA, USA) and produced 4D datasets for two scenes on the campus of Alabama A&M University (AAMU). The SOC 700 camera can record and process hyperspectral imagery at a rate of 15 megabytes of data every second (120-band elements per second at 12-bit resolution, 640 pixels per row, 100 rows per second). The imaging system’s spectral response covers the visible and near-infrared spectral range (from 0.43 to 0.9 microns), and can be used in normal to low lighting conditions with variable exposure times and display gains. More detailed about the SOC 700 system can be found at [29].

We placed the camera at two distinct locations of the AAMU campus and generated two datasets, which we call Scene-1 and Scene-2. 3D HSI cubes in Scene-1 and Scene-2 are of the same size: 640 × 640 × 120 with 21 and 16 time frames, respectively. The overall dataset sizes of Scene-1 and Scene-2 are roughly 1.70 GB and 850 MB, respectively. Compared to three time-lapse datasets discussed earlier, these two AAMU datasets are much larger, making themselves more suitable for evaluating compression efficiencies. In contrast to the time-lapse datasets, the images of the AAMU datasets were acquired at time-varying rates of approximately one per five minutes or one per minute, thereby introducing time-varying temporal correlations through the entire dataset. This special feature will allow us to investigate the relationship between prediction accuracy and correlations at different levels.

Figure 3 shows the 2D color-rendered RGB images for a few time instants for the AAMU multitemporal HSI datasets. While changing illumination conditions over time can be observed in both datasets, temporal similarity in both pixel intensity and image structure is also obvious, similar to the three time-lapse datasets shown in Figure 2. In order to quantify the potential gain on compression achievable by exploiting the temporal correlations in 4D HSI datasets, we conducted an information-theoretic analysis as detailed in the next section.

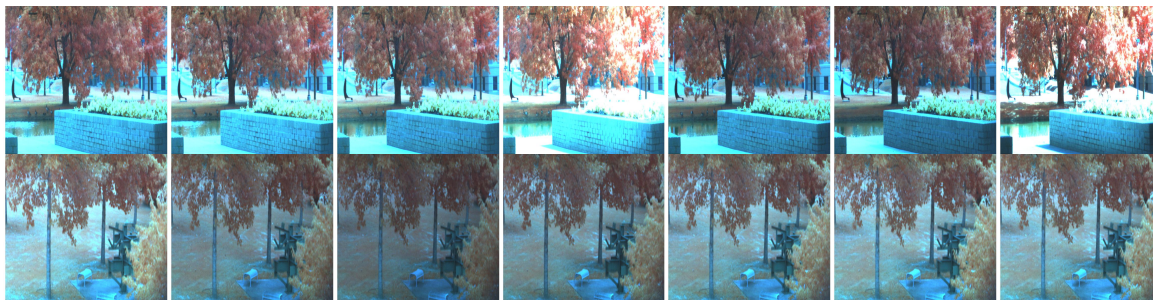


Figure 3. Sample images at different time instants from the AAMU hyperspectral image datasets (top: Scene-1, and bottom: Scene-2).

3. Problem Analysis

While the actual amount of compression achieved depends on the choice of specific compression algorithms [31], information theoretic analysis can provide us an upper bound on the amount of compression achievable. Here we focus on analyzing how temporal correlation can help improve the compression of 4D hyperspectral image datasets, as opposed to the baseline 3D-compression case where only spatial and spectral correlations are considered.

Let X_j^t be a 4D hyperspectral image at the t th time instant and j th spectral band where X represents a two-dimensional image with K distinct pixel values v_i ($i \in \{1, \dots, K\}$) within each band. Then the entropy of this source can be obtained based on the probabilities $p(v_i)$ of these values by

$$H(X_j^t) = - \sum_{i=1}^K p(v_i) \cdot \log_2 [p(v_i)]. \tag{1}$$

If we assume that there are no dependencies between pixels of X_j^t , at least $H(X_j^t)$ bits must be spent on average for each pixel of this image. However, for typical 4D hyperspectral images, this assumption does not hold given the existence of spatial, spectral and temporal correlations. The value of a particular pixel might be similar to some other pixels from its spatial, spectral or temporal neighborhoods (contexts). Considering these correlations can lead to reduced information (fewer bits to code each pixel on average) than the entropy $H(X_j^t)$. The conditional entropy of the image captures the correlations as follows:

$$H(X_j^t|C_j^t) = - \sum_{i=1}^K p(v_i|C_j^t) \cdot \log_2 [p(v_i|C_j^t)]. \tag{2}$$

where C_j^t denoted as *context*, which represents a group of correlated pixels. In general, conditioning reduces entropy, $H(X_j^t|C_j^t) \leq H(X_j^t)$.

The choice of context largely determines how much compression we can achieve by using prediction-based lossless compression schemes. One should include highly-correlated pixels into the context. Spectral and temporal correlations are typically much stronger than spatial correlations in multitemporal hyperspectral images. For example, ref. [20] claims that explicit spatial de-correlation is not always necessary to achieve good compression [31]. Ref. [31] shows that a linear prediction scheme was adequate for spectral and/or temporal prediction because of high degree of correlations, in contrast to non-linearity nature of spatial de-correlation. Therefore, we construct the context vector C_j^t using only pixels from previous bands at the same spatial location, as well as pixels from the same spectral band at the same location but at previous time points. Specifically, we denote pixels from previous N_b bands at the same spatial location as $X_{j-m}^t, m \in \{1, 2, \dots, N_b\}$ (yellow pixels in Figure 1), and pixels from the same spectral band at the same location but from previous N_t temporal positions $X_j^{t-n}, n \in \{1, 2, \dots, N_t\}$ (green pixels in Figure 1), respectively. Then, the conditional entropy in Equation (2) becomes

$$H(X_j^t | X_{j-m}^t, X_j^{t-n}) = - \sum_{i=1}^K p(v_i | X_{j-m}^t, X_j^{t-n}) \cdot \log_2 \left[p(v_i | X_{j-m}^t, X_j^{t-n}) \right]. \quad (3)$$

By using the relation between joint entropy and conditional entropy, we can further rewrite Equation (3) as

$$H(X_j^t | X_{j-m}^t, X_j^{t-n}) = H(X_j^t, X_{j-m}^t, X_j^{t-n}) - H(X_{j-m}^t, X_j^{t-n}), \quad (4)$$

which enables a simple algorithm for estimation of the conditional entropies. It suffices to estimate the above two joint entropies by counting the occurrence frequency of each $(N_t + N_b + 1)$ -tuple in the set $(X_j^t, X_{j-m}^t, X_j^{t-n})$, and $(N_t + N_b)$ -tuple in the set (X_{j-m}^t, X_j^{t-n}) , respectively. However, as pointed out in [31], the entropy estimates become very inaccurate when two or more previous bands are used for prediction in practice. The reason is that, as the entropy is conditioned upon multiple bands, the set $(X_j^t, X_{j-m}^t, X_j^{t-n})$ takes on values from the alphabet $\chi^{(N_t+N_b+1)}$, whose size can become extremely large, e.g., $(2^{12})^{N_t+N_b+1}$ for our datasets. As a consequence, a band might not contain enough pixels to provide statistically meaningful estimates of the probabilities. Similar to the “data source transform” trick proposed in [31], we consider each bit-plane of X_j^t as one separate binary source. Although binary sources greatly reduce the alphabet size, which makes it possible to obtain accurate entropy estimates, results obtained for the binary source are not very representative of the actual bit rates obtained by a practical coder since statistical dependencies between those bit-planes cannot be neglected. However, using bit-plane sources would be useful for our study since our main goal is to evaluate the relative instead of the absolute performance gain achievable by using different contexts based on various combinations of spectral and temporal bands. Therefore, we will compute the conditional entropy in Equation 3, for all the bit-planes separately, and then take their average to be the overall performance gain for a specific prediction context. In this sense, we extend the algorithm in [31] by incorporating previous temporal bands into the context vector, which allows us to estimate also the temporal correlations.

We applied this estimation algorithm on five multitemporal HSI datasets to estimate the potential compression performance of multitemporal hyperspectral image with a combination of various spectral and temporal bands. $H(p, q)$ is the entropy conditioned to p previous bands at the current time point and q bands at current spectral band but from previous time points for prediction. Using the binary-source based estimation method, we summed up the conditional entropies of all the bit-planes (a total of 12 bit-planes for all our datasets) as the estimation of $H(p, q)$ for each band of the dataset. Then the averages of $H(p, q)$ over all bands are reported in Figure 4 for all five datasets. Due to limited space, we only show results for parameters p and q chosen between 0 and 5. More detailed results for other datasets can be found in Tables A1–A5 in Appendix A.

From Figure 4, we can observe that as either p or q increases, the general trend is that the conditional entropy decreases; however, as p or q further increases (e.g, from 4 to 5), the reduction of entropy becomes smaller than the case of either p or q going from 0 to 1. This means that including a few previous bands either spectrally or temporally in the context can be very useful to improving the performance of the prediction-based compression algorithms, but the return of adding more bands from distant past will diminish as the correlations get weaker, let alone the increased computational cost associated with involving excessive number of bands for prediction. In addition, the conditional entropy tends to decrease faster with an increased p than with an increased q . This is indicative of stronger spectral correlations than temporal correlations. For example, the fourth image in the first row of Figure 3 represents a dramatic change of illumination conditions during image capturing, thereby weakening the temporal correlations. However, there still exist significant temporal correlations which, if exploited properly, can lead to improved compression by considering only spectral correlations. To this end, we propose a compression algorithm, which exploits temporal correlations in multitemporal HSI data to enhance the overall compression performance.

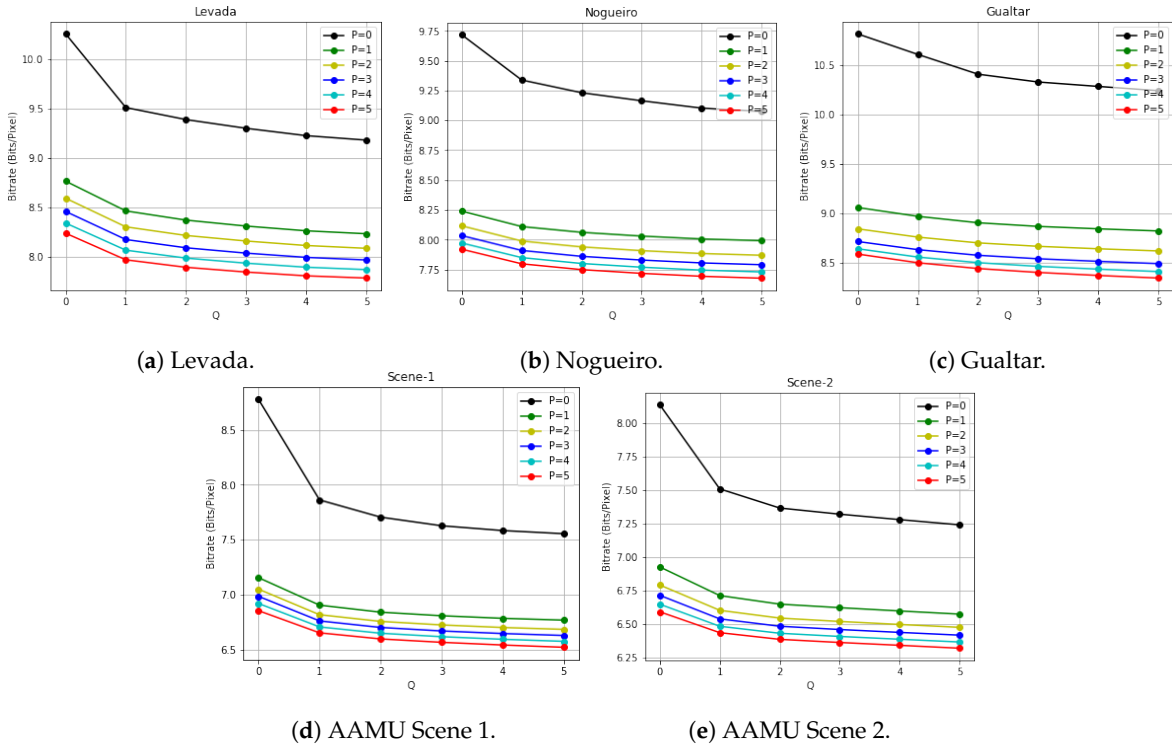


Figure 4. Conditional entropies over five datasets for different P and Q combination.

4. Proposed Algorithm

Our lossless compression algorithm is based on predicting the pixels to be coded, by using a linear combination of those pixels already coded (in a neighboring causal context). Prediction residuals are obtained by subtracting the actual pixel values from their estimates. The residuals are then encoded using entropy coders.

For multitemporal hyperspectral image, the estimate of a pixel value can be obtained by

$$\hat{x}_{p,q}^{j,t} = \mathbf{w}_{j,t}^T \mathbf{y}_{p,q}^{j,t} \tag{5}$$

where $\hat{x}_{p,q}^{j,t}$ represents an estimate of a pixel, $x_{p,q}^{j,t}$ at spatial location (p, q) , the j th band and the t th time point, while $\mathbf{w}_{j,t}$ denotes the weights for linearly combining the pixel values $\mathbf{y}_{p,q}^{j,t}$. These pixels are drawn from a causal context of several previously coded bands either at the same time point or at previous time points. More specifically, $\mathbf{y}_{p,q}^{j,t} = [\mathbf{x}_{p,q}^{j-m,t}, \mathbf{x}_{p,q}^{j,t-n}]$, $m \in \{1, 2, \dots, N_b\}$, $n \in \{1, 2, \dots, N_t\}$.

For accurate prediction, weights should be able to adapt to locally changing statistics of pixel values in the multitemporal HSI data. For this sake, learning algorithms were introduced for lossless compression of 3D HSI data [22,32]. Adaptive learning was also used in the so-called Fast Lossless (FL) method. Due to its low-complexity and effectiveness, the FL method has been selected as a new compression standard for multispectral and hyperspectral data by CCSDS (Consultative Committee for Space Data Systems) [23]. The core learning algorithm of the FL method is the *sign* algorithm, which is a variant of least mean square (LMS). In prior work, we proposed another LMS variant, called *correntropy* based LMS (CLMS) algorithm, which uses the Maximum Correntropy Criterion [6,8] for lossless compression of 3D and multi-temporal HSI data. By replacing the cost function for LMS based learning with correntropy [33], the CLMS method introduces a new term in the weight update function, which allows the learning rate to change, in order to improve on the conventional LMS based method with a constant learning rate. However, good performance of the CLMS method depends heavily on proper tuning of the kernel variance, which is an optimization parameter used by the “kernel trick” associated with the correntropy. To avoid the excessive need to tune the kernel variance for various

types of images in the multitemporal HSI datasets, we adopted the sign algorithm used by the FL predictor with an expanded context vector.

In order to exploit spatial correlations also found in hyperspectral datasets, we follow the simple approach in [22], where local-mean-removed pixel values are used as input to our linear predictor. Specifically, for an arbitrary pixel $x_{p,q}^{j,t}$ in a multi-temporal HSI, the spatial local mean is

$$\mu_{p,q}^{j,t} = [x_{p-1,q-1}^{j,t} + x_{p-1,q}^{j,t} + x_{p,q-1}^{j,t}] / 3. \quad (6)$$

After mean subtraction, the causal context becomes

$$\mathbf{U}_{p,q}^{j,t} = [x_{p,q}^{j-m,t} - \mu_{p,q}^{j-m,t}, x_{p,q}^{j,t-n} - \mu_{p,q}^{j,t-n}], \quad (7)$$

where $m \in \{1, 2, \dots, N_b\}$, and $n \in \{1, 2, \dots, N_t\}$. To simplify the notation, we represent the spatial location (p, q) with a single index, k , in that $k = (p - 1) * N_y + q$, where N_y refers to number of pixels in each row within one band. In other words, we line up the pixels in a 1D vector, where the pixels will be processed sequentially in the iterative optimization process of the sign algorithm. Now the predicted value for an arbitrary pixel in each band of multi-temporal HSI dataset is given by

$$\hat{x}_k^{j,t} = \mu_k^{j,t} + \mathbf{w}_{j,t}^T \mathbf{U}_k^{j,t}, \quad (8)$$

where $\mathbf{w}_{j,t}^T$ are the weights to be adapted sequentially for each band. It follows that the prediction residual can be obtained as

$$e_k^{j,t} = x_k^{j,t} - \hat{x}_k^{j,t} = x_k^{j,t} - (\mu_k^{j,t} + \mathbf{w}_{j,t}^T \mathbf{U}_k^{j,t}). \quad (9)$$

We apply the sign algorithm to iteratively update the weights as

$$\mathbf{w}_{j,t}^{k+1} = \mathbf{w}_{j,t}^k + \rho(k) \cdot \text{sign}(e_k^{j,t}) \cdot \mathbf{U}_k^{j,t}, \quad (10)$$

where $\rho(k)$ is an adaptive learning rate proposed in [23] to achieve fast convergence to solutions close to the global optimum. Our study found that using this adaptive learning rate can provide good results on multitemporal datasets. Note that we need to reset the weights and learning rate for each new band in the dataset to account for potentially varying statistics.

After prediction, all the residuals are mapped to non-negative values [23] and then coded into bitstream losslessly by using the Golomb-Rice Codes (GRC) [34]. Although GRC is selected as the entropy coder because of its computational efficiency [35], we observed that using arithmetic coding can offer slightly lower bitrates, albeit at a much higher computational cost. Pseudo Algorithm 1 of this 4D extension of Fast-Lossless is given to better show its structure and workflow.

Algorithm 1 Fast-Lossless-4D Predictor

Input:
 1) 4D HSI data \mathbf{X} .
 2) T (Total # of time frames of \mathbf{X}).
 3) B (Total # of spectral bands for each time frame of \mathbf{X})
 4) P (# of previous spectral bands).
 5) Q (# of bands from previous time frames).
 6) μ (learning rate).
for $t = 1:T$ **do**
 for $b = 1:B$ **do**
 Local mean subtracted data \mathbf{U} using Equations (6) and (7).
 initialize: $\mathbf{w}_{j,t} = \mathbf{0}$.
 for each pixel in this band in raster scan order **do**
 Output residual $e^{j,t}$ using Equation (9).
 Updating $\mathbf{w}_{j,t}$ using Equation (10).
 end for
 end for
end for

5. Simulation Results

We tested the proposed algorithm on all five multitemporal HSI datasets (Levada, Gualtar, Nogueiro, AAMU scene_1 and AAMU scene_2). To show the performance of our algorithm, we present the bitrates after compression in Figure 5 (Detailed results can be found in Tables A6–A10). Similar to the conditional entropy estimation results in Section 3, the bitrates were obtained by using various combinations of p (spectral) and q (temporal) number of bands for causal contexts.

We can see that for the case of $p = 0$ and $q = 0$, where we simply use mean subtraction (for spatial decorrelation) without spectral and temporal decorrelation, we can already achieve significant amount of compression on the input data by lowering the original bitrate from 12 bits/pixel to about 6 bits/pixel. If we consider either spectral or temporal correlations, or both, we can achieve additional compression gains on multitemporal HSI data. For example, the bit rate can be reduced by approximately 1 bit/pixel or 0.2 bit/pixel by including in the prediction context one more previous band spectrally or temporally. Generally, the bitrates decrease with more bands being included in the context, which agrees well with the results on condition entropy estimation in Section 3. Furthermore, if we fix the p value and increase the q value, and vice versa, we can achieve better compression. However, the return on including more bands will diminish gradually as p and q further increase. In some cases, we can even have less compression if the context includes some remote bands that might be weakly correlated with the pixels to be predicted. For example, in Table A7, when $p = 5$ and $q = 4$, the corresponding bit rate is 4.4953, which is 0.0026 higher than 4.4927 (when $p = 5$ and $q = 3$). Similar examples can be found in Table A9 (when $p = 5$ and $q = 4$) and in Table A10 (when $p = 5$ and $q = 1$). Including weakly correlated or totally uncorrelated pixels might lower the quality of the context, leading to degraded compression performance. In this same spirit, we can see that spectral decorrelation turns out to be more effective in reducing the bitrates than temporal decorrelation. This means that spectral correlations are stronger than temporal correlations in the datasets we tested. The reason can be that each hyperspectral image cube in these multitemporal HSI datasets was captured at time intervals of at least a few minutes, during which significant change of pixel values (e.g, caused by illumination condition changes) might have taken place. If the image capturing time interval is reduced, then we expect the stronger temporal correlations.

On the other hand, prediction using only one previous spectral band, and/or the same spectral band but from previous time instance can offer a low-complexity compressor with sufficiently good compression performance. The bitrate results show the wide range of tradeoffs for us to explore in order to balance compression performance with computational complexity.

We also compared the proposed algorithm (based on Fast-Lossless algorithm) with our previous work (based on correntropy LMS learning) in [6], namely CLMS, which seems to be the only existing work on lossless compression of multitemporal HSI data. For fair comparison, we use the same parameter setting in [6]. Although it would be straightforward to show the bitrates for both algorithms in multiple tables, we choose to visualize bitrates changes as number of previous bands or number of previous time frames increases for both algorithms. To reduce the complexity of this visualization, we only present $P = 3$ case since it is the default setting in the FL method. Figure 6 shows the bitrate changes for the five datasets. Note that when $Q = 0$, our method is essentially equivalent to 3D FL method. Therefore, we use green dashed line to mark FL method performance in Figure 6. While blue, green and red curves represent CLMS, FL method and ours respectively, it is clear that our methods produce the lowest bitrates consistently. Although bitrates of our method was only slightly lower than applying FL method directly to each one of time framed HSI, it outperformed CLMS method by a significant margin. The improvements on time-lapse datasets are more significant than on AAMU datasets in general. Consistent with results shown previously, we have higher compression gains in the spectral dimension than the temporal dimension. However, the results show that our algorithm can take advantage of the temporal correlations available to bring additional improvements on the overall compression performance.

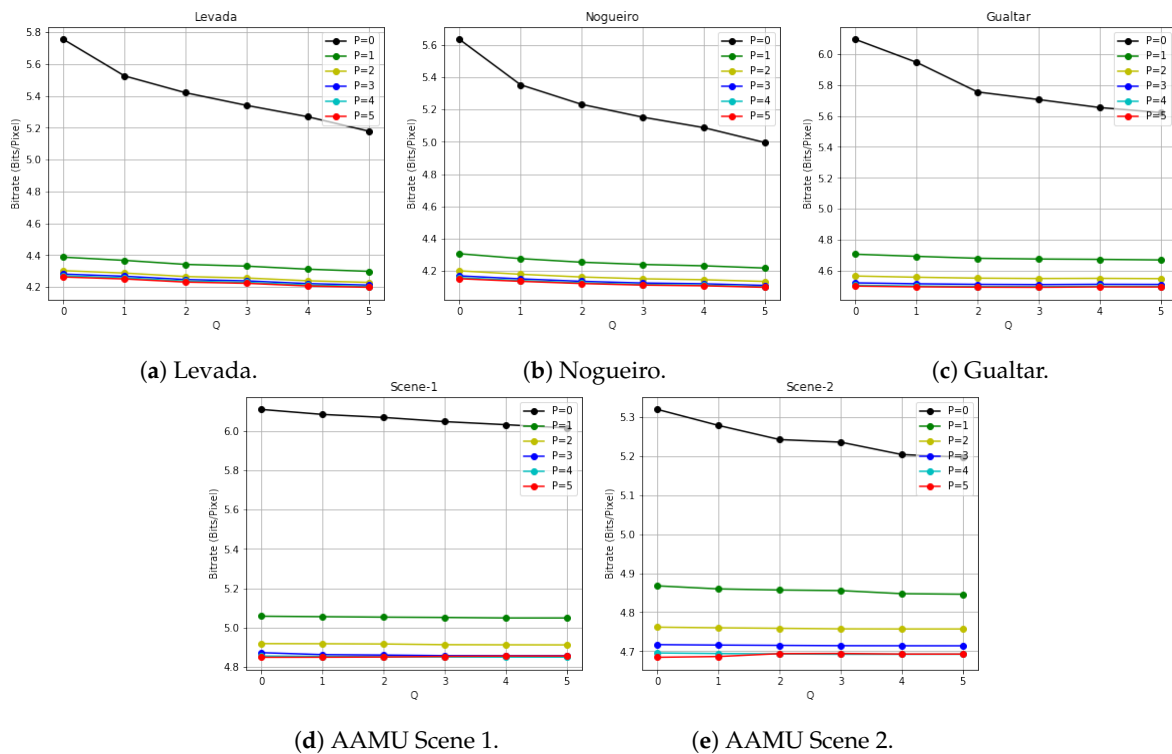


Figure 5. Bitrates over five datasets for different P and Q combination.

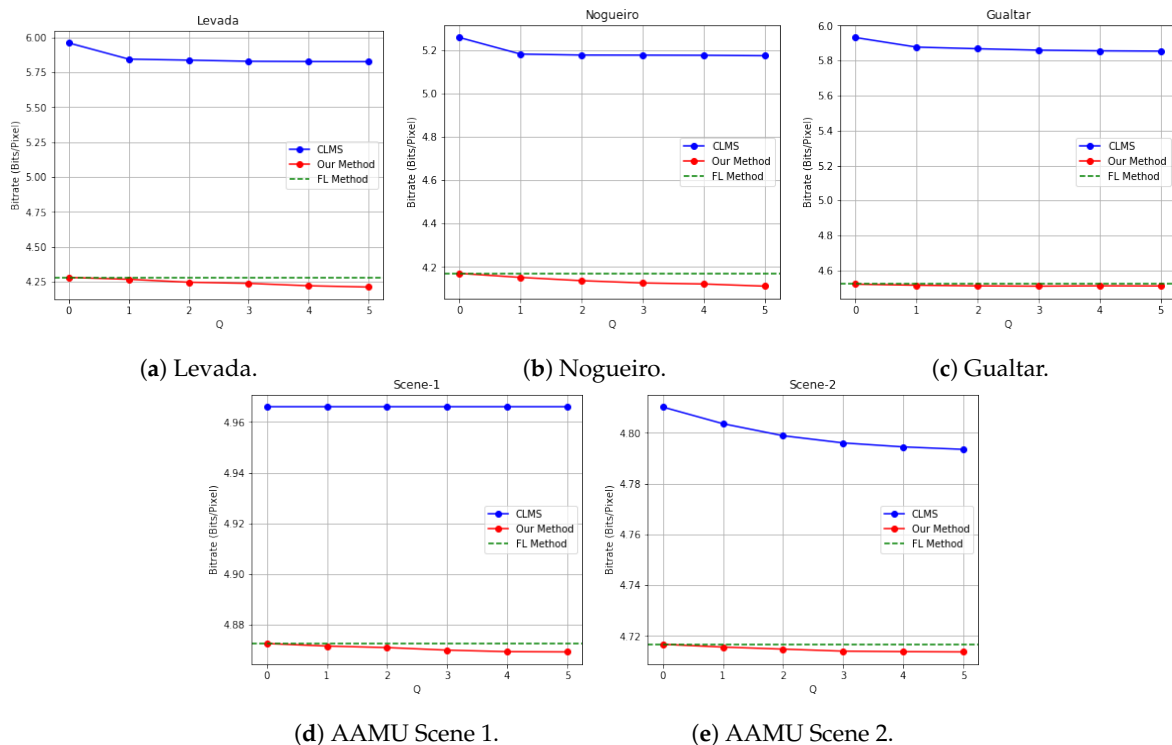


Figure 6. Bitrates comparison with CLMS and FL methods over five datasets for different Q when $P = 3$.

6. Conclusions

We have proposed a new predictive lossless compression algorithm for multitemporal time-lapse hyperspectral image data using a low-complexity sign algorithm with an expanded prediction context. Simulation results have demonstrated the outstanding capability of this algorithm to compress

multitemporal HSI data through spectral and temporal decorrelation. The actual compression results are congruent with the information theoretic analysis and estimation based on conditional entropy. We show that increasing the number of previous bands for prediction can yield better compression performance, by exploiting the spectral and temporal correlations in the datasets.

As future work, we intend to study how to adaptively select bands to build an optimal context vector for prediction. Also, we will investigate how to fully integrate the proposed algorithm and the analytic framework to achieve real-time compression on streaming hyperspectral data. Furthermore, the proposed algorithm can be extended to lossless compression of regions-of-interest in hyperspectral images, which can offer much higher compression than compressing the entire hyperspectral image dataset.

Author Contributions: Conceptualization, H.S. and W.D.P.; Methodology, H.S. and W.D.P.; Software, H.S.; Validation, H.S., Z.J., and W.D.P.; Formal Analysis, H.S.; Investigation, H.S.; Resources, H.S. and W.D.P.; Data Curation, H.S.; Writing-Original Draft Preparation, H.S.; Writing-Review & Editing, H.S., Z.J. and W.D.P.; Visualization, H.S.; Supervision, W.D.P.; Project Administration, W.D.P.; Funding Acquisition, W.D.P.

Funding: This research received no external funding.

Acknowledgments: We would like to thank Joel Fu of the Computer Science Program, Alabama A&M University, Normal, AL, for providing facilities including a SOC 700 hyperspectral camera for data collection in this research.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Condition Entropy Estimation Empirical Experimental Results

Table A1. Conditional entropies $H(p, q)$ (bits/pixels), computed for various values of p and q , over “Levada”.

p	$q = 0$	$q = 1$	$q = 2$	$q = 3$	$q = 4$	$q = 5$
0	10.2558	9.5120	9.3905	9.3022	9.2279	9.1827
1	8.7675	8.4673	8.3744	8.3141	8.2663	8.2366
2	8.5951	8.3063	8.2192	8.1633	8.1179	8.0895
3	8.4613	8.1782	8.0949	8.0410	7.9979	7.9719
4	8.3440	8.0707	7.9908	7.9394	7.8984	7.8742
5	8.2414	7.9733	7.8977	7.8497	7.8113	7.7888

Table A2. Conditional entropies $H(p, q)$ (bits/pixels), computed for various values of p and q , over “Gualtar”.

p	$q = 0$	$q = 1$	$q = 2$	$q = 3$	$q = 4$	$q = 5$
0	10.8137	10.6057	10.4057	10.3266	10.2818	10.2387
1	9.0550	8.9657	8.9007	8.8638	8.8393	8.8176
2	8.8388	8.7553	8.6971	8.6623	8.6374	8.6155
3	8.7107	8.6283	8.5715	8.5362	8.5100	8.4870
4	8.6375	8.5534	8.4964	8.4592	8.4310	8.4064
5	8.5837	8.4956	8.4371	8.3975	8.3670	8.3406

Table A3. Conditional entropies $H(p, q)$ (bits/pixels), computed for various values of p and q , over “Nogueiro”.

p	$q = 0$	$q = 1$	$q = 2$	$q = 3$	$q = 4$	$q = 5$
0	9.7167	9.3381	9.2314	9.1641	9.1023	9.0740
1	8.2400	8.1105	8.0619	8.0304	8.0064	7.9918
2	8.1165	7.9890	7.9395	7.9080	7.8841	7.8695
3	8.0343	7.9097	7.8603	7.8292	7.8054	7.7904
4	7.9722	7.8497	7.8005	7.7695	7.7453	7.7299
5	7.9200	7.7979	7.7492	7.7180	7.6936	7.6775

Table A4. Conditional entropies $H(p, q)$ (bits/pixels), computed for various values of p and q , over “Scene-1”.

p	$q = 0$	$q = 1$	$q = 2$	$q = 3$	$q = 4$	$q = 5$
0	8.7785	7.8624	7.7055	7.6271	7.5834	7.5548
1	7.1538	6.9037	6.8398	6.8070	6.7836	6.7671
2	7.0491	6.8166	6.7555	6.7233	6.6999	6.6829
3	6.9839	6.7599	6.7003	6.6683	6.6448	6.6272
4	6.9190	6.7052	6.6475	6.6160	6.5923	6.5736
5	6.8533	6.6513	6.5956	6.5646	6.5402	6.5193

Table A5. Conditional entropies $H(p, q)$ (bits/pixels), computed for various values of p and q , over “Scene-2”.

p	$q = 0$	$q = 1$	$q = 2$	$q = 3$	$q = 4$	$q = 5$
0	8.1367	7.5082	7.3654	7.3194	7.2787	7.2397
1	6.9248	6.7109	6.6466	6.6202	6.5955	6.5719
2	6.7897	6.6004	6.5421	6.5169	6.4945	6.4731
3	6.7126	6.5362	6.4812	6.4568	6.4354	6.4148
4	6.6461	6.4810	6.4289	6.4052	6.3843	6.3637
5	6.5892	6.4334	6.3836	6.3601	6.3390	6.3173

Table A6. Bit rates (bits/pixels), obtained for various values of p and q , on “Levada”.

p	$q = 0$	$q = 1$	$q = 2$	$q = 3$	$q = 4$	$q = 5$
0	5.7538	5.5248	5.4193	5.3400	5.2689	5.1784
1	4.3876	4.3674	4.3422	4.3313	4.3122	4.2987
2	4.3029	4.2878	4.2660	4.2570	4.2395	4.2293
3	4.2813	4.2679	4.2471	4.2389	4.2218	4.2125
4	4.2704	4.2578	4.2377	4.2298	4.2131	4.2043
5	4.2631	4.2508	4.2312	4.2235	4.2070	4.1985

Table A7. Bit rates (bits/pixels) obtained for various values of p and q , on “Gualtar”.

p	$q = 0$	$q = 1$	$q = 2$	$q = 3$	$q = 4$	$q = 5$
0	6.0941	5.9463	5.7547	5.7057	5.6541	5.6218
1	4.7064	4.6928	4.6802	4.6754	4.6732	4.6695
2	4.5655	4.5577	4.5518	4.5495	4.5505	4.5487
3	4.5214	4.5155	4.5117	4.5102	4.5121	4.5113
4	4.5057	4.5006	4.4977	4.4966	4.4990	4.4984
5	4.5004	4.4960	4.4936	4.4927	4.4953	4.4948

Table A8. Bit rates (bits/pixels), obtained for various values of p and q , on “Nogueiro”.

p	$q = 0$	$q = 1$	$q = 2$	$q = 3$	$q = 4$	$q = 5$
0	5.6329	5.3524	5.2312	5.1521	5.0874	4.9953
1	4.3070	4.2769	4.2545	4.2407	4.2324	4.2186
2	4.2013	4.1800	4.1631	4.1520	4.1463	4.1355
3	4.1701	4.1515	4.1364	4.1260	4.1212	4.1112
4	4.1586	4.1413	4.1272	4.1171	4.1127	4.1031
5	4.1525	4.1363	4.1228	4.1131	4.1088	4.0994

Table A9. Bit rates (bits/pixels), obtained for various values of p and q , on “AAMU Scene-1”.

p	$q = 0$	$q = 1$	$q = 2$	$q = 3$	$q = 4$	$q = 5$
0	6.1097	6.0844	6.0691	6.0478	6.0322	6.0156
1	5.0580	5.0557	5.0534	5.0518	5.0492	5.0491
2	4.9186	4.9181	4.9168	4.9133	4.9126	4.9122
3	4.8727	4.8717	4.8711	4.8701	4.8695	4.8694
4	4.8555	4.8537	4.8517	4.8507	4.8502	4.8501
5	4.8491	4.8495	4.8506	4.8538	4.8556	4.8556

Table A10. Bit rates (bits/pixels), obtained for various values of p and q , on “AAMU Scene-2”.

p	$q = 0$	$q = 1$	$q = 2$	$q = 3$	$q = 4$	$q = 5$
0	5.3202	5.2791	5.2429	5.2363	5.2045	5.1977
1	4.8675	4.8597	4.8569	4.8554	4.8475	4.8461
2	4.7615	4.7599	4.7586	4.7572	4.7569	4.7569
3	4.7167	4.7156	4.7148	4.7140	4.7138	4.7137
4	4.6953	4.6939	4.6928	4.6920	4.6919	4.6920
5	4.6842	4.6862	4.6936	4.6940	4.6925	4.6925

References

- Shen, H.; Pan, W.D.; Wu, D. Predictive Lossless Compression of regions-of-interest in Hyperspectral Images With No-Data Regions. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 173–182. [[CrossRef](#)]
- Thouvenin, P.A.; Dobigeon, N.; Tourneret, J.Y. A Hierarchical Bayesian Model Accounting for Endmember Variability and Abrupt Spectral Changes to Unmix Multitemporal Hyperspectral Images. *IEEE Trans. Comput. Imag.* **2018**, *4*, 32–45. [[CrossRef](#)]
- Marinelli, D.; Bovolo, F.; Bruzzone, L. A novel change detection method for multitemporal hyperspectral images based on a discrete representation of the change information. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017; pp. 161–164.
- Liu, S.; Bruzzone, L.; Bovolo, F.; Du, P. Unsupervised Multitemporal Spectral Unmixing for Detecting Multiple Changes in Hyperspectral Images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 2733–2748. [[CrossRef](#)]
- Ertürk, A.; Iordache, M.D.; Plaza, A. Sparse Unmixing-Based Change Detection for Multitemporal Hyperspectral Images. *IEEE J. Sel. Top. Appl. Earth Obs.* **2016**, *9*, 708–719. [[CrossRef](#)]
- Shen, H.; Pan, W.D.; Dong, Y. Efficient Lossless Compression of 4D Hyperspectral Image Data. In Proceedings of the 3rd International Conference on Advances in Big Data Analytics, Las Vegas, NV, USA, 25–28 July 2016.
- Zhu, W.; Du, Q.; Fowler, J.E. Multitemporal Hyperspectral Image Compression. *IEEE Geosci. Remote Sens. Lett.* **2011**, *8*, 416–420. [[CrossRef](#)]
- Shen, H.; Pan, W.D. Predictive lossless compression of regions-of-interest in hyperspectral image via Maximum Correntropy Criterion based Least Mean Square learning. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 2182–2186.
- Liaghati, A.; Pan, W.D.; Jiang, Z. Biased Run-Length Coding of Bi-Level Classification Label Maps of Hyperspectral. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 4580–4588. [[CrossRef](#)]
- Weinberger, M.J.; Seroussi, G.; Sapiro, G. The LOCO-I lossless image compression algorithm: Principles and standardization into JPEG-LS. *IEEE Trans. Image Process.* **2000**, *9*, 1309–1324. [[CrossRef](#)]
- Wu, X.; Memon, N. Context-based lossless interband compression-extending CALIC. *IEEE Trans. Image Process.* **2000**, *9*, 994–1001.

12. Magli, E.; Olmo, G.; Quacchio, E. Optimized onboard lossless and near-lossless compression of hyperspectral data using CALIC. *IEEE Trans. Geosci. Remote Sens.* **2004**, *1*, 21–25. [[CrossRef](#)]
13. Wu, X.; Memon, N. Context-based, adaptive, lossless image coding. *IEEE Trans. Commun.* **1997**, *45*, 437–444. [[CrossRef](#)]
14. Mielikainen, J. Lossless compression of hyperspectral images using lookup tables. *IEEE Signal Process. Lett.* **2006**, *13*, 157–160. [[CrossRef](#)]
15. Rizzo, F.; Carpentieri, B.; Motta, G.; Storer, J.A. Low-complexity lossless compression of hyperspectral imagery via linear prediction. *IEEE Signal Process. Lett.* **2005**, *12*, 138–141. [[CrossRef](#)]
16. Wang, H.; Babacan, S.D.; Sayood, K. Lossless Hyperspectral-Image Compression Using Context-Based Conditional Average. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 4187–4193. [[CrossRef](#)]
17. Said, A.; Pearlman, W.A. A new, fast, and efficient image codec based on set partitioning in hierarchical trees. *IEEE Trans. Circuits Syst. Video Technol.* **1996**, *6*, 243–250. [[CrossRef](#)]
18. Pearlman, W.A.; Islam, A.; Nagaraj, N.; Said, A. Efficient, low-complexity image coding with a set-partitioning embedded block coder. *IEEE Trans. Circuits Syst. Video Technol.* **2004**, *14*, 1219–1235. [[CrossRef](#)]
19. Aiazzi, B.; Alparone, L.; Baronti, S.; Latri, C. Crisp and Fuzzy Adaptive Spectral Predictions for Lossless and Near-Lossless Compression of Hyperspectral Imagery. *IEEE Geosci. Remote Sens. Lett.* **2007**, *4*, 532–536. [[CrossRef](#)]
20. Mielikainen, J.; Huang, B. Lossless Compression of Hyperspectral Images Using Clustered Linear Prediction with Adaptive Prediction Length. *IEEE Geosci. Remote Sens. Lett.* **2012**, *9*, 1118–1121. [[CrossRef](#)]
21. Wu, J.; Kong, W.; Mielikainen, J.; Huang, B. Lossless Compression of Hyperspectral Imagery via Clustered Differential Pulse Code Modulation with Removal of Local Spectral Outliers. *IEEE Signal Process. Lett.* **2015**, *22*, 2194–2198. [[CrossRef](#)]
22. Klimesh, M. *Low-Complexity Lossless Compression of Hyperspectral Imagery via Adaptive Filtering*; The Interplanetary Network Progress Report; NASA Jet Propulsion Laboratory (JPL): Pasadena, CA, USA, 2005; pp. 1–10.
23. Lossless Multispectral & Hyperspectral Image Compression CCSDS 123.0-B-1, Blue Book, May 2012. 2015. Available online: <https://public.ccsds.org/Pubs/123x0b1ec1.pdf> (accessed on 10 August 2018).
24. Mamun, M.A.; Jia, X.; Ryan, M. Sequential multispectral images compression for efficient lossless data transmission. In Proceedings of the 2010 Second IITA International Conference on Geoscience and Remote Sensing, Qingdao, China, 28–31 August 2010; Volume 2, pp. 615–618.
25. Muñoz-Gomez, J.; Bartrina-Rapesta, J.; Blanes, I.; Jimenez-Rodriguez, L.; Aulí-Llinàs, F.; Serra-Sagrístà, J. 4D remote sensing image coding with JPEG2000. *Proc. SPIE* **2010**, *7810*, 1–9.
26. Ricci, M.; Magli, E. On-board lossless compression of solar corona images. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31 July 2015; pp. 2091–2094.
27. Mamun, M.; Jia, X.; Ryan, M.J. Nonlinear Elastic Model for Flexible Prediction of Remotely Sensed Multitemporal Images. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 1005–1009. [[CrossRef](#)]
28. Foster, D.H.; Amano, K.; Nascimento, S.M. Time-lapse ratios of cone excitations in natural scenes. *Vision Res.* **2016**, *120*, 45–60. [[CrossRef](#)] [[PubMed](#)]
29. SOC700 Series Hyperspectral Imaging Systems. 2018. Available online: <https://surfaceoptics.com/products/hyperspectral-imaging/soc710-portable-hyperspectral-camera/> (accessed on 10 August 2018).
30. Time-Lapse Hyperspectral Radiance Images of Natural Scenes 2015. 2015. Available online: http://personalpages.manchester.ac.uk/staff/david.foster/Time-Lapse_HSIs/Time-Lapse_HSIs_2015.html (accessed on 1 March 2015).
31. Magli, E. Multiband Lossless Compression of Hyperspectral Images. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 1168–1178. [[CrossRef](#)]
32. Shen, H.; Pan, W.D.; Wang, Y. A Novel Method for Lossless Compression of Arbitrarily Shaped Regions of Interest in Hyperspectral Imagery. In Proceedings of the 2015 IEEE SoutheastCon, Fort Lauderdale, FL, USA, 9–12 April 2015; pp. 1–6.
33. Liu, W.; Pokharel, P.P.; Principe, J.C. Correntropy: Properties and Applications in Non-Gaussian Signal Processing. *IEEE Trans. Signal Process.* **2007**, *55*, 5286–5298. [[CrossRef](#)]

34. Golomb, S. Run-length encodings (Corresp.). *IEEE Trans. Inf. Theory* **1966**, *12*, 399–401. [[CrossRef](#)]
35. Shen, H.; Pan, W.D.; Dong, Y.; Jiang, Z. Golomb-Rice Coding Parameter Learning Using Deep Belief Network for Hyperspectral Image Compression. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017; pp. 2239–2242.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).