Efficient Lossless Compression of 4D Hyperspectral Image Data

Hongda Shen, David Pan and Yuhang Dong

Department of Electrical and Computer Engineering

University of Alabama in Huntsville

THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

Outline

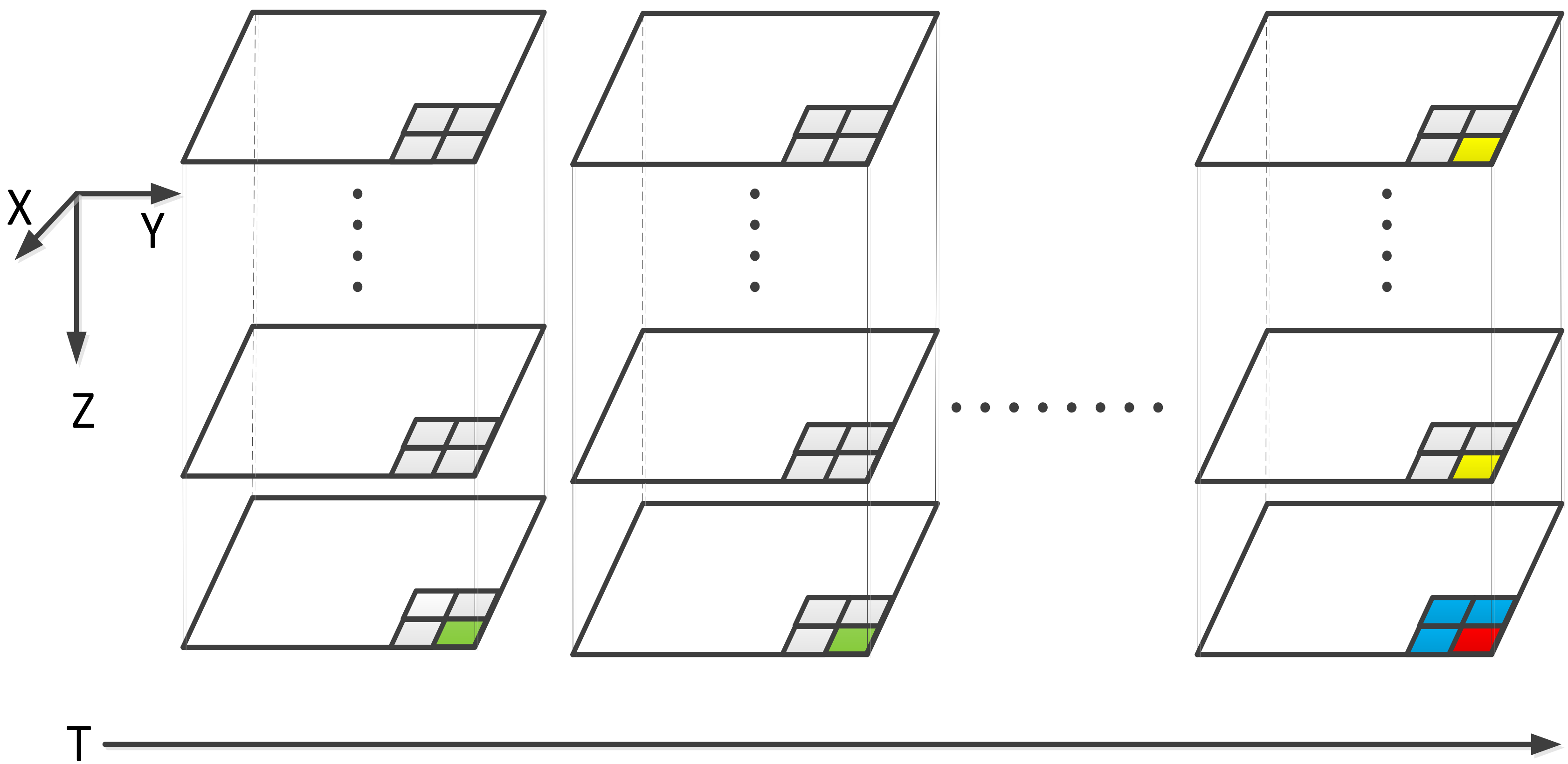
- 4D Hyperspectral Data.
- Why compression?
- Related work.
- Problem Analysis.
- The Algorithm.
- Experimental Results.
- Conclusions and Future work.

4D Hyperspectral Data

- Remote sensing (high resolution = large size).
- Hyperspectral Imaging.
  - 3D Data cube (2D Spatial + 1D Spectral)
  - 12 bits or 16 bits/pixel.
- Time-lapse Hyperspectral imagery
  - A sequence of 3D HSIs captured over the same scene but at different time stamps (often at a fixed time interval).
  - 3D Dimensions + 1D Time (4D)
  - For example, 454.78 MB (for 7 frames) to 584.71 MB (for 9 frames). See Table.

THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

4D Hyperspectral Data (Cont.)



THE UNIVERSITY OF
ALABAMA IN HUNTSVILLE

Samples

4D Hyperspectral Data (Cont.)

- Problems:
  - More stacks will be captured by the HSI sensor with the time.
  - Extreme case: in 4D HSI data streaming, the captured data volume accumulates very fast.
  - Large data volume can:
    - Slow down the data transmission within the limited bandwidth condition.
    - Requires more storage space which could be very expensive in many remote sensing applications.

THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

Why compression?

- Limited bandwidth and storage space on hyperspectral imaging sensor.
- Data compression techniques provide a good solution to these problems.
- High fidelity of captured images in the accuracy demanding applications requires lossless compression over lossy compression.

Related work

- Predictive methods:
  - 3D CALIC [1], M-CALIC [2], LUT [3], SLSQ [4] and CCAP [5].
  - "Fast Lossless (FL)" method [6]:
    - NASA Jet Propulsion Lab.
    - Achieve good compression efficiency at very low complexity.
    - CCSDS standard [9] for multi- and hyper-spectral imagery.
- Transform based methods:
  - SPIHT [7] and SPECK [8].

THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

More related work

- For hyperspectral images compression, only spatial and spectral correlations have been exploited so far.
- [10] proposed a 4D lossless compression algorithm, albeit lacking details on the prediction algorithms used for prediction.
- Karhunen-Loeve Transform (KLT), Discrete Wavelet Transform (DWT) and JPEG 2000 has been applied to reduce the spectral and temporal redundancy of 4D remote sensing image data [11].

Our contributions

- An information-theoretic analysis on the amount of compression achievable on 4D HSI based on conditional entropy, by taking into account spectral and temporal correlations.
- A predictor based on low complexity Correntropy-based Least Mean Square (CLMS) learning algorithm was proposed to better model the data.

THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

Problem Analysis

- Information theoretic analysis
- Let ${X_j}^t$ be a 4D hyperspectral image source at the t^{th} time instant and j^{th}spectral band producing K different pixel values v_i $(i = 1,2 \ldots K)$. The entropy of this source can be computed based on the probabilities $p(v_i)$ by:

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

- For 4D hyperspectral images, there exists strong spectral and temporal correlations.
- Therefore, these correlations can be exploited to reduce the $H\left({X_j}^t\right)$.

THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

Problem Analysis

- The conditional entropy of this time-lapse hyperspectral image source can be computed as follows:

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

  Where ${C_j}^t$ denoted as *Context*, a group of correlated pixels.

- As long as there is any correlation between the context ${C_j}^t$ and the current pixel, $H\left({X_j}^t|{C_j}^t\right) < H\left({X_j}^t\right)$ always holds, in other words, fewer bits are required after compression.

The Algorithm

- Linear prediction based lossless compression method uses a linear combination of those encoded pixels (causal context pixels) adjacent to the current pixel as its estimate.
- For 4D time-lapse HSI lossless compression, a linear prediction can be generalized as follows:

$$\hat{x}_{m,n}{}^{t,j} = \boldsymbol{w}_{t,j}{}^{T} \boldsymbol{y}_{m,n}{}^{t,j}.$$

where $\hat{x}_{m,n}{}^{t,j}$ represents an estimate of a pixel, $x_{m,n}{}^{t,j}$ at spatial location (m, n), j^{th} band and t^{th} time frame while $\boldsymbol{y}_{m,n}{}^{t,j}$ and $\boldsymbol{w}_{t,j}{}^{T}$ represent its causal context pixels and linear weights respectively.

Correntropy

- FL compression method for hyperspectral images is built on Least Mean Square (LMS), which is only optimal in gaussian situation.
- However, prediction residuals are more likely to follow a Laplcacian or Geometric distribution.
- So the performance of the conventional LMS predictor, for example, the FL method may degrade in presence of non-Gaussian signals, especially in those very structured regions of one image.
- To improve the robustness of the predictor, we develop a predictor based on the Maximum Correntropy Criterion (MCC), denoted as CLMS predictor.

THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

Correntropy

- A brief introduction of Correntropy [12]:
- Correntropy was developed as a local similarity measure between two random variables X and Y in
$$V_\sigma(X, Y) = E[K_\sigma(X - Y)].$$
- where K_σ is a positive definite kernel with kernel width controlled by the parameter σ, and the expectation $E[]$ is practically computed using sample arithmetic average.
- Correntropy can be viewed as a generalized correlation function containing even higher order moments of the error signal $X - Y$.

THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

Correntropy based LMS

- Assume we have a pair of random variables with a finite number of samples $\{d_i, y_i\}_{i=1}{}^{N}$ where N is the number of samples in each random variable.
- Furthermore, the estimate y_i can be computed as $y_i = \boldsymbol{w}_i{}^T\boldsymbol{X}_i$, a linear weighted average of input vector $\boldsymbol{X}_i$.
- Replace Mean Square Error (MSE) in LMS with Correntropy leads to a new adaptive learning algorithm: CLMS. The weight update function is shown as follows:

$$\boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \frac{\mu}{\sqrt{2\pi}\sigma^3}\exp\left(\frac{-e_n{}^2}{2\sigma^2}\right)e_n\boldsymbol{X}_{n.}$$

where w_n is the weight vector at n^{th} time instant and $e_i = d_i - \boldsymbol{w}_i{}^T\boldsymbol{X}_i$.

- In fact, this Correntropy-induced updating function can be viewed as LMS with a self-adjusting learning rate.

THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

CLMS based Predictor

- In Fig. 1, suppose the red pixel is the one we are predicting and the arithmetic average of the three blue pixels from its spatial causal neighborhood is computed and subtracted from the red pixel value.
- Denote N_s and N_t as the number of pixels from previous spectral bands at the current time instant (yellow pixels) and the number of pixels from the same spectral bands from previous time frames (green pixels), respectively.
- The causal context is constructed as
  $$C_j{}^t = \left[x^{t,j-1}, x^{t,j-2}, \dots, x^{t,j-N_s}, x^{t-1,j}, x^{t-2,j}, \dots, x^{t-N_t,j}\right].$$

THE UNIVERSITY OF
ALABAMA IN HUNTSVILLE

CLMS based Predictor

Algorithm 1 CLMS Predictor

Initialize:
1) T (# of time frames)
2) B (# of spectral bands for each time frame)
3) $\mu = 0.3$ and $\sigma = 50$
4) Local mean subtracted data X

```
for t = 1:T do
    for b = 1:B do
        initialize: w = 0
        for each pixel in this band do
            Output residual x − x̂ using Eq. (4).
            Updating w using Eq. (7).
        end for
    end for
end for
```

THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

Entropy Coding

- Golomb-Rice Coding is favored in this work because of its simplicity and low complexity.
- Arithmetic Coding is also optional for slightly better compression ratio.

THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

Experiments

- We conducted our experiment on three 4D time-lapse HSI test datasets, *Levada*, *Gualtar* and *Nogueiro*.
- While the size of a single dataset we tested is not very large, ranging from 454.78 MB (for 7 frames) to 584.71 MB (for 9 frames), the data can easily grow to a huge size with increased number of time frames and higher spatial and spectral resolutions.

Experiment Settings

- Learning rate $\mu = 0.3$ and kernel width $\sigma = 50$.

Table 1: Datasets Used.

Dataset	Size	# of time frames	Precision(bits)
Levada	$1024 \times 1344 \times 33$	7	12
Gualtar	$1024 \times 1344 \times 33$	9	12
Noguerio	$1024 \times 1344 \times 33$	9	12

THE UNIVERSITY OF
ALABAMA IN HUNTSVILLE

Results

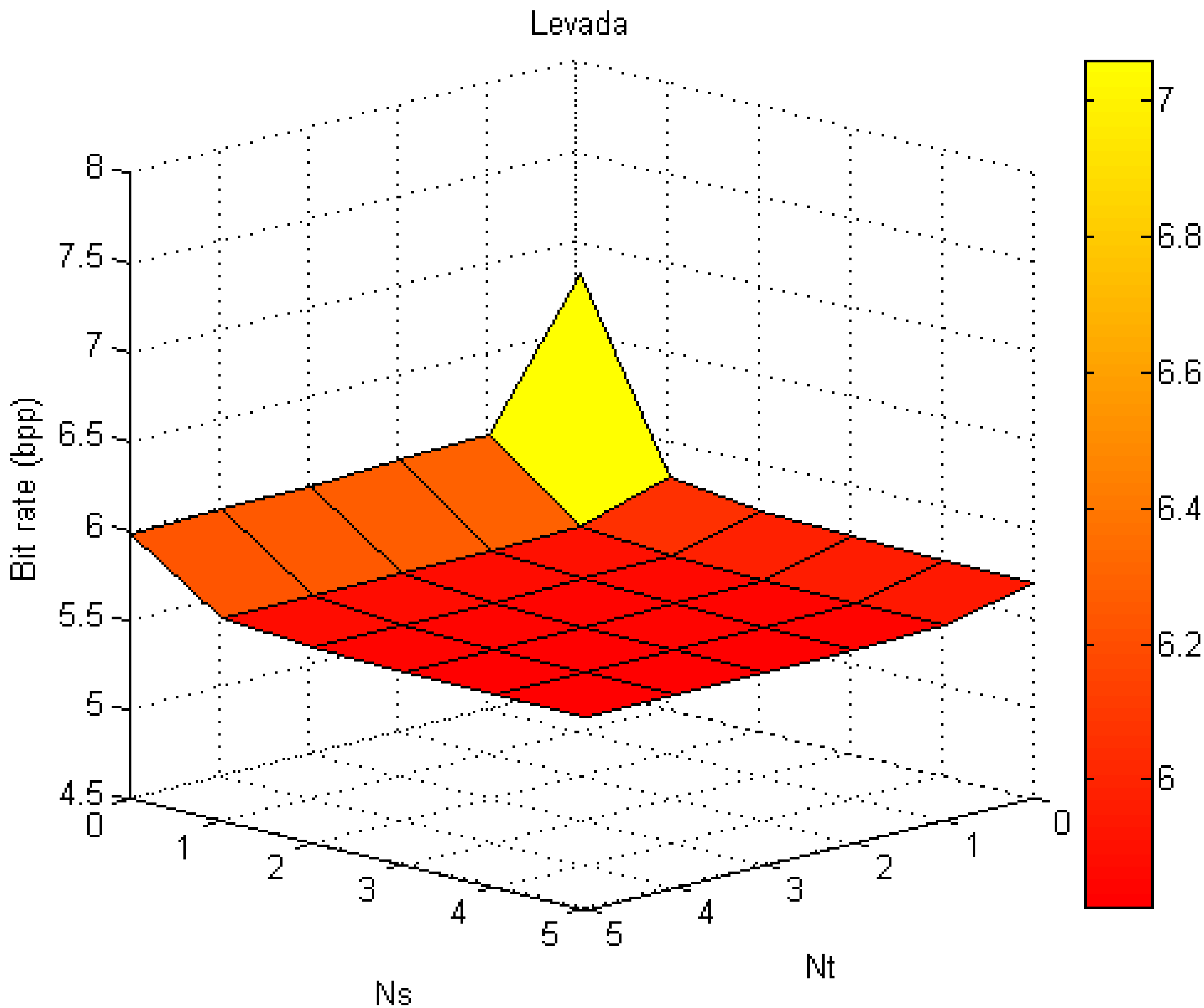
- We applied our algorithm using different combinations of N_s and N_t causal pixels from spectral and temporal bands.

Table 2: Bit rates (bits/pixels) on "Levada".

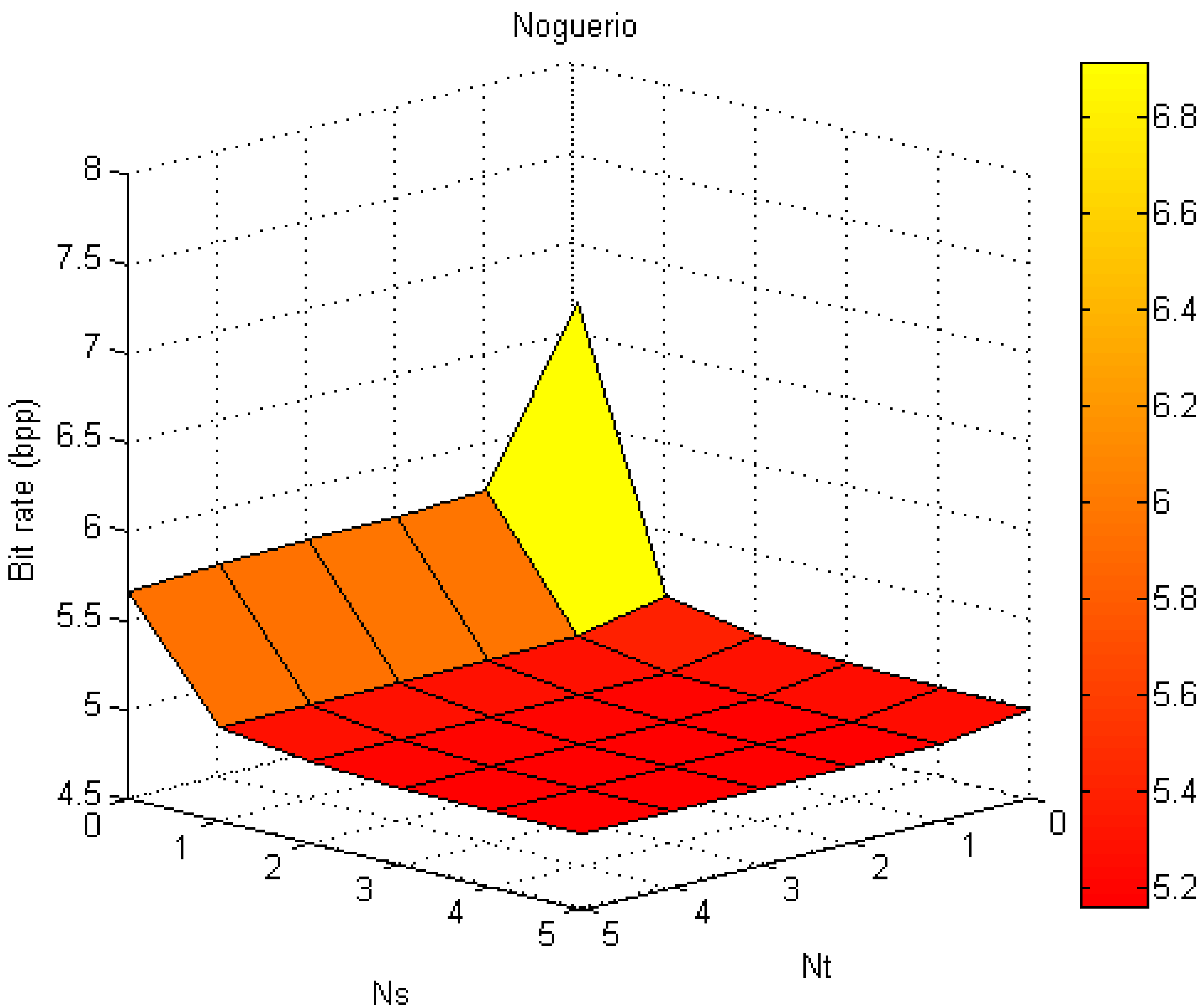
N_s	$N_t = 0$	$N_t = 1$	$N_t = 2$	$N_t = 3$	$N_t = 4$	$N_t = 5$
0	7.0604	6.2858	6.2686	6.2431	6.2400	6.2382
1	6.0476	5.8985	5.8898	5.8813	5.8795	5.8787
2	5.9807	5.8570	5.8497	5.8414	5.8395	5.8388
3	5.9592	5.8433	5.8360	5.8279	5.8260	5.8252
4	5.9487	5.8359	5.8289	5.8211	5.8191	5.8184
5	5.9410	5.8308	5.8239	5.8162	5.8146	5.8138

THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

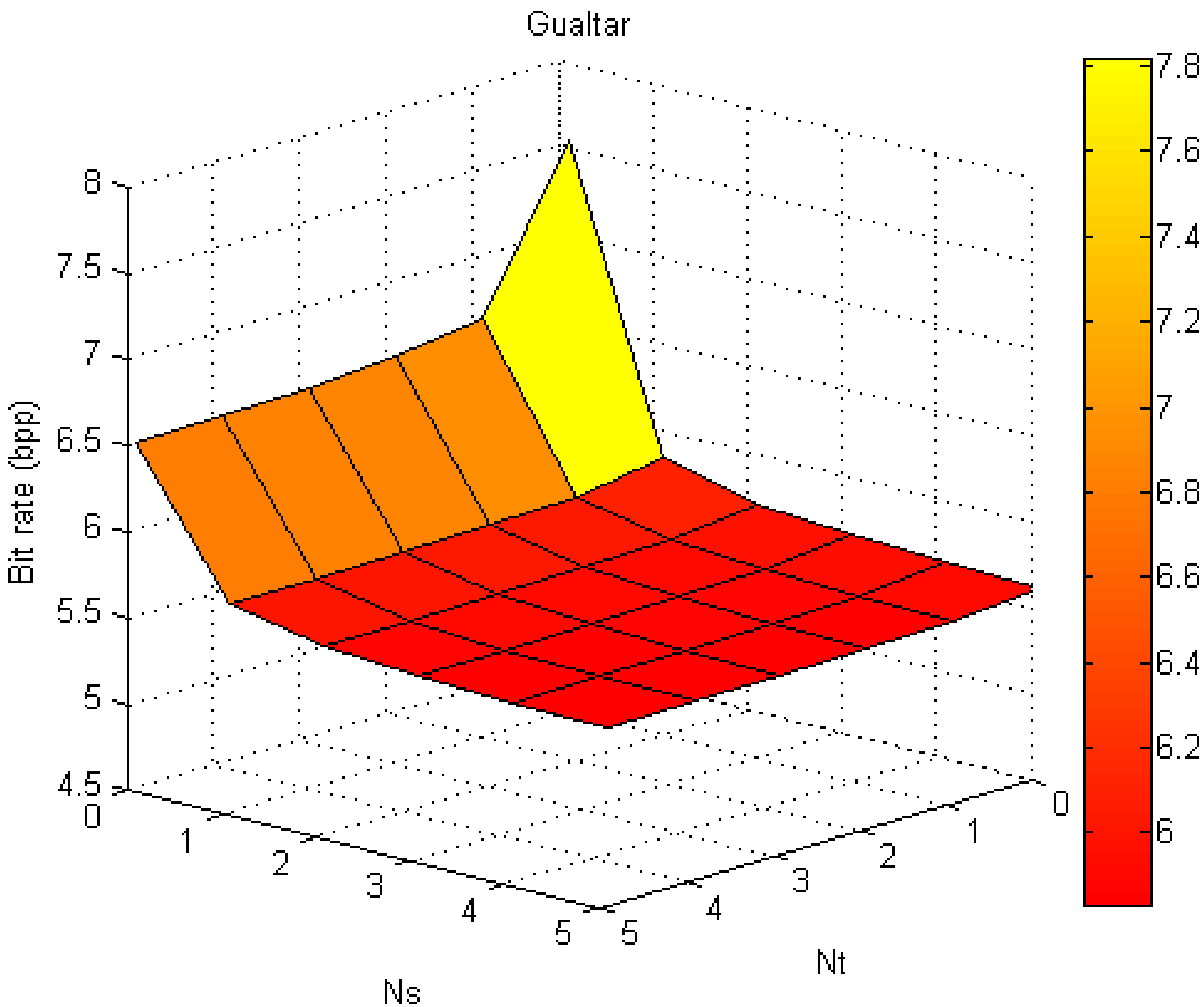
Case study: surface plot



Case study: surface plot

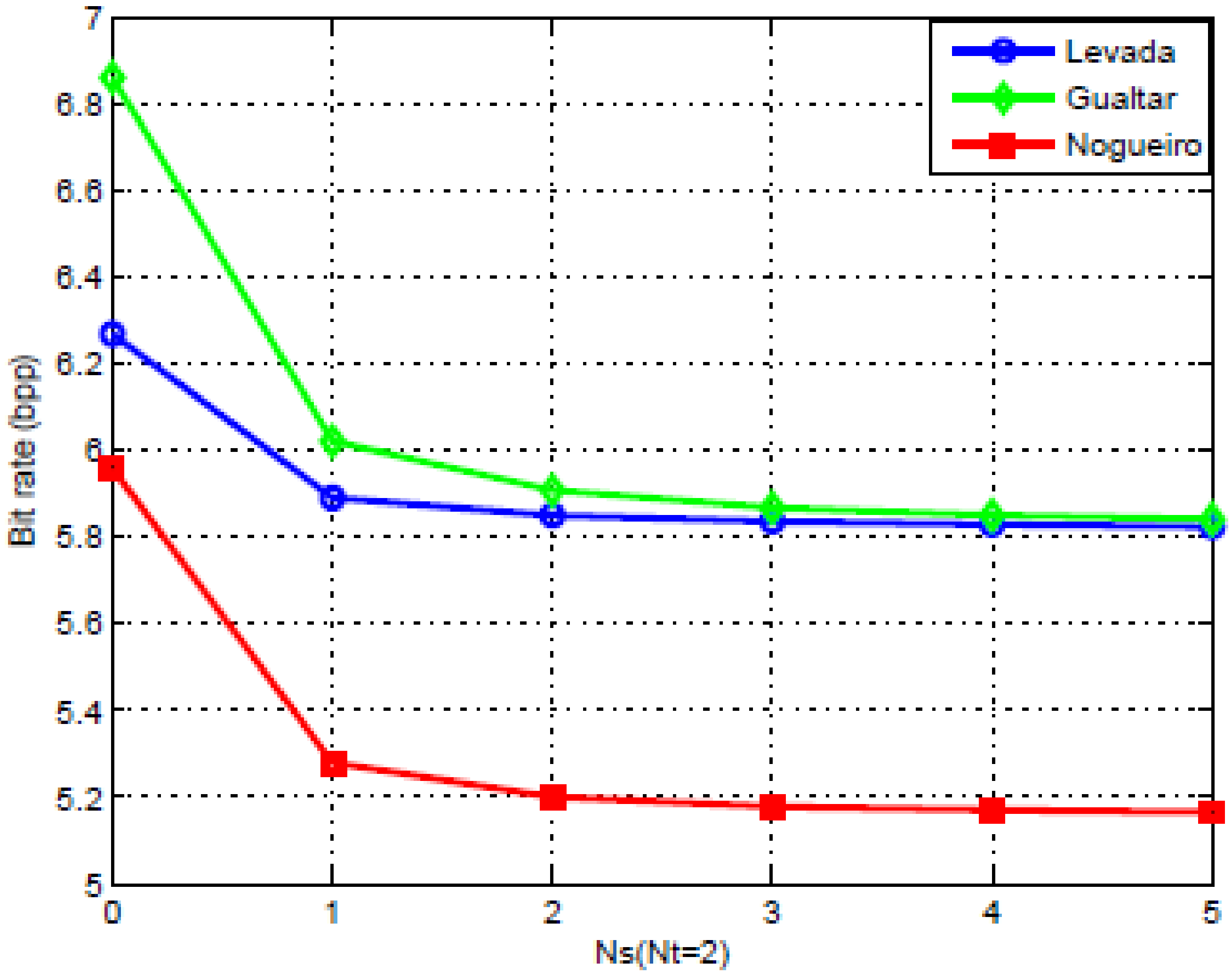


Case study: surface plot



THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

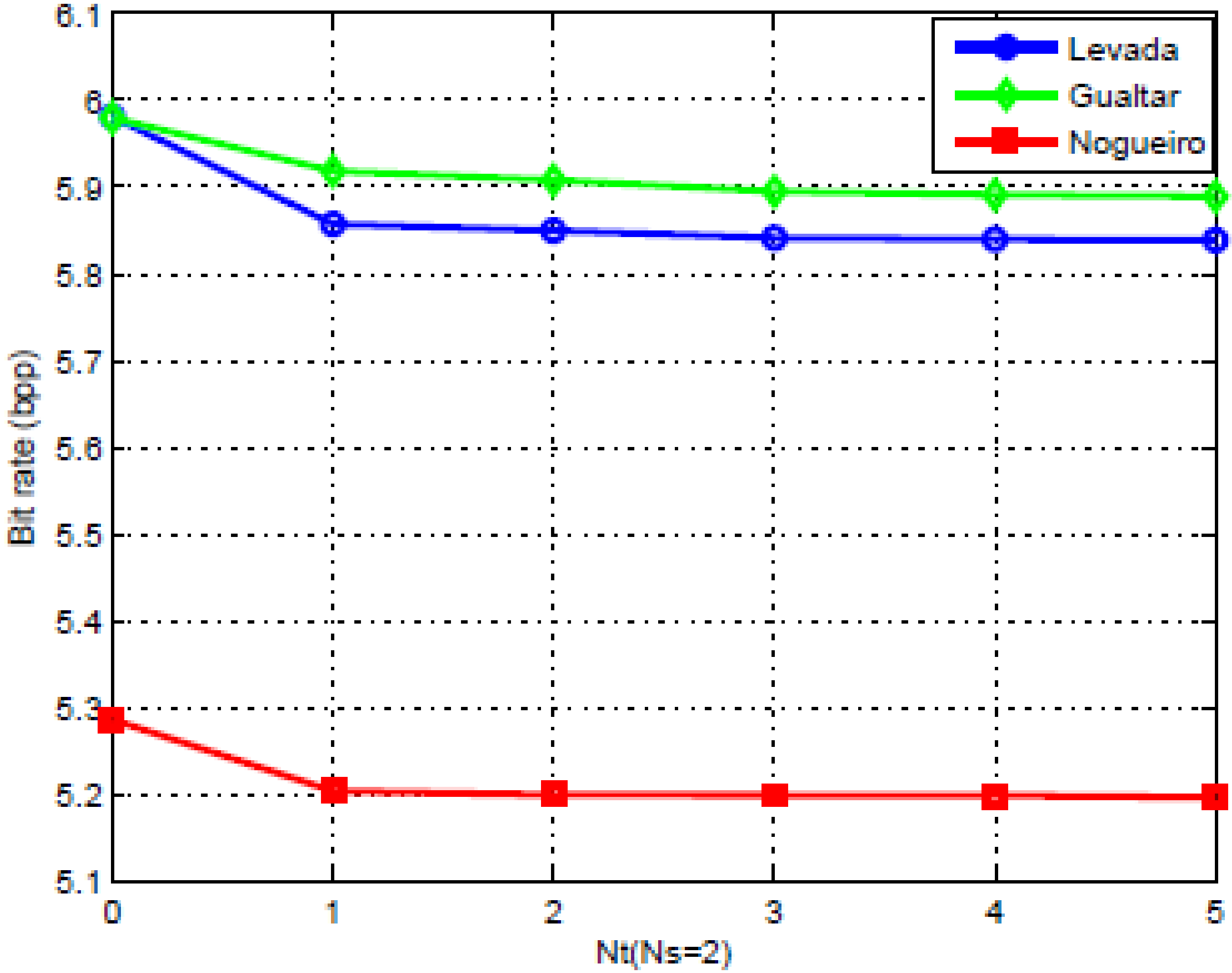
Case study: bit rate change



(a) Bit rate vs. N_s.

THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

Case study: bit rate change



(b) Bit rate vs. N_t.

Analysis

- The bit rate has been reduced by approximately 1.2 bits by just adding one previous spectral band and the same spectral band from the previous time stamp in the context.
- Furthermore, if we fix either $N_s$or $N_t$and increase only $N_s$or $N_t$accordingly, the compression bit rate will drop as well.
- However, this performance improvement gradually becomes marginal as $N_s$or N_t increases.
- Overall, it is possible to increase $N_s$or N_t to achieve higher compression ratio.
- On the other hand, prediction using only one previous spectral band and/or the same spectral band but from last time instant will also yield good compression performance at a very low computational cost.

THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

Conclusion

- Experimental results have demonstrated the outstanding capability of this proposed algorithm to compress 4D time-lapse HSI data through spectral and temporal decorrelation.
- Second, an information theoretic analysis based on conditional entropy has been made to provide a framework to guide and evaluate the actual compression.
- Increasing the number of previous bands involved in the prediction will absolutely yield better compression performance as long as they are correlated statistically with the current HSI band.

Future work

- We will investigate how to fully utilize this proposed algorithm and analytic framework to handle HSI data streaming, which is more challenging but also in better need for compression.
- Additionally, ROI lossless compression of HSI has begun to gain attention from researchers. Recently, some work has been done to handle ROIs in HSI data.

Reference

[1] 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.

[2] X. Wu and N. Memon, "Context-based, adaptive, lossless image coding," IEEE Trans. Commun, vol. 45, no. 4, pp. 437–444, Apr 1997.

[3] J. Mielikainen, "Lossless compression of hyperspectral images using lookup tables," IEEE Signal Process. Lett., vol. 13, no. 3, pp. 157–160, March 2006.

[4] 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.

[5] 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.

[6] 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.

Reference

[7] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," IEEE Trans. Circuits Syst. Video Technol., vol. 6, no. 3, pp. 243–250, Jun 1996.

[8] W. A. Pearlman, A. Islam, N. Nagaraj, and A. Said, "Efficient, lowcomplexity image coding with a set-partitioning embedded block coder," IEEE Trans. Circuits Syst. Video Technol., vol. 14, no. 11, pp. 1219–1235, Nov 2004.

[9] "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).

[10] M. A. Mamun, X. Jia, and M. Ryan, "Sequential multispectral images compression for efficient lossless data transmission," in 2010 Second IITA Intl. Conf. on Geosci. Remote Sens., vol. 2, Aug 2010, pp. 615–618.

[11] J. Munoz-Gomez, J. Bartrina-Rapesta, I. Blanes, L. Jimenez-Rodriguez, F. Auli-Llinas, and J. Serra-Sagrista, "4D remote sensing image coding with JPEG2000," Proc. SPIE, vol. 7810, pp. 1–9, 2010.

[12] A. Singh and J. C. Principe, "Using correntropy as a cost function in linear adaptive filters," in Intl. Joint Conf. on Neural Netw., June 2009, pp. 2950–2955.