GOLOMB-RICE CODING PARAMETER LEARNING USING DEEP BELIEF NETWORK FOR HYPERSPECTRAL IMAGE COMPRESSION

Hongda Shen, W. David Pan, Yuhang Dong and Zhuocheng Jiang

Dept. of Electrical and Computer Engineering
University of Alabama in Huntsville
Huntsville, AL 35899, USA

ABSTRACT

While Golomb-Rice codes are optimal for geometrically distributed source, the practically achievable coding efficiency depends on the accuracy of the coding parameter estimated from the input data. Most existing methods are based on the assumption of geometric distribution and thus would suffer from a loss in coding efficiency if the underlying distribution deviates from the geometric distribution, which is usually the case in practice. We proposed a data-driven parameter estimation method without assuming the underlying distribution. We formulated the problem of choosing the best coding parameter for the given input data as a pattern classification problem. To this end, we trained a deep belief network using the data segments to be coded, along with their "labels", which are the optimal coding parameters that yield the shortest codewords. Simulations on data synthesized using statistical models, as well as data in hyperspectral image coding showed that the proposed deep learning method tended to be more robust than several state-of-the-art parameter estimation methods, with the capability to further improve the accuracies of these methods.

Index Terms— Golomb-Rice code, hyperspectral image compression, machine learning, deep belief network.

1. INTRODUCTION TO CODING PARAMETER ESTIMATION

Golomb-Rice (GR) codes [1, 2] are often used to compress prediction residuals found in many data compression applications. Particularly, GR codes have been used in the "Fast Lossless" (FL) method, the Consultative Committee for Space Data Systems (CCSDS) new standard for Multispectral and Hyperspectral Data Compression [3]. Parameterized with a single integer, Golomb-Rice codes are easy to implement, and are optimal for sources with geometric distribution. However, coding parameter has to be estimated from the input data, which might deviate from the assumed geometric distribution. And the efficiency of the Golomb-Rice codes depends directly on the estimation accuracy [4].

Specifically, a geometrically distributed source X is defined as follows:

$$P(X = k) = (1 - p)p^k, \tag{1}$$

where $p \in (0, 1)$ and k is a non-negative integer. Given p, the minimal expected code length is achieved by selecting a coding parameter m as follows [5]:

$$m = \log_2 \left(\left\lceil -\frac{\log(1 + p)}{\log p} \right\rceil \right), \tag{2}$$

which plays a central role in the actual coding process as described in [1]. In practice, we often estimate m from the input data. To our best knowledge, so far, there have been only three existing methods on this problem. These three state-of-the-art methods are presented as follows. [6] used the following formula to estimate m from the prediction residuals for hyperspectral data compression.

$$m = max \left(0, \left\lceil \log_2 \left(-\frac{\mu}{2} \right) \right\rceil \right), \tag{3}$$

where μ is the arithmetic average of the data to be coded. In [7] and [8], two similar formulas involving the arithmetic average μ have been proposed, respectively:

$$m = max \left(0, 1 + \left\lfloor \log_2 \left(\frac{\log(\phi - 1)}{\log(\frac{\mu}{\mu+1})} \right) \right\rfloor \right), \tag{4}$$

where ϕ denotes the golden ratio $(\sqrt{5} + 1)/2$, and

$$m = max \left(0, \left\lceil \log_2(\mu) - 0.05 + \frac{0.6}{\mu} \right\rceil \right). \tag{5}$$

All these parameter estimation methods were derived or optimized based on the assumption that the underlying distribution is geometric distribution. Nevertheless, the actual data could deviate from a perfect geometric distribution. For example, in the work on predictive hyperspectral image compression [9], we sought to code the prediction residuals using Golomb-Rice code. Deviations from the geometric distribution can be seen from some representative data histograms (as an approximation of the actual distribution) shown in Fig. 1. The "tails" in the histogram (to the left) makes the distribution more resemble a mixture of two distinct geometric distributions, due to the existence of edges, contours or corners in the original images. Histogram to the right of Fig. 1 shows more deviations from the standard geometric distribution, due to the presence of impulsive noise in the hyperspectral images. On the other hand, to avoid long processing delays, data to be coded are typically partitioned into short segments [3], making accurate parameter estimation difficult due to insufficient statistics. To remedy this problem, modified Golomb-Rice codes have been proposed by using more complex data source modeling [5, 10], which are beyond the scope of this work, which aims at improving the accuracy of parameter estimation for Golomb-Rice codes. In this paper, we proposed

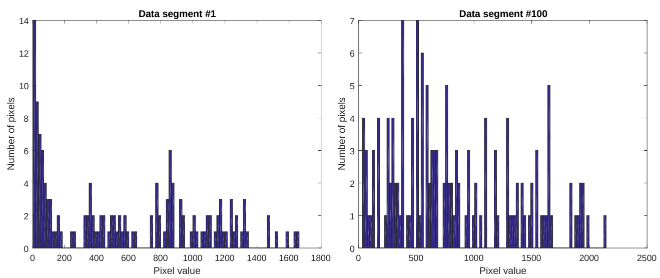

Fig. 1: Histograms of two data segments from a sample dataset, "Indian Pines".

a data-driven parameter estimation method without assuming the underlying distribution. The novelty lies in our formulation of the problem of choosing the best coding parameter for the given input data as a pattern classification problem. Traditional machine learning algorithms require pre-extracted features prior to the actual classifier. Hence, it is impossible to apply traditional machine learning methods to this problem. However, motivated by the success of deep machine learning methods in solving many classification problems without feature extraction, we considered one specific deep learning method known as the Deep Belief Network (DBN), which has been applied to classification in lower dimensions [11] and data-driven classification [12, 13] for image, speech and other data types. The rest of this paper is organized as follows. Section 2 presents the proposed method of learning the optimal Golomb-Rice coding parameters using deep belief network. Simulation results are then given in Section 3. Section 4 concludes the paper.

2. THE PROPOSED DEEP LEARNING METHOD

We first explain why the problem of choosing the best coding parameter can be reformulated as a supervised pattern classification problem. Then we describe how we use the deep machine learning method to solve the classification problem. In most practical applications, there are only a finite number of parameter values to choose from for Golomb Rice codes. For example, coding a typical image with 8 bits/pixel would require the coding parameter m to be chosen from one of only 9 possible integers in the set of $[0, 8]$. Similarly, the an image with 16 bits/pixel would require only a slight larger set $[0, 16]$ of 17 integers. Therefore, we can train a classifier, where the input is a data segment to be Golomb-Rice coded, and its "label" is the m value in the set, such that Golomb-Rice coding the data segment will give the shortest codeword length, among all the possible m values in the set of admissible values. In the testing phase, we feed the new data segment to be coded into the classifier, which will output the m value we will use for actual coding of this data segment. We expect the coding parameters thus chosen would yield good coding efficiency if the classifier has been reasonably well trained. Such a data-driven method does not require any knowledge about the underlying distribution of the input data, and thus would be generally more robust than methods that presume a certain distribution of the data. Furthermore, the classification performance can be readily measured by comparing the output of the classifier during testing with the ground truth – the actual m value that can best compress the test data segment using the Golomb-Rice code.

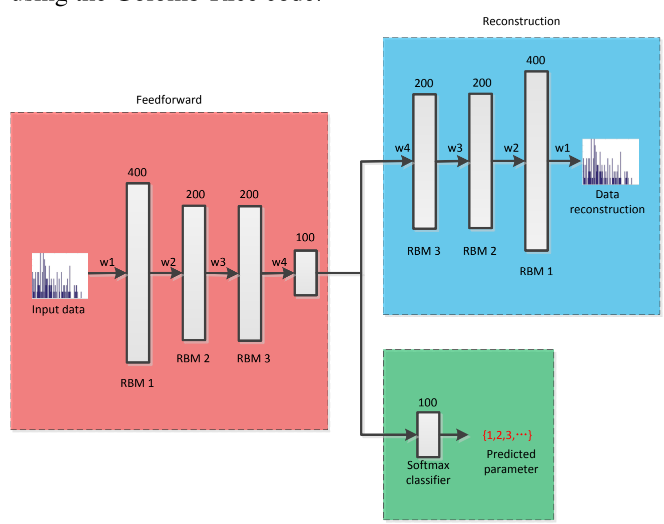

Fig. 2: The architecture of the deep belief network for Golomb-Rice coding parameter estimation. Note that the reconstruction layers were used only for pretraining of RBM's.

The deep belief network (DBN) is a generative graphical model composed of multiple stacked Restricted Boltzmann Machines (RBMs). Using DBN to model image data has been extensively studied in [12]. Motivated by DBN's remarkable data modeling capability, we designed a data-driven method based on DBN to estimate the parameter value for the Golomb-Rice codes. Although DBN is a unsupervised learning method, the automatically learned features can be utilized for supervised learning to achieve pattern classification. Specifically, on top of the stacked RBM, a discriminative softmax layer was added to enable classification using the learned features of the data. The DBN was pretrained and unrolled in a greedy layer-wise manner [14]. Then back propagation was carried out to fine-tune the entire DBN based on the negative log-likelihood of the labels. Since Gaussian-Bernoulli RBM (GB-RBM) is mainly designed for real-valued data [11], we selected the Bernoulli RBM (B-RBM) for nonnegative discrete data. All the data samples were first mapped to the interval $[0, 1]$ and then sent to the DBN. A four-layer DBN was constructed to learn from the data. There are 400, 200, 200 and 100 neurons within each of the four layers of DBN. The top layer is a softmax layer, which outputs the class label (the m value), as shown in Fig. 2. Note that while this 400-200-200-100 architecture was found to provide accurate parameter estimation, it is not unique and other architectures might provide better results. All the data segments to be coded came with the ground truth (the optimal coding parameter value that gives the shortest codeword). In the training phase, the DBN can learn the underlying distribution of the data and adjust the weights within each layer of the network. Once the training is completed, new data segments can be directly forward fed through the network, thereby yielding the estimated parameter value at the output of the network.

3. SIMULATION RESULTS

In order to study the efficiency of the proposed method, we conducted two simulations. The first simulation used synthesized data generated by using a certain distribution that deviates from the standard geometric distribution. The second simulation used actual residual data obtained by applying the “Fast Lossless (FL)” predictor on the well-known hyperspectral image dataset, “Indian Pines” [6]. Comparisons were made against three existing parameter estimation methods, namely, Method 1, Method 2, and Method 3, corresponding to Eq. (3), Eq. (4), and Eq. (5), respectively.

As shown in Fig. 1, data segments to be coded might not follow exactly the geometric distribution. Therefore, we synthesized data using distributions that deviate from the standard geometric distributions. Motivated by the work [4, 9], we generated the synthetic dataset following a mixture of two distinct geometric distributions. We first generated 5,000 data segments, with each segment containing 100 samples, following the standard geometric distribution with $p = 0.8$. Then, we added more data segments generated with another geometric distribution ($p = 0.999$). These additional data segments can be treated as “outliers” (in a loose sense) and they range from 1% to 10% of the total number of data segments, indicating different degrees of mixtures. Note all these values were chosen to simulate typical scenarios encountered in lossless coding of hyperspectral data in our case study.

Fig. 3 shows that the proposed method has the lowest misclassification rates for most of the cases for the synthesized dataset except for the case where the outlier percentage is 9%. Overall, the proposed method tends to be most robust in achieving high estimation accuracy.

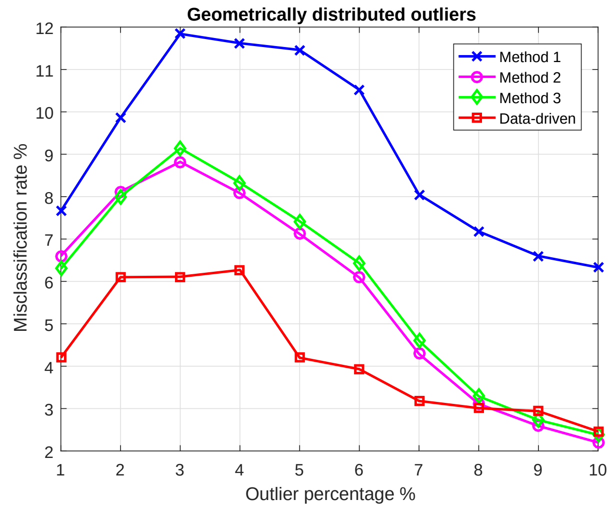

Fig. 3: Simulation results on synthetic data.

The so-called Fast Lossless (FL) predictor has been proposed to de-correlate the hyperspectral image in the spatial and spectral directions due to its low-complexity [6]. After the prediction, residual data segment within a spectral band is encoded using the Golomb-Rice code. For the hyperspectral dataset “Indian Pines” [15] used in our case study, we treated each row of a spectral band image as one data segment (containing 145 data samples), based on which the coding parameter was estimated. Since the dataset has 200 spectral bands, corresponding to a total of $29,000 = 145 \times 200$ data segments, a reasonably large number of data would be available for both training and testing of the deep belief network. We randomly selected $14,500$ data segments for training, and the remaining $14,500$ segments for testing. Since each pixel has 16 bits, there will be a total of 17 possible classes, $\{0, 1, \cdots, 16\}$. However, only eight classes ($\{2, 3, \cdots, 9\}$) showed up in the ground truth, whereas the remaining coding parameters were never chosen due to their inferior coding efficiency (less compression) than the parameters falling in the set of eight winning classes.

Fig. 4 shows the proposed method achieved very high accuracy, ranging from 96.57% (for Class 7) to 98.88% (for Class 9), indicating that the distribution of each data segments was learned very well by the deep belief network.

Table 1: False estimation rate for all four methods.

Method 1	Method 2	Method 3	**Proposed**
6.21%	4.28%	4.33%	2.01%

Table 1 shows the proposed method has the lowest false estimation rates than the other three methods. This means that the proposed deep learning method can further improve the accuracy of the existing parameter estimation methods (already with higher than 90% accuracy) on the real data.

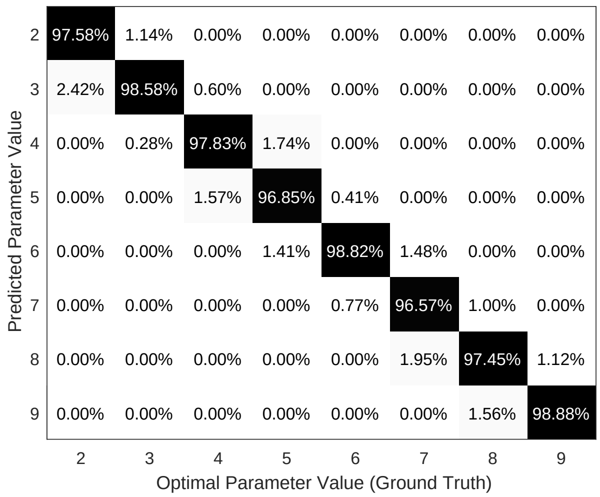

Fig. 4: Classification accuracy confusion matrix.

4. CONCLUSIONS

We proposed a data-driven parameter estimation method for Golomb-Rice coding by learning from the data using a deep belief network. To the best of our knowledge, this might be the first time the Golomb-Rice coding parameter estimation problem was formulated as a supervised learning problem. Simulations of the proposed method on both synthesized and real data demonstrated its advantages in terms of robustness and accuracy over several other parameter estimation methods that presumes the input data to be geometrically distributed. As the next step, we will study how varying the data segment size and the size of the training dataset will affect the estimation accuracy.

5. REFERENCES

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

[2] R. F. Rice, "Some practical universal noiseless coding techniques," in *Tech. Rep. JPL-79-22*, Jet Propulsion Laboratory, Pasadena, California, Mar. 1979.

[3] "Lossless multispectral & hyperspectral image compression CCSDS 123.0-B-1, ser.Blue Book, May 2012," `https://public.ccsds.org/Pubs/123x0b1ec1.pdf`, 2012 (accessed September 1, 2016).

[4] H. Shen, W. D. Pan, D. Wu, and M. Lubna, "Fast Golomb coding parameter estimation using partial data and its application in hyperspectral image compression," in *Proc. of IEEE SoutheastCon 2016*, March 2016, pp. 1–7.

[5] J. J. Ding, H. H. Chen, and W. Y. Wei, "Adaptive Golomb Code for joint geometrically distributed data and its application in image coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 4, pp. 661–670, April 2013.

[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.

[7] A. Kiely, "Selecting the Golomb parameter in Rice coding," in *The Interplanetary Network Progress Report*, Jet Propulsion Laboratory, Pasadena, California, Nov. 2004, pp. 1–18.

[8] A. Said, "On the determination of optimal parameterized prefix codes for adaptive entropy coding," `http://www.hpl.hp.com/techreports/2006/HPL-2006-74.pdf`, 2006 (accessed August 31, 2016).

[9] H. Shen, W. D. Pan, and D. Wu, "Predictive lossless compression of regions of interest in hyperspectral images with no-data regions," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 1, pp. 173–182, Jan 2017.

[10] A. Y. Reznik, "Low-complexity lossless codes for image and video coding," *Proc. SPIE*, vol. 7798, pp. 77980X–77980X–9, 2010.

[11] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[12] N. Wang, J. Melchior, and L. Wiskott, "Gaussian-binary Restricted Boltzmann Machines on modeling natural image statistics," *Computing Research Repository*, vol. abs/1401.5900, 2014.

[13] Y. J. Hu, Z. H. Ling, and L. R. Dai, "Deep belief network-based post-filtering for statistical parametric speech synthesis," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 5510–5514.

[14] Y. Bengio, P .Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Conference on Neural Information Processing Systems*, 2007.

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