IGARSS 2017
2017 IEEE International Geoscience and Remote Sensing Symposium
July 23–28, 2017
Fort Worth, Texas, USA
International Cooperation for Global Awareness
IEEE GRSS

Learning the Optimal Golomb-Rice Coding Parameters from Data using Deep Belief Network and Its Application in Hyperspectral Image Compression

Hongda Shen, ***W. David Pan***, Yuhang Dong and Zhuocheng Jiang
Department of Electrical and Computer Engineering
University of Alabama in Huntsville
Huntsville, Alabama

July 26, 2017

THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

In Memory of Solomon Golomb (1932 – 2016)

Quotes from the original paper of Golomb, a pioneer of communication technology (then at JPL) "Run-Length Encodings," IEEE Trans. on IT, 1966.

- Secret Agent 00111 is back at the Casino again, playing a game of chance, while the fate of mankind hangs in the balance.
- Each game consists of a sequence of favorable events (probability p), terminated by the first occurrence of an unfavorable event (probability $q = 1 - p$).
- The bartender, who is a free-lance agent, has a binary channel available, but he charges a stiff fee for each bit sent. The problem perplexing the Service is how to encode the vicissitudes of the wheel so as to place the least strain on the Royal Exchequer.

Golomb in White House to receive the National Medal of Science (2013)

Model of binary source for Run Length Coding:
1111111…1**0**11111**0**111111111**0**…

p = Prob[1], q = Prob[0] = $1 - p$

Outline

- Motivation
- Review of Golomb-Rice Codes
- Application in Compression of Hyperspectral Images
- Challenge: Coding Parameter Estimation
- Proposed Data-Driven Method
  - We call it “Golomb meets Big Data.”
- Deep Belief Network for Parameter Estimation
- Simulation Results
  - Synthesized Data
  - Actual Hyperspectral Image
- Conclusion

Motivation

ARSC | Alabama Remote Sensing Consortium

an industry-higher education consortium & a unique resource for the state of alabama.

- Hyperspectral data are “big”.
- Efficient compression on hyperspectral data needed to reduce delays associated with real-time transmission of the data, especially for “streaming” applications.
- We are in the Alabama Remote Sensing Consortium (ARSC) http://www.nsstc.uah.edu/arsc/
  - ARSC In agreement with Teledyne Brown Engineering (TBE) for the provision of hyperspectral data from the company’s Multi-User System for Earth Sensing (MUSES), an Earth-observation platform built for use on the International Space Station (ISS).
  - ARSC will use the data to pursue collaborative opportunities in education, research, and outreach with emphasis on remote sensing technology and applications.
- Focus on lossless compression of hyperspectral data.

Why Golomb Codes for Data Compression?

- The better known **Huffman code** is an optimal variable-length code; however,
  - at the encoder, code tree needs to be trained based on the source distribution;
  - The codeword dictionary have to be stored as side info., resulting in loss of compression efficiency.
  - The decoder is complex.
- **Golomb code** is also a variable-length code
  - Designed for compression of non-negative integers
  - <u>Optimal</u> for geometric source with a certain parameter

    Probability Mass Function (PMF): $G(n) = p^n q = p^n (1 - p)$,

  - No need to store codeword dictionary
  - Simple and fast decoding

Applications of Golomb Code

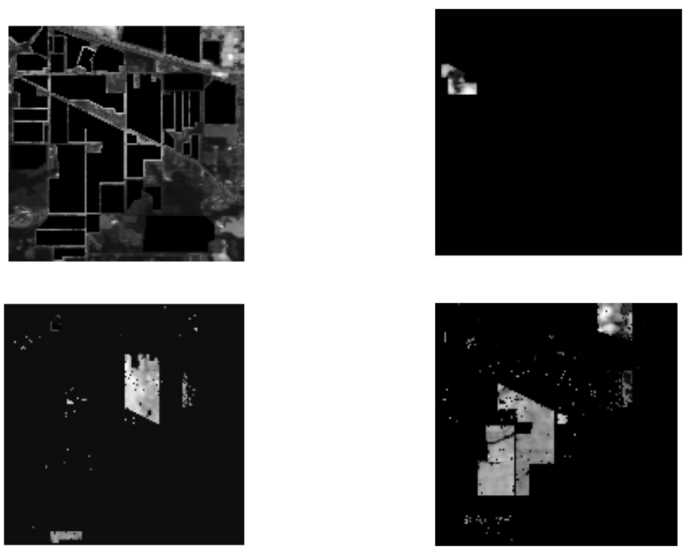
- NASA JPL Low-Complexity **Lossless Compression** of Hyperspectral Imagery via Adaptive Filtering

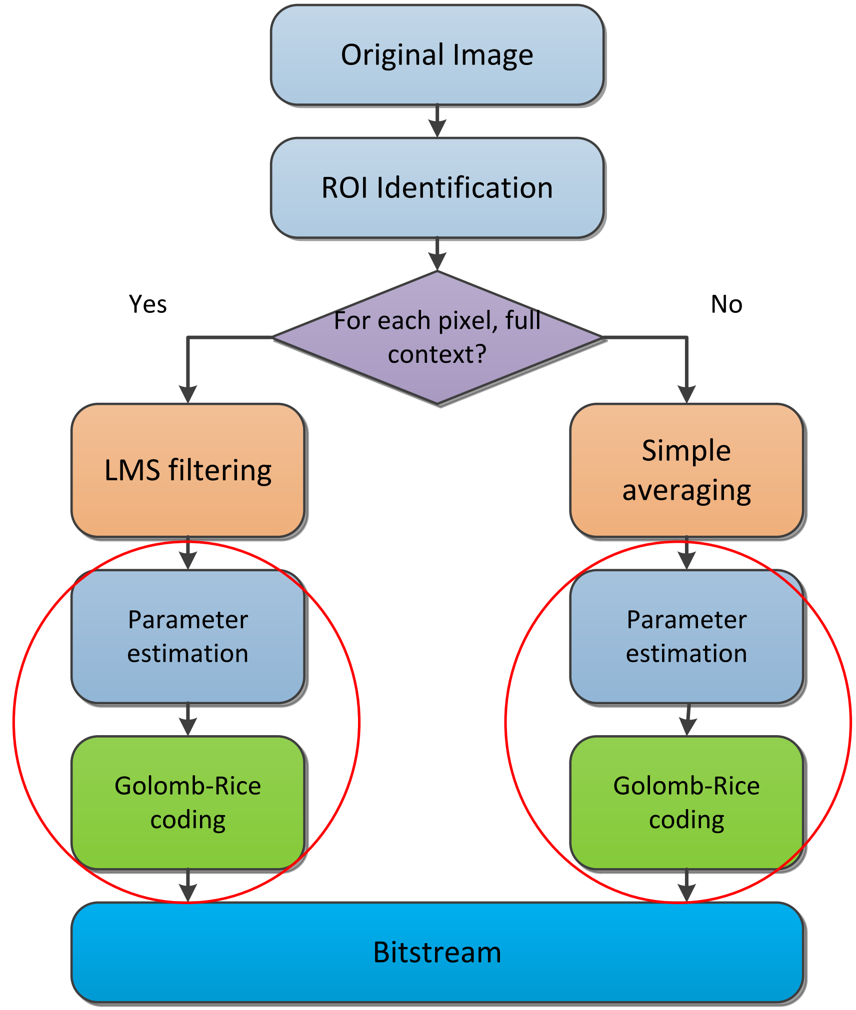
  http://ipnpr.jpl.nasa.gov/progress_report/42-163/163H.pdf
- Emerging CCSDS (Consultative Committee on Space Data Standards) Recommended Standard for Multispectral and Hyperspectral **Lossless Image Coding**
- JPEG-LS (lossless)
  The LOCO-I **lossless image compression** algorithm: principles and standardizations into JPEG-LS
  http://www.hpl.hp.com/research/info_theory/loco/HPL-98-193R1.pdf

  http://www.jpeg.org/jpeg/jpegls.html

Coding the Prediction Residuals

Regions of Interest





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, Jan. 2017.

Golomb-Rice (GR) Coding Scheme

- Golomb code is a family of codes parameterized by an integer ***m*** > 0. $p^m = ½$
  - For example, $p_1 = 0.5$, $p_2 = 0.707$, $p_3 = 0.7937$ for $m = 1, 2, 3$.
- Represent an non-negative integer to be coded as $n = mq + r$, where q is the quotient of (n/m), r is the remainder.
- The case with $m = 2^s$ also known as **Golomb-Rice (GR) Code**, where the remainder r is simply the s LSB's of n to be coded.
- Unary code for q (# of bits = $q + 1$)
- Binary code for r (# of bits = $\log_2 m = s$)

	$m = 1$			$m = 2$	
n	$G(n)$	Codeword	n	$G(n)$	Codeword
0	1/2	0	0	0.293	00
1	1/4	10	1	0.207	01
2	1/8	110	2	0.116	100
3	1/16	1110	3	0.104	101
4	1/32	11110	4	0.073	1100
5	1/64	111110	5	0.051	1101
6	1/128	1111110	6	0.036	11100
7	1/256	11111110	7	0.025	11101
8	1/512	111111110	8	0.018	111100
9	1/1024	1111111110	9	0.013	111101
10	1/2048	11111111110	10	0.009	1111100

Coding Parameter Estimation

- Coding Parameter m is critical to performance of GR code
- Given a sequence of nonnegative integers n, how to estimate the $m = 2^s$ such that $p^m = ½$?
  - Estimate $\mu \approx$ Sample Mean
  - $\mu = p/(1 - p)$ for geometric distribution, thus $p = \mu /(1 + \mu)$
  - $p^m = ½$, thus $\mu^m/(1 + \mu)^m = ½$
  - If $\mu >>1$, then $\mu^m/(1 + \mu)^m \approx 1 - m/\mu = ½$ by dropping higher-order terms in the Binomial Series
  - Thus $m = 2^s = \mu/2$

$$s = \max\left\{0, \left\lceil \log_2\left(\frac{\mu}{2}\right)\right\rceil\right\}$$

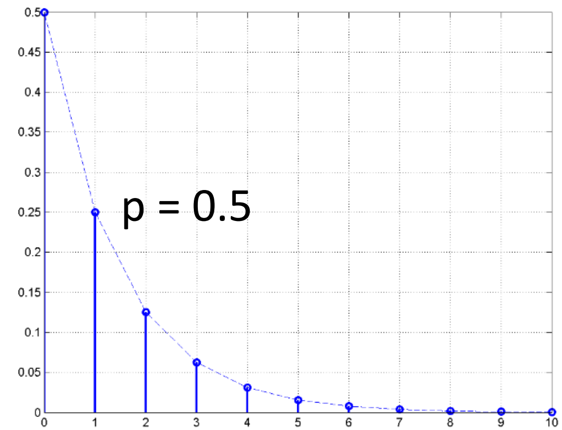
Challenge with Practical Data

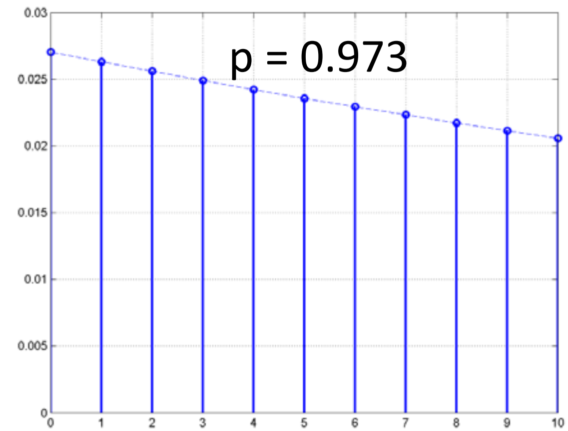
- There are other more accurate coding parameter estimation methods:
  - **Method 1** (A. Klimesh, 2005)
    $$s = \max\left\{0, \left\lceil \log_2\left(\frac{\mu}{2}\right)\right\rceil\right\}$$
  - **Method 2** (M. Kiely, 2004)
    $$s = \max\left\{0{,}1 + \left\lfloor \log_2 \frac{\log(\phi-1)}{\log\left(\frac{\mu}{\mu+1}\right)}\right\rfloor\right\}$$, where $\phi = \frac{\sqrt{5}+1}{2}$, the Golden Ratio.
  - **Method 3** (A. Said, 2006)
    $$s = \max\left\{0, \left\lceil \log_2 \mu - 0.05 + \frac{0.6}{\mu}\right\rceil\right\}$$
- All these parameter estimation methods were derived or optimized based on the assumption that <u>the underlying distribution is geometric distribution</u>.
- Nevertheless, the actual data could deviate from a perfect geometric distribution.

Deviation from the Geometric Distribution

In theory:

$G(n) = p^n q = p^n (1 - p)$

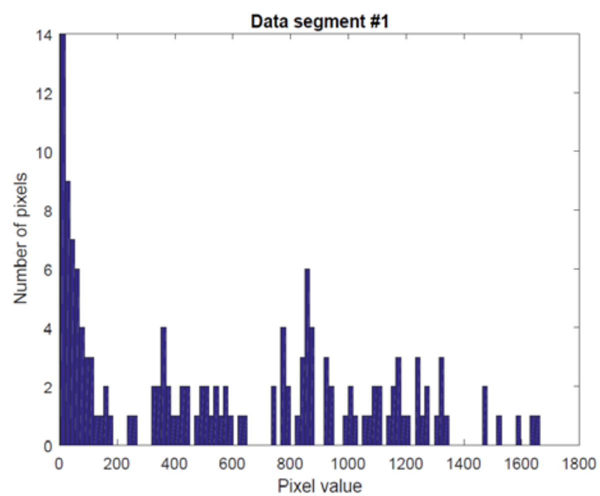


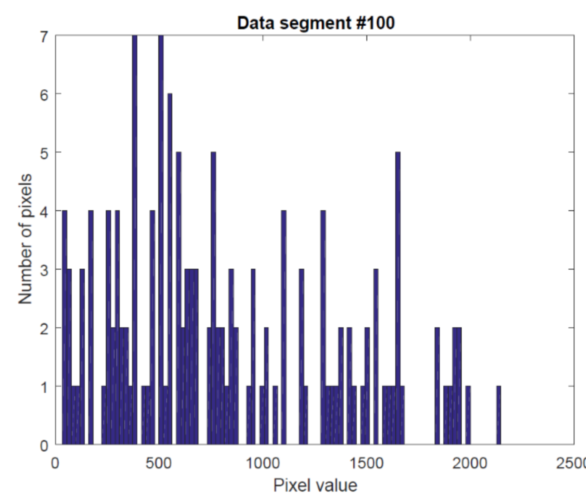


In Practice:

Significant deviation due to existence of in hyperspectral images:

- Edges, contours, corners, etc.
- Impulsive noise





Two prediction residual data segments from “Indian Pines”

Formulation as a Pattern Classification Problem

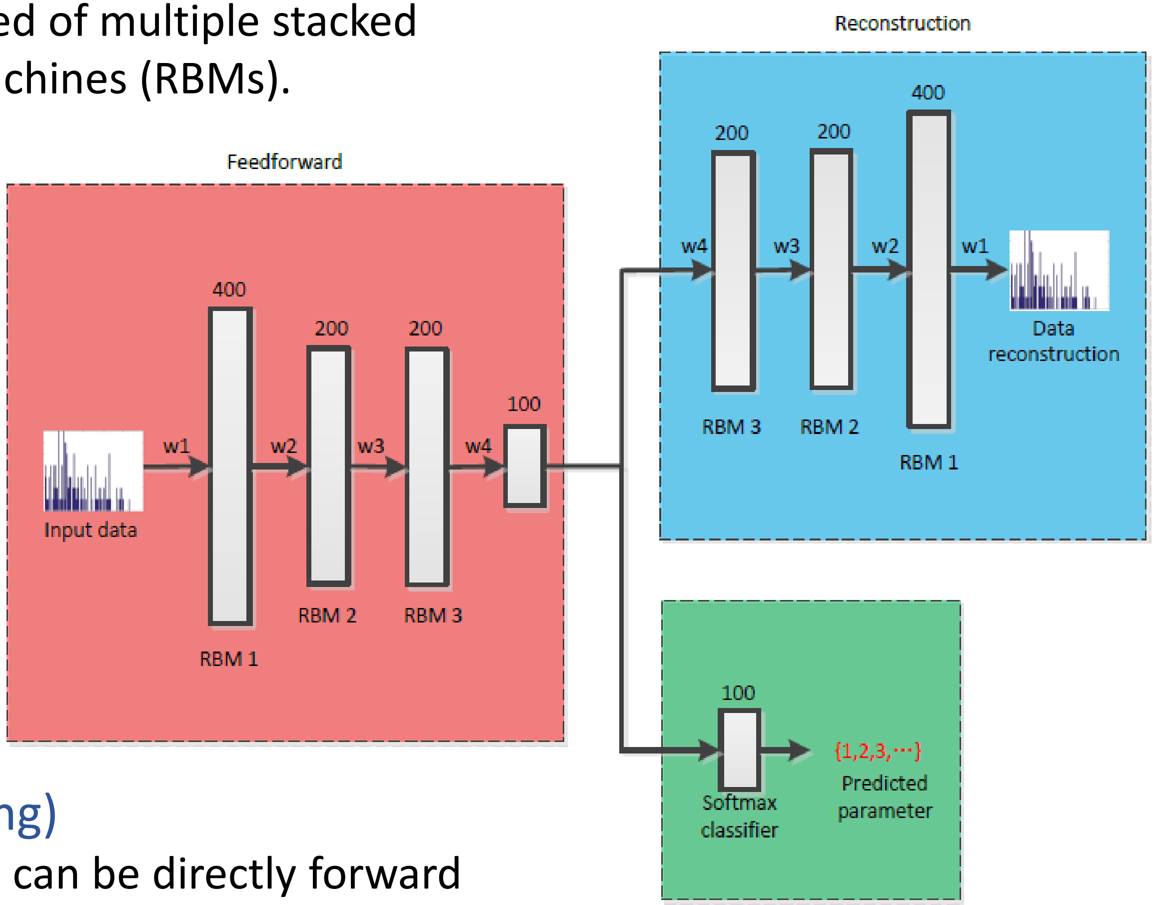
- we proposed a data-driven parameter estimation method without assuming any underlying distribution.
- The novelty lies in our formulation of the problem of choosing the best coding parameter for the given input data as a pattern classification problem.
- Traditional machine learning algorithms require 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).

Supervised Learning

- In most practical applications, there are only a finite number of parameter values to choose from for GR codes.
  - For example, coding a typical image with 16 bits/pixel would require only a set [0, 16] of 17 integers.
- Therefore, we can train a classifier, where the input is a data segment to be coded, and its “label” is the *m* value in the set, such that GR 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.
- Such a data-driven method does not require any knowledge about the underlying distribution of the input data, and thus would be generally <u>more robust</u> than methods that presume a certain distribution of the data.

Stacked Restricted Boltzmann Machines

The deep belief network (DBN) is a generative graphical model composed of multiple stacked Restricted Boltzmann Machines (RBMs).



Training

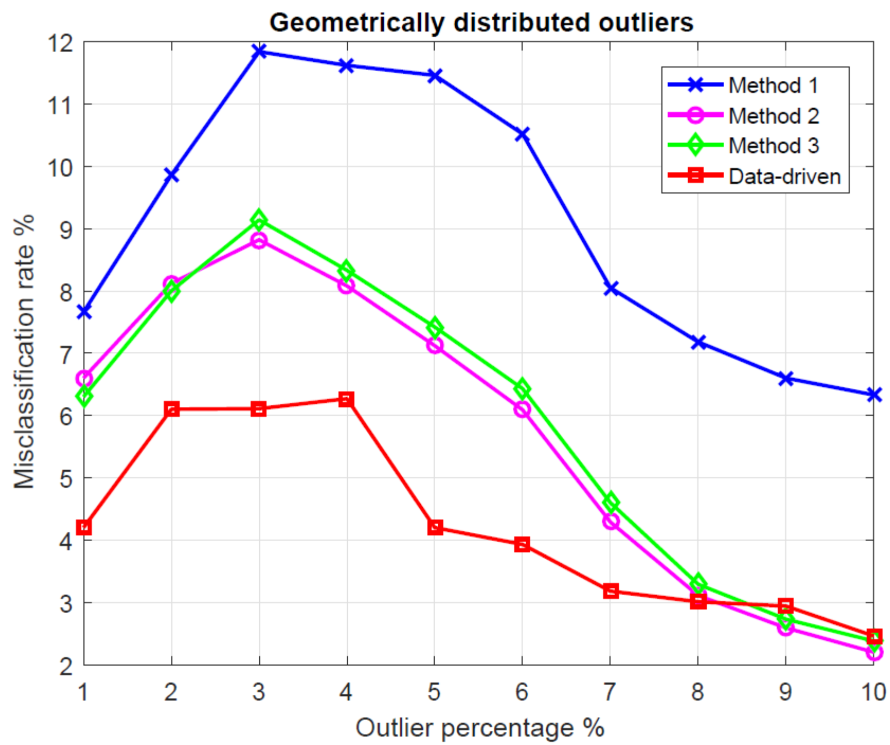
- Data segments with the ground truth (the optimal coding parameter value that gives the shortest codeword).
- the DBN can learn the underlying distribution of the data and adjust the weights within each layer of the network.

Testing (Actual Coding)

- new data segments can be directly forward fed through the network, thereby yielding the estimated parameter value at the output of the network.

Simulation Results (Synthesized Data)

- we synthesized data using distributions that deviate from the standard geometric distributions.
- The synthetic dataset follows 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



Actual Data from Hyperspectral Image Compression

- We used the so-called Fast Lossless (FL) predictor on "Indian Pines".
- Each row of a spectral band image as one data segment (containing 145 data samples), based on which the GR coding parameter was estimated.
- Out of 29,000 data segments from 200 bands, we randomly selected half for training, and the remaining half for testing.
- Only eight classes {2,3, ,,,,, 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.

Confusion Matrix

Rows: Predicted Parameter Value; Columns: Optimal Parameter Value (Ground Truth)

	2	3	4	5	6	7	8	9
2	97.58%	1.14%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
3	2.42%	98.58%	0.60%	0.00%	0.00%	0.00%	0.00%	0.00%
4	0.00%	0.28%	97.83%	1.74%	0.00%	0.00%	0.00%	0.00%
5	0.00%	0.00%	1.57%	96.85%	0.41%	0.00%	0.00%	0.00%
6	0.00%	0.00%	0.00%	1.41%	98.82%	1.48%	0.00%	0.00%
7	0.00%	0.00%	0.00%	0.00%	0.77%	96.57%	1.00%	0.00%
8	0.00%	0.00%	0.00%	0.00%	0.00%	1.95%	97.45%	1.12%
9	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.56%	98.88%

The proposed method achieved very high accuracy, from about 96% (for Class 7) to 98% (for Class 9), indicating the distribution of each data segment was well learned by the deep belief network.

False Estimation Rates

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

- This table 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.

Conclusion

- 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.
- Our Mission: to harness the power of machine intelligence to push the envelope of big data compression.