Proceedings of the
International Conference on Mechanical Engineering and Renewable Energy 2017
(ICMERE2017) 18 – 20 December, 2017, Chittagong, Bangladesh

ICMERE2017-PI-404

A TREE-BASED ARITHMETIC CODING BY ESTIMATING EXACT PROBABILITY DISTRIBUTION FOR DATA COMPRESSION

Tahmina Khanam[1], Pranab Kumar Dhar[2*] and Golam Moktader Daiyan [3]

[1]*Institute of ICT, Chittagong University of Engineering and Technology (CUET), Chittagong-4349, Bangladesh*
[2]*Department of CSE, Chittagong University of Engineering and Technology (CUET), Chittagong-4349, Bangladesh*
[3]Department of CSE, East Delta University, Chittagong, Bangladesh
tahminacse0904079@gmail.com, pranabdhar81@gmail.com, and daiyan@eastdelta.edu.bd
Corresponding author*

Abstract- *In recent era, with the vast increase of digital data, the need of digital data compression is increased day by day. Since storage area for digital data and also bandwidth for digital data communication is limited, compression technique can be utilized for storage and bandwidth effectively and efficiently. In this paper, a tree-based arithmetic coding method for data compression is proposed by estimating the exact probability distribution from tree. The main idea of this paper is to estimate the exact probability of the data by reducing the decoding time of encoded data. In addition, a tree approach is introduced which takes less time for estimating probability. Initially, a tree is created using four neighbourhoods from two dimensional data. During tree generation phase, each node is labelled with a probability. Then backtracking is applied to the tree to generate the exact probability of the data. Simulation results indicate that the proposed tree-based arithmetic coding has superior performance than the conventional arithmetic coding in terms of time complexity.*

Keywords: Arithmetic coding, exact probability distribution, tree, neighbourhood

1. INTRODUCTION

Now-a-days, digital data are increased profoundly all over the world. Multimedia data compression has been widely used to handle the large amount of digital data. It can be applicable for image, audio, video, and text also. Many state-of-the-art methods have been proposed in literature. A comprehensive survey on data compression can be found in [1]. Masmoudi *et al.* [2] proposed a new arithmetic coding based on exploiting inter-block correlation. However, it estimates uniform probability distribution in first block but it may not be uniform in real and also other blocks are estimated using dependency not exactly. Besides, an adaptive arithmetic coding based on adjacent data probability was proposed by Chuang *et al.* [3]. However, here, probabilities of blocks are estimated using adjacent data probability which introduces dependency. Carpentieri [4] proposed a technique to appropriately select each pixel position and encoded the current pixel prediction error by using the distribution as a model for AC. The scheme presented here is very slow and requires large amount of computations, especially for large images. Golchin *et al.* [5] proposed a new method consisting in combining context classification scheme with adaptive prediction and entropy coding in order to produce an adaptive lossless image encoder. Matsuda *et al.* [6] proposed to encode image prediction errors using a kind of CAAC. They proposed a model that approaches the probability density of errors by a generalized Gaussian function. The encoding algorithm is very slow and the encoder takes between 10 and 20 min for the encoding process. Kuroki *et al.* [7] have presented an AAC of prediction errors in lossless image compression. They proposed a model that estimates the probability density of each error pixel using Laplacian distribution with zero mean. Ye *et al.* [8] presented an experimental study on issues related to parametric probability modeling for entropy coding. Lynch *et al.* [9] combined wavelet transform and Huffman coding for lossless data compression and transfer to alleviate the power demand on wireless SHM systems. Zhang and Li [10] studied the wavelet-based and LPC methods, respectively, for compression of structural vibration sensor data. Bao *et al.* [11], Mascarenas *et al.* [12], Sadhu *et al.* [13], and O'Connor *et al.* [14] and Yongchao *et al.* [15] studied a new compression technique called compressive sampling for structural responses. Whereas encouraging results are seen, a high compression ratio is not available in lossless compression applications nor is accurate reconstruction achieved in a lossy compression algorithm. Considering the voluminous data that are measured from the dense sensor network in infrastructures, it would be desired to seek more effective methods realizing a high compression ratio as well as accurate reconstruction for efficient data transfer and further applications. Yang *et al.*

[16] proposed a data compression technique for calculating structural seismic responses using principled independent component analysis. It is first shown that independent component analysis (ICA) is able to transform a multivariate data set into a sparse representation space where is optimal for coding and compression, such that both the intra-dependencies and interdependencies (i.e., redundant information) between the multichannel data are removed for efficient data compression. The data compression technique is therefore of particular importance to manage such large data sets; it reduces the size of the original data from the acquisition station for transfer, and then reconstructs them at the data analysis station. An effective compression scheme contributes to efficient data transfer and fast access to measured data for real-time analysis and evaluation of the structure, which is critical for online monitoring and control.

In this paper, a tree-based arithmetic coding method for data compression is proposed by estimating the exact probability distribution from tree. The probability distributions used here include the distributions most commonly used in image modeling, such as Laplacian distribution, Gaussian distribution, t-distribution and generalized Gaussian distribution. Simulation results indicate that the proposed tree-based arithmetic coding has superior performance than the conventional arithmetic coding in terms of time complexity.

The rest of this paper is organized as follows. The proposed tree-based arithmetic coding is presented in section 2. Experimental results of the proposed method are discussed in section 3. Finally, section 4 concludes this paper.

2. PROPOSED METHOD

Lossless data compression is very much effective for sensible data compression such as medical signal data compression. Arithmetic coding technique is a famous one in the field of lossless data compression. In arithmetic coding, it uses data occurrence probability to encode and decode data. However, the probability estimation is often done with uniform probability distribution since calculating the probability of data among large data set is time consuming. In addition, this approach makes the decoding process lengthy by increasing the decoding steps. To overcome this limitation, in this paper a tree based arithmetic coding by estimating exact probability distribution is introduced. It has the following steps:

Step 1: Generate a tree T from 2D data $X = \{x_1, x_2, x_3, ..., x_n\}$ by considering the center element as root R and expand it with 4 neighborhoods as child $\{C_1, C_2, C_3, C_4\}$. Expand this tree until all data are added to it.

Step 2: Assign a probability to each node using Eq. (1) where, *C* is the current element, *n* is the total number of data, and *P* (*C*) is the probability of *C*.

$$\begin{array}{ll} p(C) = 1/n, & if\ \ C \neq R \\ p(C) = 1/tn + p(R)\,\&\,set(R = INF), & if\ C == R \end{array} \qquad (1)$$

Step 3: Backtrack to root and at the same time estimate the exact probability of the data using Eq. (2).

$$p(each\_element) = sum(p(each\_element \neq INF)) \qquad (2)$$

Step 4: Encode the each data using arithmetic coding and the probability of the data itself.

Step 5: Decode data with the same probability values.

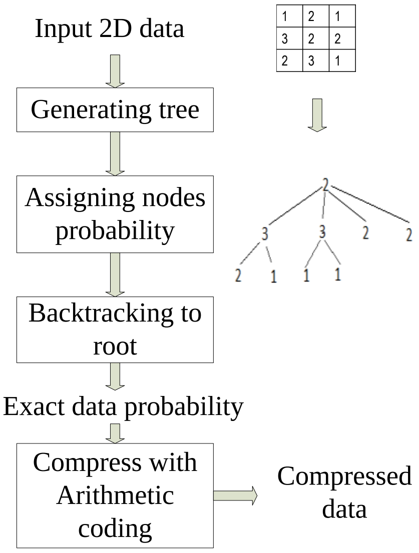


Fig. 1: Proposed compression technique with probability estimation scheme

3. EXPERIMENTAL RESULTS

A compression algorithm can be evaluated in a number of different ways. We could measure the relative complexity of the algorithm, the memory required to implement the algorithm, how fast the algorithm performs on a given machine, the amount of compression, and how closely the reconstruction resembles the original. In this work, we mainly concern with the one criterion that is time complexity since we are dealing with lossless compression. The proposed method is implemented in MATLAB environment. The system is simulated under core i5 processor and 4 GB RAM. The system efficiency in terms of time complexity for calculating data probability is shown in Table 1 where various images are used and Table 2 shows the required time for 2D matrix data.

After calculating the probability, the data are encoded using arithmetic coding. In the time of decoding the data, required time for decoding is decreased compared to the decoding scheme using uniform distribution as shown in Table. 3. This is due to the use of exact probability of the data. The proposed method takes less time for calculating probability due to the use of the tree approach. In addition to this, decoding time has decreased for exact probability distribution scheme.

For our proposed method, the time complexity remains approximately constant with the increase of data size. For both matrix data and image data, time complexity graph view is presented in Fig. 2. This graph shows the change of time complexity with respect to the increasing of data size. From this figure, it is shown that the complexity almost remain same for any data size. Here, for each data size, average time complexity of samples that have the specific data size are shown graphically. Since this approach is applied for lossless compression the main parameter to describe the system efficiency is time. From this Figure and the experimental results we observed that the proposed method shows sufficient efficiency in saving time. In other words, the proposed tree based arithmetic coding method provides better result than the conventional arithmetic coding.

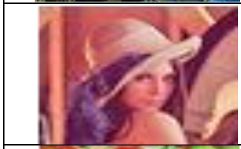
Table 1: Time required time for calculating the probability for various images

Image data	Required time (in sec.)
	1.213405
	1.045112
	0.984376
	0.951532
	1.125436
	1.015103
	1.334291
	1.340187
	0.995103

Table 2: Time required for calculating the probability of sample matrix data

Sequence of data	Estimated probability	Required time (in sec.)
$\begin{pmatrix} 2 & 2 & 3 & 1 \\ 2 & 2 & 3 & 1 \\ 3 & 3 & 2 & 1 \\ 1 & 2 & 2 & 3 \end{pmatrix}$	P(1)=0.2500 P(2)=0.4375 P(3)=0.3125	0.060134
$\begin{pmatrix} 1 & 2 & 2 & 3 & 1 \\ 1 & 2 & 2 & 3 & 1 \\ 2 & 3 & 3 & 2 & 1 \\ 3 & 1 & 2 & 2 & 3 \\ 3 & 1 & 2 & 2 & 3 \end{pmatrix}$	P(1)=0.4375 P(2)=0.6250 P(3)=0.5000	0.334291
$\begin{pmatrix} 3 & 1 & 2 & 2 & 3 & 1 \\ 2 & 1 & 2 & 2 & 3 & 1 \\ 2 & 2 & 3 & 3 & 2 & 1 \\ 3 & 3 & 1 & 2 & 2 & 3 \\ 1 & 3 & 1 & 2 & 2 & 3 \\ 3 & 1 & 2 & 2 & 3 & 1 \end{pmatrix}$	P(1)=0.6250 P(2)=0.8750 P(3)=0.7500	0.340187

Table 3: Comparison between the proposed method and conventional encoding scheme

Image data	Decoding time of proposed method	Decoding time with uniform distribution
	1.213405	2.912216
	1.045112	2.043215
	0.984376	1.982351
	0.951532	2.315943
	1.125436	2.463219
	1.015103	2.321102
	1.334291	2.819534

	1.340187	2.912216
	0.995103	2.043215

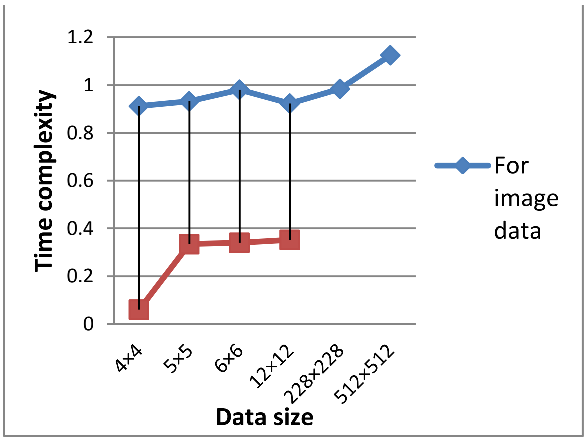


Fig. 2: Time complexity graph

4. CONCLUSION

In this paper, an arithmetic coding compression scheme is implemented with exact probability distribution. Initially, from 2D data, a tree is built up using four neighbourhoods. In tree generation phase, each node is labelled with a probability. Then backtracking in the tree generates the exact probability of the data. Simulation results demonstrate that the proposed tree-based arithmetic coding outperforms conventional coding in terms of time complexity, because of taking less time for estimating exact probability distribution. Simulation results verify that the proposed scheme can be effectively utilized for data compression. In future, the probability estimation of the proposed scheme can be further improved by dividing data into blocks and generating tree from blocks.

5. REFERENCES

[1] Sayood K., "Introduction to Data Compression. Morgan Kaufman Series", *Elsevier,* 2006.

[2] Masmoudi A. and William P., "An improved lossless image compression based on arithmetic coding using mixture of non-parametric distribution", *Multimedia Tools and Applications*, vol. 74, no. 23, pp. 10605-10619, 2015.

[3] Chuang C. P., Chen G-X., Liao Y-T., Lin C-C., "A lossless color image compression algorithm with adaptive arithmetic coding based on adjacent data probability", *in Proc. of the International Symposium on Computer, Consumer and Control*, pp. 141-145, 2012.

[4] Carpentieri B., "A new lossless image compression algorithm based on arithmetic coding", *in Proc. of the 9th International Conference on Image Analysis and Processing*, pp 54–61, 1997.

[5] Golchin, F., Paliwal, K., "A lossless image coder with context classification, adaptive prediction and adaptive entropy coding", *in Proc. of the IEEE International Conference on Acoustics Speech and, Signal Processing*, pp. 2545–2548, 1998.

[6] Matsuda, I., Shirai, N., Itoh, S., "Lossless coding using predictors and arithmetic code optimized for each image", *in Proc. of the International Workshop Visual Content Processing and Representation*, pp. 199–207.

[7] Kuroki, N., Manabe, T., and Numa, M., "Adaptive arithmetic coding for image prediction errors", *in Proc. of the IEEE International Symposium on Circuits and Systems*, pp. 961–964, 2004.

[8] Ye H., Deng G., Devlin JC., "Parametric probability models for lossless coding of natural images", *in Proc. of the EUSIPCO,* pp 514–517.

[9] Lynch, J. P., and Loh, K. J., "A summary review of wireless sensors and sensor networks for structural health monitoring", *Shock Vib. Digest*, vol. 38, no. 2, pp. 91–128, 2006.

[10] Zhang, Y., and Li, J., "Wavelet-based vibration sensor data compression technique for civil infrastructure condition monitoring", *J. Comput. Civ. Eng.*, vol. 6, pp. 390–399, 2006.

[11] Bao, Y., Beck, J. L., and Li, H., "Compressive sampling for accelerometer signals in structural health monitoring", *Int. J. Struct. Health Monitor.*, vol. 10, pp. 235–246,2011.

[12] Mascarenas, D., Chong, S. Y., Park, G., Lee, J.-R., and Farrar, C., "Application of compressed sensing to 2-D ultrasonic propagation imaging system data", *in Proc. of the 6th European Workshop on Structural Health Monitoring*, 2012.

[13] Sadhu, A., Hu, B., and Narasimhan, S., "Blind source separation towards decentralized modal identification using compressive sampling", *in Proc. of the 11th Int. Conf. on Information Science, Signal Processing, and their applications*, 2012.

[14] O'Connor, S. M., Lynch, J. P., and Gilbert, A. C., "Implementation of a compressive sampling scheme for wireless sensors to achieve energy efficiency in a structural health monitoring system", *in Proc. of the SPIE Smart Structures and Materials and Nondestructive Evaluation and Health Monitoring*, San Diego, 2013.

[15] Yongchao Y., Satish N., Yi-Qing N., "Data compression of very large-scale structural seismic and typhoon responses by low-rank representation with matrix reshape", *Thesis work*, Houston, USA.

[16] Yang, Y. and Nagarajaiah, S., "Data Compression of Structural Seismic Responses via Principled Independent Component Analysis", *Journal of Structural Engineering*, vol. 140, no. 7, 2014.