ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 8, August 2014

Numerical Data Compression With Data Representation Through Combinations

Elda CINA[1], Esmerald ALIAJ[2], Habib HAMAM[3,4]

Information Technology Faculty,"Aleksander Moisiu", University of Durres, Alabania[1]

Information Technology Faculty,"Aleksander Moisiu", University of Durres, Alabania[2]

Faculty of Engineering, University of Moncton, Canada[3]

School of Engineering, Canadian Institute of Technology, Albania[4]

Abstract: Digital technology is developing very rapidly, and the demand for devices to preserve data of large scale is increasing at an equal intensity. As such, the storage and transmission of data in efficient ways have become indispensable. The early solution seems to be the compression of data. With this article we aim to introduce a new technique of numerical data compression based on the Data Representation through Combination theory. Our approach is applied to lossless and lossy image compression. It presents better results compared to pre-existing methods, especially regarding lossless data compression. According to this technique, which uses the index representation of data instead of the original image, we gain not only compression but also other security benefits such as cryptography, steganography and watermarking.

Keywords: Data compression, compression rate, index, security

I. INTRODUCTION

With the widespread use of computers the efficient storage and transmission of large scale data has become necessary and is now very much in demand.

To minimize the size in bytes of information, compression can be used. It can be achieved by the removal of one or more of the three basic data redundancies [1]:

1. Coding Redundancy
2. Inter-pixel Redundancy
3. Psycho-visual Redundancy

Rapid changes in technology lead to new demands in compression techniques [2]. Information can exist in three different forms, 1D- (speech), 2D- (image), and 3D- (object in space). Our approach is applied to the second type of data— images. However, it can also very simply be applied to the other types of data without any significant changes. Image compression means the minimization of the size of graphical files while maintaining their quality to an acceptable level for the human visual system. The various compression techniques used can help reduce the storage and transmission costs of this type of data [3].

There are lossy and lossless techniques [4].

Lossless techniques are very important in applications where precision is valued, such as in medical imagery or deep space research. Some of the most known techniques for lossless compression are:

- ✓ Run Length Encoding (RLE) [5]
- ✓ Huffman Encoding [6,9]
- ✓ Lempel-Ziv-Welch Coding (LZW) [7,9]
- ✓ Arithmetic Coding [8]
- ✓ Delta encoding [8]
- ✓ Transform-based compression [8]
- ✓ Data Representation Through Combinations [9]

Lossy techniques on the other hand, are more widely used in aspects from web development to personal everyday use. Some of the most important standards are: JPEG, JPEG2K, JPEGXR, Fractal Compression, Embedded Zero tree[10][11], Wavelet Encoding, SPIHT Compression [12]etc.

Most recent methods regarding both lossy and lossless types of compression are a combination of traditional methods and new mathematical theories such as [13]. Other fields of study regarding data compression include neural networks described in[14]. Neural networks seem to be well suited for this particular task as they have a capacity to preprocess input patterns to produce simpler patterns with fewer components[15],[16]. In any case, lossless techniques have a compression rate of about2:1 or even slightly higher, but usually they cannot go any further than that. Except from[9] and another technique that is closely based on it called Spiral Compression [17], that can give a compression rate of 4:1 for lossless compression and 5.33 for an almost lossless compression [18].

With this paper, we are going to introduce another technique based on data representation through combination called "Numerical Data Compression." It is a technique that can result in lossless and adaptively lossy compression depending on the user's needs. It has a simpler algorithm than the two previously mentioned techniques, a better compression rate, as well as a much smaller work load for the processor.

II. DATA REPRESENTATION THROUGH COMBINATIONS

An image is generally represented by a matrix of samples. Each sample is expressed in binary way by a sequence of bits. If we take in consideration images we can say that an

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 8, August 2014

image with width W and height H is represented by ***W*H*** pixels. Each of the pixels (samples) could take values from 0 to 255 (for gray levels images). Ergo, to memorize an image we need ***W*H* 8*** bits.

The "Data Representation through Combinations" theory [9] presents a new way of defining an image. Instead of analyzing the pixels and writing them in the file, a unique number is assigned for each combination of pixels. This number is referred to as the "index." Thus an image is nothing but a combination of pixels and this combination is numbered. As such, we put this number in the file instead the original image.

If we consider the set of Images with 2 x 2 pixels (W=2 and H=2): Pixel 1 to Pixel 4 as pointed out by Table 1. Each pixel may have a value between 0 and 255=L-1(where L expresses the gray levels).The number of combinations is $L^{WxH}= 256^{2x2}= 2^{32}$, which means that the index of any image is between 0 and 2^{32}-1. To store this index in a file, we need between 0 and 32 bits (where 32 bits is the worst case).The lower is the index, the lower is the number of bits needed.

Index	Pixel 1	Pixel2	Pixel3	Pixel4
0	**0**	**0**	**0**	**0**
1	**1**	**0**	**0**	**0**
2	**2**	**0**	**0**	**0**
.				
.				
.				
4421768	**68**	**0**	**68**	**0**
.				
.				
.				
256^4-1	**255**	**255**	**255**	**255**

Table 1: all possible combinations for Image 2 x 2 pixels [19]

Therefore, the compression ratio varies from (example: combination # comb-1)

$$\min = \frac{M \times N \times roundup(log_2(L))}{roundup(M \times N \times log_2(L))} : 1$$

to (combinations 0 and 1):

$$\max = M \times N \times roundup(log_2(L)) : 1$$

Thus, depending on the combination, the compression ratio may be low (min) or very high (max).

For more details you can refer to the original article[9].

This theory gives great results not only in data compression but it intrinsically brings forth opportunities in other fields, such as information security, as it can easily be used for cryptography, steganography and watermarking.

III.PROPOSED APPROACH

According to Data Representation through Combinations, we extract a sole index for the entire given file data (1D signal, the image, or the object in space). In reality such a thing cannot be applied due to hardware limits. Instead we can split the object into small pieces and find the index of each piece. Therefore, the result is a group of indexes instead of the original signal, which would evidently be smaller than the original. In the following text, we introduce a new technique called "Numerical Data Compression" which operates with integers. Our approach has been tested on one type of data, being 2D signals (images), but it can also be applied to other-dimensional types of data. The principal concept of Numerical Data Compression is expressed by the following algorithm:

Encoding process:

1. Turn image into vector
2. Save positions of negative numbers
3. Spit vectors into groups with 6 pixels
4. For each group find index
5. Repeat the same process for Signs vector
6. At the end we have the size of the vector, a vector of indexes and a vector of sign indexes.
7. Send to the receiver

Decoding process:

1. Receive the file
2. Divide the vector of indexes from vector of signs
3. Convert indexes to pixels values
4. Repeat the same process for vector of signs
5. Join groups of pixels
6. Return negative values
7. Turn vector into image

This technique can be used to obtain lossless data compression as described in the algorithm above but also lossy data compression. In this case, the algorithm will change slightly. We add a division step between step 2 and 3 of the encoding process and a multiplier step between step 6 and 7 of the decoding process. The factor of division is altered form 1 (lossless) to 10 for a satisfactory compression. Therefore, we lose from 0 to 9 gray levels by which the difference is hardly detectable by the human visual system. Clearly, this factor could be more than 10 but in this case the artefacts would be noticeable by the naked eye.

IV. RESULTS

A number of experiments have been carried out in order to evaluate the performance of the proposed algorithm. Different contents of varying sizes have been tested. In the image below, we can find three samples of different qualities of compression accompanied by their corresponding results.

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 8, August 2014

.

Compression multiplier = 1	Compression multiplier = 2	Compression multiplier = 4
		
Compression multiplier = 8	Compression multiplier = 10	Compression multiplier = 12
		

Image	**Lena**					
Size (pixels)	W-256 H-256					
Compression multiplier	1	2	4	8	10	12
Compression rate [CR]	5,35	6.13	7,61	8.32	9	9.66
CPU time used for compression	0.3	0.094	0.094	0.062	0.078	0.094
CPU time used for decompression	0.9	0.265	0.265	0.234	0.218	0.218
(RMS)	0	0.71	1.88	4.16	5.34	6,53
(PSNR) dB	Infinity	51.17 dB	42.66 dB	35.75 dB	33.58 dB	31.84 dB
Structural Similarity (SSIM)	1,000	0,998	0.992	0.970	0.956	0,943

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 8, August 2014

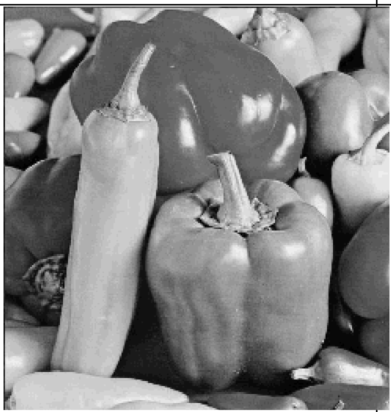
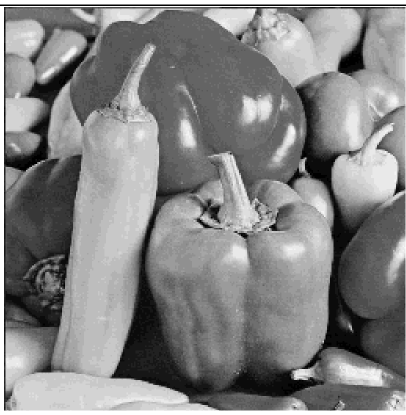
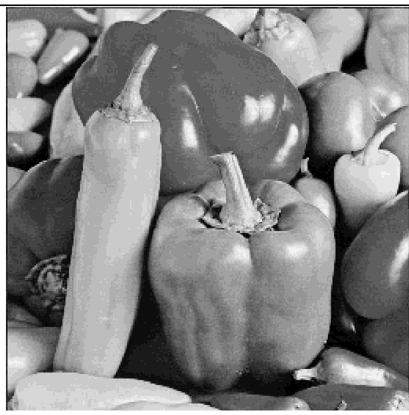
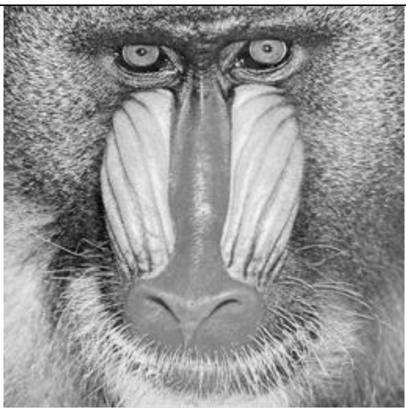
Compression multiplier = 1	Compression multiplier = 2	Compression multiplier = 4
		
Compression multiplier = 8	Compression multiplier = 10	Compression multiplier = 12
		

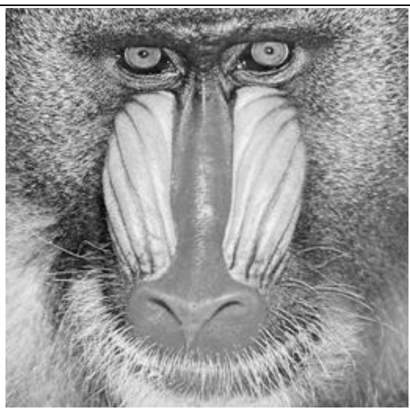
Image	**Peppers**					
Size (pixels)	W-512 H-512					
Compression multiplier	1	2	4	8	10	12
Compression rate [CR]	5	6.17	7.67	8.39	9.08	9.75
CPU time used for compression	0.30	0.39	0.25	0.22	0.20	0.09
CPU time used for decompression	1.30	1.19	0.98	0.89	0.87	0.17
(RMS)	0	0.71	1.87	4.17	5.36	6.53
(PSNR) dB	Infinity	51.14 dB	42.69 dB	35.72 dB	33.55 dB	31.83 dB
Structural Similarity (SSIM)	1,000	0.998	0.989	0.957	0.937	0.913

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 8, August 2014

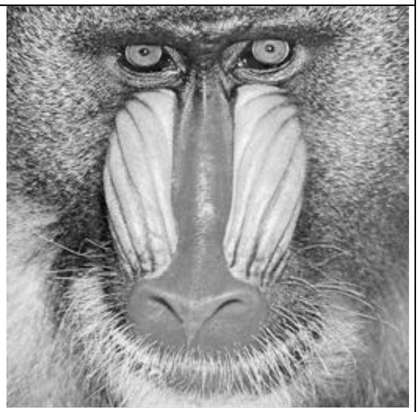
Compression multiplier = 1



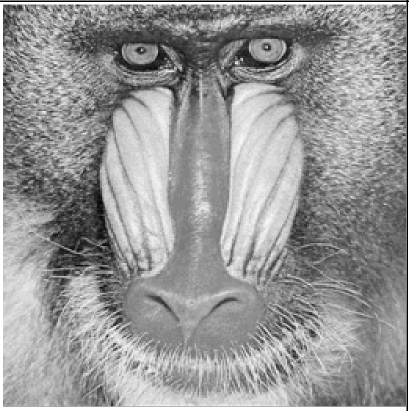
Compression multiplier = 2



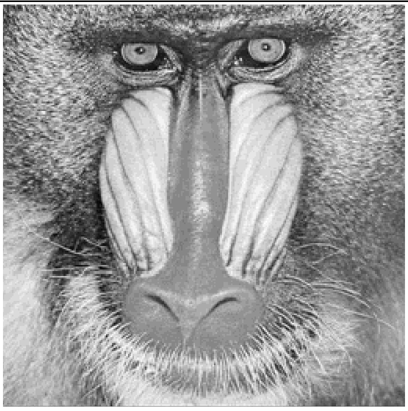
Compression multiplier = 4



Compression multiplier = 8



Compression multiplier = 10



Compression multiplier = 12

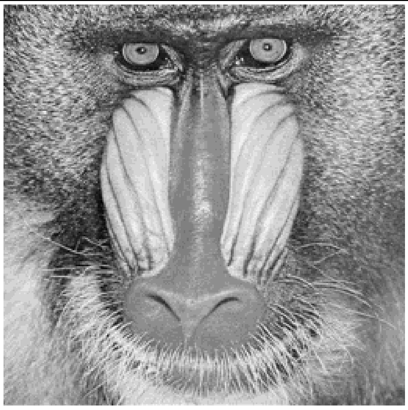


Image	**Mandril**					
(pixels)	W-600-420					
Compression multiplier	1	2	4	8	10	12
Compression rate [CR]	5.35	6.13	7.61	8.32	9	9.65
CPU time used for compression	0.36	0.09	0.06	0.08	0.08	0.13
CPU time used for decompression	0.09	0.33	0.27	0.25	0.23	0.56
(RMS)	0	0.71	1.87	4.17	5.34	6.51
(PSNR) dB	Infinity	51.14 dB	42.69 dB	35.72 dB	33.58 dB	31.87 dB
Structural Similarity (SSIM)	1,000	0.999	0.997	0.987	0.980	0.971

You can see from the results that a high compression rate with small error metrics is obtained. We get a compression rate over 5:1 for lossless compression which is better from the usual compression methods which have a compression factor of 2:1, but also better from our previews research on [17] which had a compression factor of 4:1. It also gives significant compression rates for lossy compression. The quality factor is adaptable by the user's needs in terms of image quality or compression rate.

V. CONCLUSIONS

With this paper, we have described a new technique of image data compression based on "Data Representation through Combinations." This method has been tested on images producing satisfactory experimental results, but it can also be used another numerical data without changes. It offers lossless and lossy data compression at rates similar to, or higher than existing methods. It starts from a compression rate factor of 5 for lossless data compression. Our suggestion is that the lossy variant should not surpass the quality factor of 10 which leads to a compression rate of about 9, and according to the user's needs in terms of image quality, we can obtain higher and higher compression rates.

Optimization of this technique and further development of it in the space and frequency domain will be the subject of future work. Another avenue to be explored consists of extending our technique towards security practices such as cryptography, steganography and watermarking.

VI. REFERENCES

[1] Digital Image Processing Using MATLAB by Rafael C.Gonzalez,RichardE.Woods& Steven L.Eddins
[2] P.Nirupama BVRIT "Analysis of Various Image Compression Techniques" . vol,2no.4, May 2012. ARPN Journal of science and Technology
[3] M. I. Khalil, "Image compression using new entropycoder", International Journal of Computer Theory andEngineering, vol. 2, no. 1, pp. 39-41, February, 2010
[4] Rajaravivarma, V., Lord, E., & Barker, J. (2001). Data compression techniques in image compression for multimedia systems
[5] "Comparison of JPEG image coders". Grgic, S. Mrak, M., &Grgic, M. (2001) Retrieved October 23, 2003, from http://www.vcl.fer.hr/dtv/sgrgic_vip2001.pdf
[6] " A Hybrid Image Compression Technique," Acoustics Speech & Signal Processing, Dr. Charles F. Hall, , IEEE International Conference on ICASSP' 85, Vol.10, pp 149- 152, Apr, 1985
[7] "Optimized RTL design and implementation of LZW algorithm for high bandwidth applications". Electrical Review 2011 (4): 279–285.
[8] "Generalized Kraft Inequality and Arithmetic Coding". IBM Journal of Research and Development 20: 198–203. doi:10.1147/rd.203.0198 J. Rissanen (1976).
[9] H. Hamam "A new representation of data through combinations", ICCAT 2013
[10] C. Valens ,Embedded Zerotree Wavelet Encoding – ©,1999 http://140.129.20.249/~jmchen/wavelets/Tutorials/c.valens/ezwe.pdf
[11] Jerome M Shapiro, Embeded Image coding Using zorotreesof wavelet coefficients, IEEE Transcations on Signal Processing Vol 41 No.12 December 1993 https://diuf.unifr.ch/courses/03-04/imaging /shapiro-93.pdf Amir Said A New Fast and E�cient Image Codec Based
[12] on Set Partitioning in Hierarchical Trees , IEEE Transactions on Circuits and Systems for Video Technology, Vol 6, June 1996 http://www.cipr.rpi.edu/research/SPIHT/EW_Code/csvt96_sp.pdf
[13] S. Smith, The Scientist and Engineer's Guide to Digital Signal [10]A. Alarabeyyat1, S. Al-Hashemi1, T. Khdour1, M. Hjouj Btoush1, S. Bani-Ahmad1, R. Al-Hashemi2 Lossless Image Compression Technique Using Combination Methods Journal of Software Engineering and Applications, 2012, 5, 752-763 http://dx.doi.org/10.4236/jsea.2012.510088 Published Online October 2012 (http://www.SciRP.org/journal/jsea)
[14] Aslam Khan, Sanjay K Mishra, Shiv K. Sahu, Amit Mishra "A Technique for Image Compression by using GSOM Algorithm" International Journal of Computer Applications (0975 – 8887) Volume 87 – No.5, February 2014
[15] Martin Riedmiller and Heinrich Braun "A Direct Adaptive Method for Faster Back propagation Learning: The RPROP Algorithm" Institute fur Logic, Komplexitat und Deduktionssyteme, University of Karlsruhe, W-7500 Karlsruhe, FRG, 1993.
[16] Sangeeta Mishra TCET, MumbaiThakur Village, Kandivli(E)-101 "Image Compression Using Neural Network" International Conference & workshop on Recent Trends in Technology, (TCET) 2012 Proceedings published in International Journal of Computer Applications(IJCA)
[17] Compressing 2D Images through data representation using spiral path, E. Cina, E. Aliaj, H. Hamam, ISTI 2014
[18] E. Cina, E. Aliaj, H. Hamam, Increasing compression rate of Images through Data Representation Using Spiral Path and Biplanes,International Journal of Engineering Research & Technology Vol. 3 - Issue 7 (July - 2014) e-ISSN: 2278-0181
[19] Said, R. Souissi, H. Hamam "A New Representation of Image Through Numbering Pixel Combinations"J. pg. 46 Journalof Information Security Research Volume 4 Number 1 March 2013

BIOGRAPHIES

Elda CINA has graduated the Faculty of Electric Engineering, Computer engineering of Polytechnic University of Tirana in 2006. She also holds a second level diploma in Business Administration from 2009 at the Faculty of Economy, University of Tirana. Now she is a PhD student in Faculty of Economy, University of Tirana and her research field is Information Systems Security. She is working as a lecturer in the "Äleksandër Moisiu" University of Durrës for 7 years now. Currently she is head of Information Technology department in the Information Technology Faculty. She is the author of 1 book in Networking and some different articles in the field of information Systems.
Her fields of interest are networking, computer architecture, signal processing, etc.

Esmerald ALIAJ has received a B.S. from the Technological Institute of Athens in 2013 in Electronics Engineering. He is working as a Part Time Lecturer at Aleksandër Moisiu"University of Durrës. His research interests center on improving the understanding and efficiency of Genetic Algorithms, mainly through the use of distributed systems and GPU implementations. He has also shown interest in 2D data analysis, specifically image compression and machine vision for use in intelligent systems.

Habib Hamam obtained the B.Eng. and M.Sc. degrees in information processing from the Technical University of Munich, Germany 1988 and 1992, and the PhD degree in Physics and applications in telecommunications from Université de Rennes I conjointly with France Telecom Graduate School, France 1995. He also obtained a postdoctoral diploma, "Accreditation to Supervise Research in Signal Processing and Telecommunications", from Université de Rennes I in 2004. He is currently a full Professor in the Department of Electrical Engineering at the Université de Moncton and a Canada Research Chair holder in "Optics in Information and Communication Technologies". He is an IEEE senior member and a registered professional engineer in New-Brunswick. He is among others associate editor of the IEEE Canadian Review, member of the editorial boards of Wireless Communications and Mobile Computing - John Wiley & Sons - and of Journal of Computer Systems, Networking, and Communications – Hindaw-i Publishing Corporation. His research interests are in optical telecommunications, Wireless Communications, diffraction, fiber components, RFID, Human-Machine interaction and E-Learning.