КОМП'ЮТЕРНІ НАУКИ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

УДК 004.4 ' 23 DOI: 10.31388/2078-0877-19-3-276-288

ANALYSIS OF THE EFFICIENCY OF DATA COMPRESSION IN A THREE-DIMENSIONAL SCANNING SYSTEM USING THE RLE ALGORITHM

Miguel Hurtado Madrid, research Laboratory in Digital Systems and Renewable Energy,
Gregorio Trinidad Garcia, research Laboratory in Digital Systems and Renewable Energy,
Jose I. Cortez, doctor, Research Laboratory in Digital Systems and Renewable Energy,
Jose Luis Ameca, research Laboratory in Digital Systems and Renewable Energy,
Fernanda Merlo Simoni, research Laboratory in Digital Systems and Renewable Energy,
Cesar Antonio Aguilar Rodríguez, research Laboratory in Digital Systems and Renewable Energy
Benemerita Universidad Autonoma de Puebla, Mexico
Hanna Alieksieieva, PhD (pedagogical sciences), associate professor
Berdiansk State Pedagogical University, Berdiansk, Ukraine
Natalya Sosnytska, doctor of pedagogical sciences, professor,
Head of the Department of Higher Mathematics and Physics
Tavria State Agrotechnological University, Melitopol, Ukraine

***Summary* – the techniques used to compress data allow a bigger transfer of information due to the processing that is performed, by means of numerical and statistical transformations, it is possible to represent the same information in a smaller space. One of the main problems that exist in communications systems lies in the use of the channel. The current studies focus on the study of the characteristics and effects of the noise present in the channel and the disturbances generated in the transmitted information. This paper studies the efficiency of the Run Length Encoding (RLE) compression algorithm in a wired communication channel, by measuring some parameters it is possible to determine the efficiency of the use of the channel, using the RLE algorithm, a methodology is used for the study, which allows determining the efficiency of the use of the channel, prior to the use of**

the compression algorithm and afterwards, in this way it is possible to determine the efficiency of the use of the channel, in this work, there is an increase in the use of the channel due to the RLE compression algorithm.

***Keywords* – RLE algorithm, data compression, entropy, BER.**

I. INTRODUCTION

In communication systems, the amount of data that is transmitted is a critical aspect to consider because as more data is sent in the shortest possible time, the communication channel is used more efficiently.

There are two ways to increase the amount of data sent in a communications system, the first one is to increase the speed of transmission and the second one is the coding of the information sent [1].

The speeds used for data transfer are limited by communication standards, which requires the use of several coding techniques that allow an increase in the rate of sending information.

Encoding is classified into two types: encryption and compression, the first is intended to provide protection to the information, however, the encryption process generates longer data frames.

The purpose of compression is to reduce the number of symbols used in the representation of information, this is done through the development of statistical processes and mathematical models applied to the information being compressed [1] [2].

There are two types of compression algorithms, lossless compression algorithms and lossy compression algorithms, the difference between these two types of compression lies in the type of processing given to the information [3] [4].

One of the main advantages of lossless compression algorithms is the integrity of the information during the coding process that does not present implementation complexity since their architectures are based on mathematical models previously demonstrated [5] [6] [7].

Lossless compression methods are classified into two types: substitutional (or dictionary-based) and methods based on mathematical transformations. In statistical methods, the coding of a symbol is based on the context in which it occurs, while substitutional methods group symbols to create an implicit type of context [8]. A compression system based on a statistical algorithm achieves better compression ratios than a substitutional compressor but the computational complexity and memory requirements for statistical algorithms are much greater than for substitutional compressors [4] [9] [10] [11].

Currently, the studies that focus on compression algorithms consider diverse applications of knowledge: medical sciences, communications systems, image processing, among others [12] [13] [14]

[15].

The present article describes the study of the efficiency of the RLE compression algorithm, through the application of a methodology will be determined the efficiency of the algorithm used in a three-dimensional scanning system of high precision.

Section II describes the RLE compression algorithm, its characteristics, advantages and limitations of use. In addition, the programming model that defines the algorithm is studied. Section III describes the methodology used in the development of the work, defining the parameters used in the methodology.

Section IV presents the development of the methodology applied to the RLE compression algorithm, describing the simulations and experiments performed. In section V the results obtained are discussed and the conclusions are established in section VI.

II. RLE COMPRESSION ALGORITHM

RLE is an algorithm that bases its operation on the search for consecutive redundant sequences of characters. When this sequence appears, it is replaced by a special character, which indicates the compression and is stored in a byte, then there is the number of characters that are compressed and stored in another byte [16].

When the dataset arrives at the receiver, it identifies the special character, which indicates that the compression has been carried out, then it receives the symbol that indicates how many characters have been compressed, allowing in this way to reconstruct the original sequence.

The data compression algorithm physically reduces any type of repeated character sequence once the characters reach a predetermined level of occurrence.

In the special case where the repeated character is null, the algorithm works in the same way as the algorithm known as Null Suppression [5].

Compression is made by scanning the source file for repeated characters, and usually requires the use of a special character to indicate that compression is in progress.

When a compression indicator character (SC) is used, it is accompanied by one of the repetitive characters on which the compression will be carried out (x) and finally, a numeric type character is added, which will indicate the frequency of appearance of such character within a string (CC) [4] [17].

The flowchart describing the behaviour of the algorithm (Fig. 1) describes the behaviour of the algorithm and is characterised by the use of dynamic memory arrangements to improve data transmission. The principle of operation lies in the study of the redundancy of the information that allows a compressed frame to be generated.

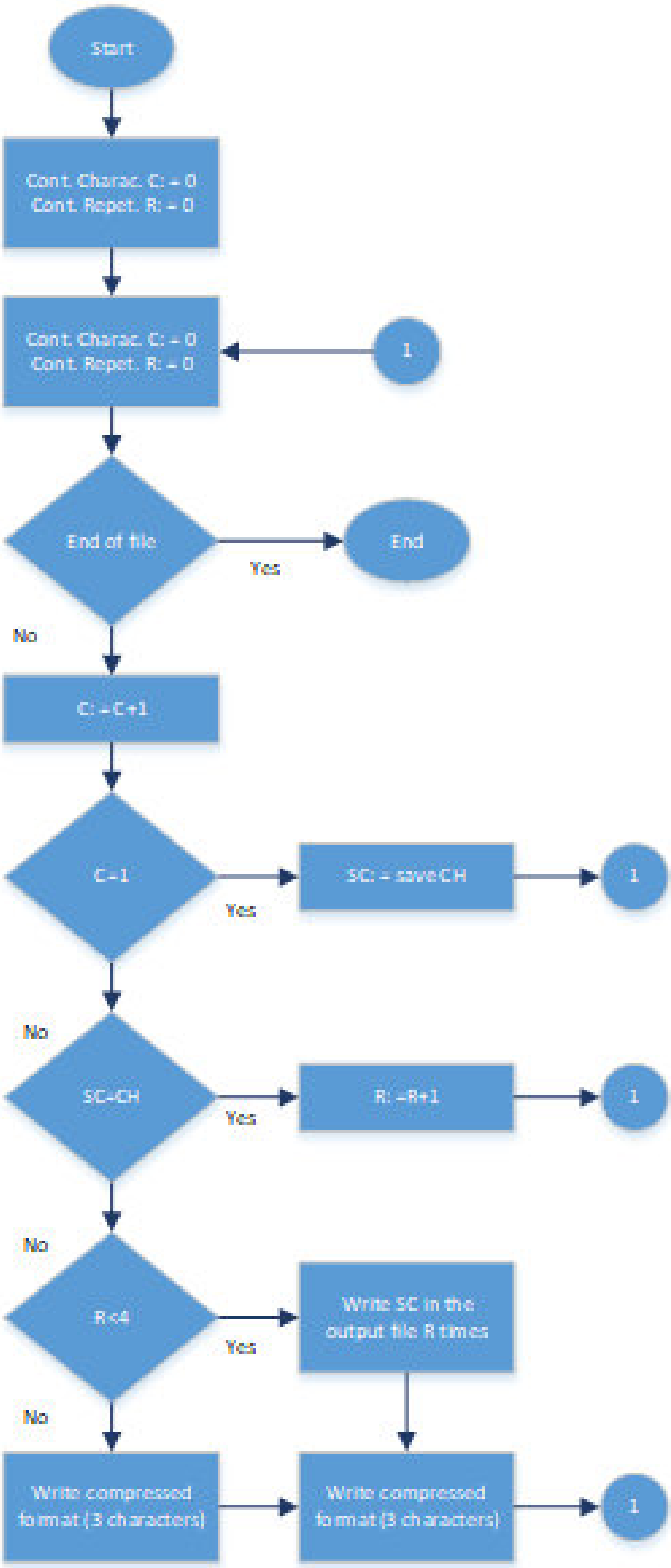

Fig. 1. Compression algorithm RLE

Because of the architecture of the algorithm, it is possible to implement it in programmable logic devices because it has iterative control structures. One of the advantages is that it does not require a previous analysis of the information because it performs processing in the data while generating the compression.

The Compression Rate (CR) is defined as:

$$CR = \frac{NCod}{NOri} x100\% \quad (1)$$

In the equation, $NCod$ is the total number of bits in the coded word, while $NOri$ is the total number of bits in the original word. To get an idea of the compression ratios produced by RLE, we consider a string of N characters, which needs to be compressed. It is assumed that the string contains M repetitions, with an average length of L elements each. Each of the M repeats, is replaced by 3 characters (code change, counter and data), so the size of the compressed string $(N - M)\,(L + M)\,3 = N - M)\,(L - 3)$, and the compression factor is:

$$\frac{N}{(N-M)(L-3)} \quad (2)$$

The implementation of the algorithm presents limitations for its implementation in digital systems with low memory resources, because of the need to use dynamic memory for the development of the compressed frame.

III. Methodology For the Study of Efficiency

The efficiency of a compression algorithm allows obtaining information regarding the quality after the transmission process of the compressed information, in this way it is possible to quantitatively determine the advantages and disadvantages of using the compression algorithm for a certain application.

The methodology proposed for the study of the efficiency of the algorithm is divided into five parts:

- Development of simulations based on the data to be compressed.
- Measurement of the evaluation criteria of the results obtained in the simulations.
- Elaboration of hardware experiments.
- Measurement of the evaluation criteria of the results obtained in the experiments performed.
- Analysis of the results obtained for the formulation of optimization proposals.

a. Compression Ratio

The compression algorithms allow reducing the number of symbols in an information chain, such reduction is measured by mathematical relations that determine the percentage of reduction obtained by compression, this relation is called compression relation or compression ratio, which is fundamental to quantify the percentage of reduction of information reached by the compression algorithm RLE, if this is greater than 1 it is called expansion and is produced for two reasons: the first one is explained by the fact that the algorithm was not implemented correctly, and the second one is because the type of information is not adequate for the implementation carried out. Since the data do not have the same characteristics in images, text or audio [18] [12].

The compression ratio is defined by the mathematical expression (3), which is expressed as a function of its specific scalar information [19].

$$\boldsymbol{Ratio} = \frac{\textbf{\textit{number of data compressed}}}{\textbf{\textit{number uncompressed data}}} \qquad (3)$$

The compression ratio can also be called bit per bit (bpb) since it is equal to the number of bits needed to represent a symbol in a compressed string.

In image compression, the term bpp means bits per pixel. Consequently, the main purpose of data compression is to represent any data at low bit rates. The term cost per bit refers to the role played by individual bits in the compressed chain.

b. BIT ERROR RATE

Another parameter that identifies the efficiency of a compression algorithm is the Bit-Error-Rate (BER), which measures the number of bits lost in the transmission of a digital system as a result of the noise of the channel used to transmit the information. [16]

Digital transmission systems have advantages over analogue systems in terms of the ability to combine and transport data from different applications, processing power to compress information, and reduction in channel bandwidth.

One of the disadvantages is the abrupt degradation of the recovered digital information in front of the reduction of the signal power or increase of the noise and/or interference, when the received signal is contaminated with Gaussian noise, the binary error rate can be determined analytically for systems without coding, when additionally there are interferences and/or coding for the correction of errors, the process of analytical determination of the error rate is a complex problem, for this reason it is usual to resort to computational simulations for the evaluation of the performance of the transmission systems. [17]

The transmission channel is used as the means by which electrical signals are transmitted, the relationship between the use of the transmission channel and a compression algorithm is measured by entropy, and determines how effective is the channel used by an algorithm is.

c. ENTROPY

Entropy is a concept that represents the limits of coding, where data is processed without the need to know the nature of the data. Entropy denotes the minimum number of bits per symbol needed to represent a string. It is an index that indicates the amount of information that exists in a data source. Although the calculation of general entropy cannot be analysed, in a practical way first-order entropy is usually used as an approximation. If you have a set of possible events whose probability of occurrence is $p_1, p_2, p_3, \dots, p_n$, these probabilities are known, but that is all you know about when the event occurs.

If a measure of the quantity of choice is found, for example, H ($p_1, p_2, \dots, p_n$,), you get the following properties:

1. H must be continuous in p_i.

2. If all p_i are the same, $p_i = \frac{1}{n}$, then H must be a monotonous increasing function of n. With equally probable events there is more possibility of choice, or uncertainty, when there are more possible events.

3. If the option is split into two successive choices, the original H must be the weighted sum of the individual values of H. (Fig. 2).

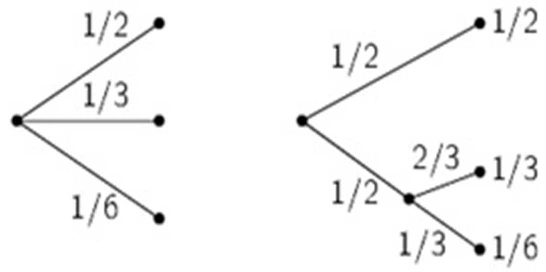

Fig. 2. Decomposition of the possibility of choosing between three options

DEVELOPMENT

The development of the methodology for the study of the efficiency of the algorithm contemplates the accomplishment of simulations and practical experiments that will provide the necessary information to determine in a quantitative way its efficiency.

In the simulations that were conducted, a numerical and statistical analysis platform was used, which integrates the tools for the development of the study considering previous data obtained from the three-dimensional reconstruction system.

The data used both for the performance of these simulations and for the practical tests was made on the basis of a sample file, the selection of the file was made by means of a statistical study that better represents the entire mean of study whose data represent the samples in a format of integers without signs with a length of 10 bits.

The arithmetic means of the files and the general average were compared, and the file was selected whose arithmetic mean approached the value of the general average, thus ensuring that the results obtained will be valid for the whole set of samples.

Chart 1 – Numerical results obtained from the statistical analyses to the files of the scanned samples

General mean of the samples	394.6791
Size of file 4	394.3826

The simulation of the RLE compression algorithm consisted in the coding of the flow chart, see Fig.1, the tests applied to the algorithm were based on the determination of parameters such as uncompressed number of symbols, compressed number of symbols, data compression ratio, and data compression percentage, where the following results were obtained:

The compression ratio obtained for the algorithm is presented in, in order to compare the number of symbols without compression with the number of symbols after compression, it is observed that the total number of symbols after compression is less than the total number of symbols to transmit without compression by a factor of 25.8 %.

The numerical results obtained are shown in and are used in the study of the efficiency of the algorithm.

Chart 2 – Results obtained from the simulation of the RLE compression algorithm

Number of uncompressed symbols	3000 symbols
Number of compressed symbols	2226 symbols
Data compression ratio	0.7420
Data compression percentage	25.8000%
BER	0 bits

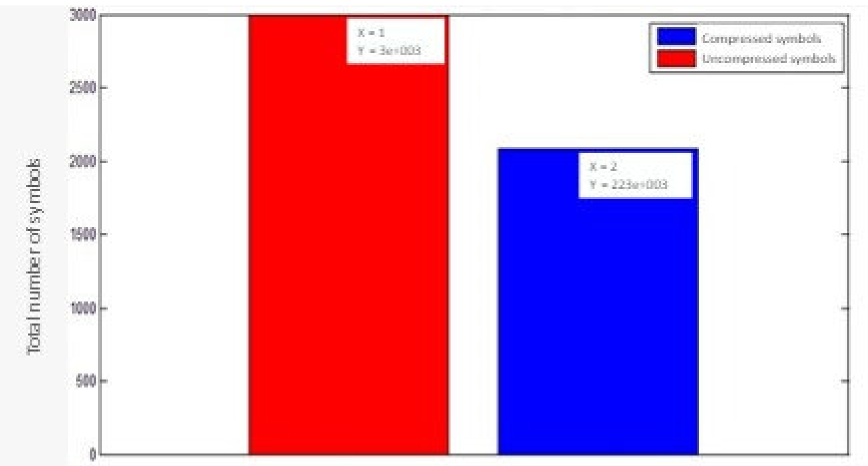

Fig. 1. Results obtained from data compression using the RLE algorithm

A. EXPERIMENTAL VALIDATION

Practical tests were performed on the algorithm using as input data the first 1000 data from file 4.

The tests conducted on the algorithm consisted of the transmission of previously compressed data over a wire channel, the data were stored in a plain text file, and subsequent calculations were made on the information collected.

The calculations performed are the number of uncompressed symbols, the number of compressed symbols, the data compression ratio, the BER, and finally the data compression percentage, resulting in the following results:

The compression ratio obtained for the algorithm is presented in, in order to compare the number of symbols without compression with the number of symbols after compression.

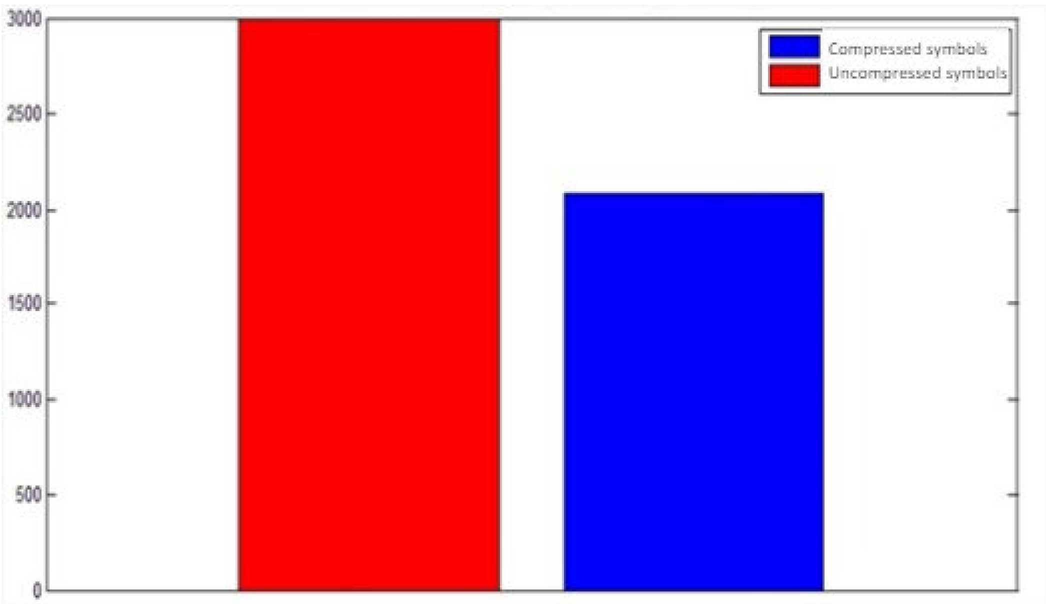

Fig. 2. Results obtained from experiments performed using the RLE algorith

The values obtained to determine the most efficient algorithm are shown in chart 3.

Chart 3 – Numerical results obtained from the simulation of the RLE compression algorithm

Number of uncompressed symbols	3000 symbols
Number of compressed symbols	2226 symbols
Data compression ratio	0.7420
Data compression percentage	25.8000%
BER	0 bits

IV. RESULTS

In order to determine the efficiency of the algorithm, a quantitative evaluation of the results obtained in the simulations as well as in the experiments performed was conducted.

Chart 4 – Practical results obtained

	Information	Results	
Criteria	Uncompressed	Simulation	Practical
Compression Ratio	0	0.7420	0.7410
Compression Time	15 min	30 sec	30 sec
Entropy	4.2146	4	4
BER	0 bits	0 bits	0 bits

It is observed that the compression achieved is 26%, the compression time is reduced significantly in consideration of the time needed for the transmission of uncompressed data, it is observed that the values of uncompressed entropy and post-compression entropy is reduced by 25% which translates into increased use of the channel.

Finally, in the Bit Error Rate (BER) criteria, it is observed that no error was found in the experiments performed because a speed of 9600 bits/second and using a twisted pair cable is obtained robustness in terms of Gaussian noise and 1/f noise.

According to these results, it can be concluded that the implementation of the RLE compression algorithm favours the transmission of information coming from the three-dimensional scanning system, because it reduces the transmission time of the information, the transmission channel is used efficiently and there is no loss of information due to the compression effect or the transmission channel effect.

V. CONCLUSIONS

During the development of this work, a methodology was implemented to evaluate the performance of the RLE compression algorithm, which is characterised by belonging to the lossless compression algorithms. The parameters that were considered for this purpose make it possible to determine efficiency in a quantifiable manner since the results can be measured and verified through experimentation.

The information that was considered for the development of the work comes from a mechatronic scanning system whose purpose is a three-dimensional reconstruction of solid elements, with a precision of the mechanical type, the system delivers digital samples of 10 bits, where each sample represents a pixel of an image of dimensions of 10000 pixels per side, because of this it was required the design of a specialized dictionary for the type of information.

The importance of the work lies in the proposal of a methodology that allows the study of compression algorithms at a deeper level due to the fact that not only is its design and implementation studied, but the efficiency is determined from criteria that allow evaluating its performance in more complex communications system, allowing to obtain specific information of the use of the channel, the number of bits lost during the transmission of the information and the time required for the compression of the information.

This information can be used to improve future design techniques in digital communications systems, starting from an approach not only functional but considering aspects of resource optimization.

VI. References

[1] M. D., Information Theory, Inference, and Learning Algorithms, Cambridge University Press, 2004.

[2] D. A Gudeto Salomon, Data Compression Methods, New York: Springer Verlag, 2002.

[3] S. HAYKIN, Digital Communication, John Wiley Sons, 1988.

[4] S. Haykin, Communication Systems, McMaster University: John Wiley Sons, 2001.

[5] B. Sklar, Digital Communications, Fundamentals and Communications, Prentice Hall, 2001.

[6] J. Proakis and M. Salehir, Communication Systems Engineering, Upper Saddle River, New Jersey 07458: Prentice Hall, 2002.

[7] A. S. William and . A. Pearlman, Digital Signal Compression, Cambridge University Press, 2011.

[8] D. Salomon, A guide to Data Compression Methods, Springer, 2002.

[9] B. M. a. J. Abrahams, "On the redundancy of optimal binary prefix-condition codes for finite and infinite sources," *IEEE Transactions on Information Theory,* vol. 33, no. 1, pp. 156-160, Jan. 1987.

[10] B. M. J. E., Transmisión de Datos, ULA: Taller de Publicaciones de la Facultad de Ingeniería, 2005.

[11] J. Fowler and . R. Yagel, Lossless Compression of Volume Data, Symposium on Volume Visualization, 2007.

[12] K. S. N. M. J. D. a. P. J. M. T. Asif, "Near-Lossless Compression for Large Traffic Networks," *IEEE Transactions on Intelligent Transportation Systems,* vol. vol. 16, no. no. 4, pp. pp. 1817-1826, Aug. 2015.

[13] P. D. Symes, Video Compression, McGraw-Hill, 1998.

[14] S. B., C. A. Bibyk Kenneth and M. J. Shalkhausert, Digital Compression Algorithmsf For Hdtv Transmission, The Ohio State University Columbus: Dept. of Electrical Engineering.

[15] Y. W. W. X. a. Q. D. J. Yang, "Image Coding Using Dual-Tree Discrete Wavelet Transform," *IEEE Transactions on Image Processing,* vol. 17, no. 9, pp. 1555-1569, Sept. 2008..

[16] E. h. Y. a. L. Wang, "Joint Optimization of Run-Length Coding, Huffman Coding, and Quantization Table With Complete Baseline JPEG Decoder Compatibility," *IEEE Transactions on Image Processing,* vol. 18, no. 1, pp. 63-74, 2009.

[17] J. M. H. Madrid, Caracterización de Algortimos de Compresión de Datos en la Comunicación de un Sistema de Detección de Defectos, Puebla, México, 2014.

[18] B. J. a. W. P. Burleson, "Real-time VLSI compression for high-speed wireless local area networks," *Data Compression Conference,* pp. pp. 431-., 1995.

[19] D. W. Jones, "Practical evaluation of a data compression algorithm," in *Data Compression Conference,*, Snowbird, UT, 1991.

АНАЛИЗ ЭФФЕКТИВНОСТИ СЖАТИЯ ДАННЫХ В ТРЕХМЕРНОЙ СИСТЕМЕ СКАНИРОВАНИЯ С ИСПОЛЬЗОВАНИЕМ RLE-АЛГОРИТМА

Мигель Хуртадо Мадрид, Грегорио Тринидад Гарсия,
Хосе Итало Кортез, Хосе Луис Амека, Фернанда Мерло Симони,
Сезар Антонио Агилар Родригес, Анна Алексеева, Наталья Сосницкая

Аннотация **– методы, используемые для сжатия данных, обеспечивают большую передачу информации благодаря выполненной обработке. С помощью числовых и статистических преобразований возможно представить ту же информацию в меньшем пространстве. Одна из основных проблем, которые существуют в системах связи, заключается в использовании канала. Текущие исследования сосредоточены на изучении характеристик и эффектов шума, присутствующего в канале, и помех, создаваемых в передаваемой информации. В этой статье изучается эффективность алгоритма сжатия Run Length Encoding (RLE) в проводном канале связи. Измеряя некоторые параметры, можно определить эффективность использования канала, используя алгоритм RLE. Методология используется для исследования, которое позволяет определить эффективность использования канала, до использования алгоритма сжатия и после. Таким образом можно определить эффективность использования канала. В этой работе наблюдается увеличение в использовании канала с помощью алгоритма сжатия RLE.**

АНАЛІЗ ЕФЕКТИВНОСТІ СТИСНЕННЯ ДАНИХ У ТРИВИМІРНІЙ СИСТЕМІ СКАНУВАННЯ З ВИКОРИСТАННЯМ RLE-АЛГОРИТМУ

Мігель Хуртадо Мадрид, Грегоріо Тринідад Гарсія, Хосе Італо Кортез, Хосе Луїс Амека, Фернанда Мерло Сімоні, Сезар Антоніо Агілар, Ганна Алєксєєва, Наталія Сосницька,

***Анотація* – методи, що використовуються для стиснення даних, дозволяють здійснювати більшу передачу інформації за рахунок обробки, що здійснюється за допомогою числових і статистичних перетворень. Можна представити одну і ту ж інформацію в меншому просторі. Одна з основних проблем, що існують у системах зв'язку, полягає у використанні каналу. Поточні дослідження спрямовані на вивчення характеристик і ефектів шуму, присутнього в каналі, і порушень, що генеруються в переданній інформації. В даній роботі досліджено ефективність алгоритму стиснення Run Length Encoding (RLE) у проводному каналі зв'язку. Шляхом вимірювання деяких параметрів можна визначити ефективність використання каналу, використовуючи алгоритм RLE. Дослідження, яке дозволяє визначити ефективність використання каналу дає використання алгоритму стиснення. Таким чином, можна визначити ефективність використання каналу завдяки алгоритму стиснення RLE.**

На підставі проведених досліджень були зроблені наступні висновки:

Під час розробки даної роботи була впроваджена методика для оцінки працездатності алгоритму стиснення RLE, яка характеризується належністю до алгоритмів стиснення без втрат.

Інформація, яка враховувалась при розробці роботи, надходить із системи мехатронного сканування, метою якої є тривимірна реконструкція твердих елементів, з точністю механічного типу система доставляє цифрові зразки 10 біт, де кожен зразок представляє пікселя зображення розмірами 10000 пікселів на сторону, через це було потрібно розробка спеціалізованого словника для типу інформації.

Важливість роботи полягає в пропозиції методики, яка дозволяє вивчати алгоритми стиснення на більш глибокому рівні через те, що не тільки вивчається його розробка та реалізація, але ефективність визначається за критеріями, що дозволяють оцінити її ефективність у більш складну систему зв'язку, що дозволяє отримати конкретну інформацію про використання каналу, кількість біт, втрачених під час передачі інформації, і час, необхідний для стиснення інформації.