УДК 378.011.3-051-057.21:377:004

Jose Italo Cortez,
Gregorio Trinidad Garcia,
Pedro Garcia Juarez,
Jose Miguel Hurtado Madrid
(Autonomous University of Puebla, Mexico)
Ganna Alieksieieva
(Berdiansk State Pedagogical University)

A NOVEL DATA COMPRESSION ALGORITHM FOR A THREE-DIMENSIONAL RECONSTRUCTION SYSTEM

Анотація

У статті представлені певні спільні результати роботи викладачів Автономного Університету Пуебла (Мексика) та Бердянського державного педагогічного університету (Україна) стосовно нового алгоритму стиснення даних для системи тривимірної реконструкції. Важливість алгоритмів стиснення в системах передачі інформації пов'язана з необхідністю більш високої швидкості передачі інформації за допомогою невеликої кількості символів або бітів. У цій роботі представлений новий алгоритм стиснення даних для передачі даних в системі тривимірної реконструкції. Алгоритм Хаффмана виконується з метою досягнення найкращого ступеню стиснення та мінімальної кількості втрачених біт для системи тривимірної реконструкції. Ці параметри дозволяють вимірювати ефективність алгоритму, через що необхідно проводити симуляції, експериментальні випробування та вивчати результати, визначати ефективність. Беручи до уваги цю інформацію, результати цієї роботи дозволять запропонувати алгоритми та оптимальні архітектурні рішення для типу інформації, що використовується в цій програмі.

Ключові слова: вища освіта, засоби комп'ютерної техніки, інформаційні технології.

Summary

The importance of compression algorithms in data transfer systems is due to the need for a higher rate of information transfer using a small number of symbols or bits. In this work, a new data compression algorithm is presented, in order to transfer data with a three-dimensional reconstruction system. A Huffman algorithm is performed, with the purpose of achieving the best compression ratio and the minimum number of lost bits for a three-dimensional reconstruction system. These parameters allow to measure the efficiency of the algorithm, because of this it is necessary to perform simulations, experimental tests and by studying the results is determined the efficiency considering this information, the results of this work will allow proposing algorithms and optimal architectures for the type of information used in this application.

Key words: higher education, computer engineering means, information technologies.

Introduction. The basic principle of compression algorithms is to encode data in a communication system for the transmission in a communication system, this encoding is done by developing mathematical models and statistical processes applied to the data to be compressed [1] [2].

There are two types of compression algorithms, algorithms for lossless compression and lossy compression algorithms, the difference between these two types of compression lies in the type of processing that is given to the information.

The lossless compression algorithm is one of the most widely used because of the advantages presented in contrast to lossy compression algorithms. One of the main advantages is the integrity of the information during the coding process and also no present implementation complexity, this is due to their architectures being based on mathematical models previously demonstrated.

Currently the study that is performed on compression algorithms provides applications in various areas of knowledge: medical sciences, communications systems, image processing and others.

In the area of communications systems a work was performed [3] where the comparison of different techniques is proposed to store code prefixes Huffman algorithm, where the problem of space required memory for storing codes arises prefixes for this algorithm, the proposal raised in this paper are various techniques for optimizing prefix codes that reduce the memory space required for storage.

The importance of this work lies in the analysis of compression algorithms, specifically the Huffman algorithm. The results of this study show an optimization of space required for storing prefix codes, however, this reduction leads to an increase in the length of code required to implement compression, which directly influences the compression factor reducing its efficiency, this article highlights the importance of design in the dictionary as this directly influences the compression ratio of the algorithm, this paper presents a study focusing on the efficiency of the algorithm where a canonical architecture is contemplated with the design of a dictionary because the data required to compress have certain statistical characteristics.

In the area of medical sciences a study was done [4] where a compression technique proposed in real time based on the coupling of a delta algorithm second order and the compression algorithm Huffman, this technique will be applied in the plethysmographs, with the purpose of making medical diagnoses concerning the circulatory system, cardia functions based on the study of change in blood volume.

Contributions concerning this work are observed in the implementation of new compression techniques using coupling various techniques for improving the transmission rate information.

This paper reports a low rate of compression and justified mainly due to the presence of motion artifacts and non-uniformity of data. This work resumes the need for a study.

This article describes and characterizes a Huffman algorithm, which is the development of a methodology proposed for the study of the efficiency. To validate this methodology, the algorithm is applied in the three-dimensional reconstruction since it is an application where a transfer of information is required and necessary compression for performance improvement in data transmission in a wired channel, also due the type of data calculating required arithmetic mean, variance, plus the probability of each element, for the design and construction of a dictionary for this information.

In section II of this article describes the main characteristics of a compression algorithm and presented a Huffman compression algorithm model. In section III the behavior of the transfer of information of a three-dimensional system is studied and the importance of a Huffman compression algorithm is shown, also the main parameters of an algorithm for this application are defined, moreover, it defined a methodology for determining the efficiency of a Huffman compression algorithm. In section IV a dictionary for Huffman compression algorithm is developed. Subsequently, a study of the efficiency of a compression algorithm through simulation and experimental part is done, in section V a discussion of the results is made and the conclusions of the study are set out in Section VI.

II. Huffman compression algorithm

Compression algorithm

Data compression is to process a frame of symbols and by transformations represent the same message using different codes, in order to store the signal in a smaller space, some of the applications are in digital communications, in computer systems is used to reduce data volume [1] [3].

The resolution is imposed by the digital system with which it works and cannot alter the number of bits arbitrarily; therefore, compression is used to transmit the same amount of information that would occupy a large resolution on a lower number of bits, compression is a particular case of coding [2] [5].

The probability of transmission error is reduced, increasing system benefits, in addition to the ability to recover the original signal based on the processed signal without loss of information.

A. Huffman algorithm and its importance

The compression algorithm Huffman proposes a method of encoding a set of messages consisting of a finite number of symbols, in order to obtain a code of minimum redundancy which is constructed such that the average number of digits coding message is minimized [1] [6].

The importance of Huffman compression algorithm is observed in various applications like intercom information systems, medical diagnosis, and characterization of sensors, among others. In these applications, the compression of information is of great importance since it allows the transfer of information in real time [7] [9].

Huffman compression algorithm is defined as a lossless compression without loss of the statistical type, that is, performs data compression by using a dictionary replaces redundant information coded by variable-length binary codes. This dictionary is developed through a preliminary statistical analysis that identifies redundancy, entropy and probabilities of symbols in the information, further definitions and statistics restrictions are contemplated, which must ensure the validity of the analysis and therefore the dictionary. In Theorem 1, the statistical properties of the information to develop the dictionary [7] [8] [11].

Theorem 1. For an alphabet-source A, whose source-symbols are associated with a probability distribution P, there is a minimum binary code redundancy and free prefix, C, which satisfies the following properties:

i If $p_j > p_i \; then \; |c_j| \leq |c_i|$

ii Code words that correspond to the two least probable symbols have the same length.

iii the two longer code words are identical except for the last digit.

It is expressed mathematically by (Eq. 1). This expression determines the order of probability depending on the element, this allows data to be sorted according to their probability to determine the dictionary for this type of information [9].

$$\begin{cases} p(x_k) & |si| \; k \leq n-2 \\ p(x_{n-1} + p(x_n)) & |si| \; k = n-1 \end{cases} \qquad \text{Eq. 1}$$

To make the dictionary takes a finite sequence α_M^ℓ of length M, with real integer values.

$$\alpha_M^\ell = \{\alpha_{M,i}^\ell\} = (\alpha_{M,0}^\ell, \ldots, \alpha_{M,i}^\ell, \ldots, \alpha_{M,M-1}^\ell), \alpha_{M,i}^\ell \in \mathbb{R} \qquad \text{Eq. 2}$$

Where $\alpha_{M,i}^\ell = 0, i < 0, i > M-1$, i denotes the order of the variable, l is the number of domain sequence of real numbers R. A real value sequences of finite length, α_M^ℓ can be written as:

$$A = \{\alpha_M^0, \ldots, \alpha_M^\ell, \ldots, \alpha_M^{L-1}\} \qquad \text{Eq. 3}$$

$$= \{\{\alpha_{M,i}^0\}, \ldots, \{\alpha_{M,i}^\ell\}, \ldots, \{\alpha_{M,i}^{L-1}\}\} \qquad \text{Eq. 4}$$

Where L denotes the sequence number, that is, the family size, function aperiodic correlation between the sequences α_M^ℓ and $\alpha_M^{\ell'}$ with a shift in i' is defined as:

$$\rho_{M,i'}^{\ell,\ell'} = \frac{1}{M}\sum_{i=0}^{M-1} \alpha_{M,i}^\ell \alpha_{M,i-i'}^{\ell'} \qquad \text{Eq. 5}$$

Where $\rho_{M,i'}^{\ell,\ell'}$ for $|i'|$ > M-1, if the aperiodic autocorrelation function satisfies:

$$\rho_{M,i'}^{\ell,\ell'} = \begin{cases} 1 & ; i' = 0 \\ 0 & ; i' \neq 0, \pm(M-1) \\ \epsilon_{M-1} & ; i' \pm (M-1) \end{cases} \qquad \text{Eq. 6}$$

Where ϵ_{M-1} is called offset value and $|\epsilon_{M-1}| \leq 1/2$, then the sequence α_M^ℓ sequence is called Huffman for actual values.

$$\alpha_{M,i}^\ell = K(\alpha_{2,i}^{\ell_1} \otimes b_{2,i}^{\ell'_2} \otimes b_{3,i}^{\ell'_3} \otimes \ldots \otimes b_{m,i}^{\ell'_m} \otimes b_{\frac{M+1}{2}}^{\ell' \frac{M+1}{2}}) \qquad \text{Eq. 7}$$

All individual symbols are considered groups, obviously one size and the probability of the group as the sum of the probabilities of its members.

The two groups least likely merge into one and a bit selector preceding the code words involved two symbols is inserted: Bit 0 for codes of the symbols of a group and bit 1 for the other.

Following this fusion the K (C) factor and the number of groups are reduced by one unit.

The previous step is repeated using the modified groups, until only one remains, in this case, K (C) = 1 which became a free prefix code and will be created code for each symbol of the alphabet, according to SOA Huffman architecture is determined by the flowchart of FIG. 1.

```
Process Huffman Coding
            Divide the probabilities of each element in decreasing order
loop:       Combine the two lowest odds
            if number of elements = 2 then
                goto coding
                else
                goto loop
            end if
coding:     Assign zero and one to the two codewords
loop2:      If Is any element the result of combining two elements? Then
                goto EndProcess
                else
                Complete the code word with zero or one
                goto loop2
            End If
EndProcess
```

Fig. 1 Huffman Coding implementation

III. Study of a Huffman algorithm and methodology for a three-dimensional system.

A. DIMENSIONAL SYSTEM AND THE USE OF HUFFMAN ALGORITHM

The system that we have is trestle type robot three degrees of freedom Fig. 2 and is characterized by a displacement axis of a micrometer, which is achieved with the use of a gearbox with a ratio of 4: 1 [10].

For displacement the robot in the cartesian axes various types of motors are used; in the Y axis two DC motors are connected; for movement in the Z axis is has a stepper motor; with the X axis it was used a high speed DC motor for a shorter displacement. The required accuracy is obtained by an analog-digital converter (ADC) with a sampling frequency, depending on the number of data required, the distance required for the scanning and the number of revolutions per minute of the DC motor employed [2].

The total time required for scanning is determined in 7 seconds using 3 axes, the acquisition and transmission of samples to the PC, this time was reduced by modifying the algorithm to correct mechanical effects present in small displacements, achieving a linear scanning speed and greater.

The disadvantage of this modification is reflected in the number of samples obtained at an instant of time. The results obtained in scans showed a mean of 15,000 samples with a length of 10 bits in a time interval between scanning 0.3 seconds. This volume information is transmitted via USB protocol and due to the large amount of data is insufficient protocol after 5 continuous scanning system [6].

This problem leads to the need of implementing a system for reducing the volume of data to be transmitted by using compression algorithms, without loss of information resulting from compression.

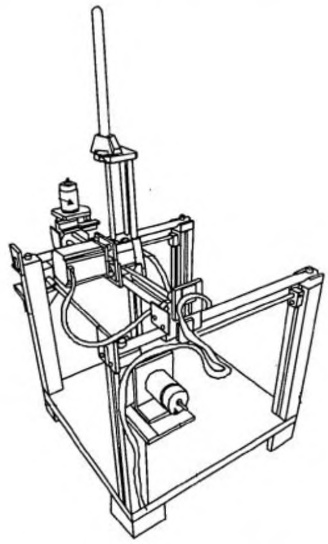

Fig. 2 Robot type trestle used for validation of the Huffman algorithm.

The use of compression algorithms presents a solution to this problem, without altering the information where mathematical transformations and statistical analyzes that allow the replacement and reduction of symbols in the elements of information are used.

To use these algorithms effectively requires a study to determine the algorithm with the highest efficiency, this requires measuring various parameters necessary for evaluating an algorithm called evaluation criteria and are performed based on the study of the nature of the information to be compressed and the type of algorithm used, the results of this evaluation will determine the efficiency for the type of information in a wireless transmission channel.

B. CRITERIA FOR COMPRESSION ALGORITHM

One criterion is the compression ratio, which is determined by mathematical relationships the reduction ratio of the information obtained in the algorithm, this criterion is essential to measure saving space required for the representation of data, if the is greater than 1 is called expansion, and occurs for two reasons: the selected algorithm is not adequate or not a correct implementation was done, or is a different type of information to the implementation performed on the algorithm, because the data don't have the same characteristics in images, text or audio [14] [15].

The compression ratio is expressed by the mathematical relationship (Eq. 8), which is expressed in terms of a scalar of the own information [6].

$$Ratio = \frac{data\,compressed}{data\,uncompress} \qquad \textit{Eq. 8}$$

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

In image compression, the same term, bpb means bits per pixel. Consequently, the main objective of data compression is to represent any data at low bit rates, the term bit cost refers to the role of the individual bits in the compressed string.

Another parameter that identifies the efficiency of a compression algorithm is the Bit-Error-Rate (BER) relation, which measures the number of bits that are lost in transmission of a digital system as a result of channel noise that is used for transmission of information [11].

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

One disadvantage is the abrupt degradation of the digital information recovered from the decline of the signal strength or increase in noise or interference, when the received signal is contaminated with Gaussian noise, the rate of binary error can be determined analytically without coding systems, when there are further interference or encoding for error correction, the process of analytical determination of the error rate is a complex problem, for this reason, is often used computer simulations to evaluate the performance of systems transmission.

The transmission channel is used as the means by which signals are transmitted power, the relationship between the use of the transmission channel and a compression algorithm is measured by entropy. This parameter determines how effective is the use of the channel used by an algorithm [11].

C. METHODOLOGY FOR EFFICIENT HUFFMAN ALGORITHM

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

• Development of simulations based on the data to be compressed.

• Measurement criteria evaluation of the results obtained in the simulations.

• Development of experiments by developing a system firmware level.

• Analysis of the results obtained for the formulation of proposals for optimization.

IV Development of methodology and experimental validation

The methodology development is performed using a dictionary that is a database in which each possible value of information is stored, the relative probability and a unique binary code substitution for each value [2].

The code is obtained by drawing a binary tree, then a grouping of the two least likely probabilities, assigning the code 1 to the highest probability and 0 the lowest, and both values are summed. This process is repeated with all possible information values, the resulting binary codes are of variable length for each value of information.

After obtaining binary codes, the implementation of the algorithm is performed using the canonical architecture [1], which contemplates the use of the dictionary for processing the data to be compressed.

The development of the methodology for the study of the efficiency of the algorithm aims to develop practical simulations and experiments that will provide the necessary information to determine quantitatively efficiency.

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

The data used for performing these simulations as for practical tests was performed on the basis of a sample file, the file selection was made by a statistical study that best represents the entire universe of the study whose data represent the samples in a format of integers numbers with no signs with a length of 10 bits.

Arithmetic means of the files and the overall mean were compared, and the file selected was whose arithmetic value measure was the closer to the overall average, with this ensures that the results will be valid for the entire sample. In Table 1 the results of this study are shown.

Table 1. Numerical results obtained from statistical files of scanned samples analysis.

Overall mean samples	394.6791
Measure of file 4	394.3826

Simulation Huffman compression algorithm type, consisted in the coding the flowchart, see fig.1, the tests applied to this algorithm settled in determining the parameters including number of symbols uncompressed, number of symbols compressed, ratio of data compression and compressed data rate, where the following results were obtained:

The compression ratio obtained for this algorithm is shown in Fig. 3, in order to compare the number of symbols without compression by a number of symbols after compression, it is noted that the total number of symbols subsequent to the compression it is less that the total symbols to be transmitted without compression by a factor of 30.6%, the numerical results are shown in Table 2 and used in the study of the efficiency of the algorithm.

Table 2. Results obtained from the simulation Huffman compression algorithm

Number of symbols uncompressed	3000 symbols
Number of symbols compressed	2082 symbols
Compression ratio data	0.6940
Percentage of compressed data	30.600%
BER	0 Bits

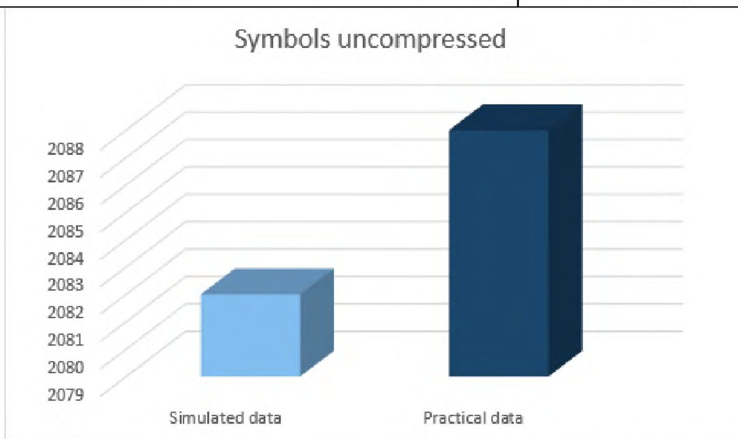

Fig. 3. Data without compression and data compression by Huffman algorithm

A. EXPERIMENTAL VALIDATION

Practical tests on the algorithm using as input the first 1000 data file 4 were carried out, the tests were performed at this algorithm consisted in the transmission of previously compressed data over a wireless channel, data is stored in a plain text file, and then calculations on the data collected were made, calculations are the number of symbols uncompressed, the number of symbols compressed, and the compression ratio of the data, the BER, and finally the compressed data rate, with the following results:

The relation of compression ratio obtained for this algorithm is shown in Fig. 4, in order to compare the number of symbols uncompressed by the quantity of symbols after the compression.

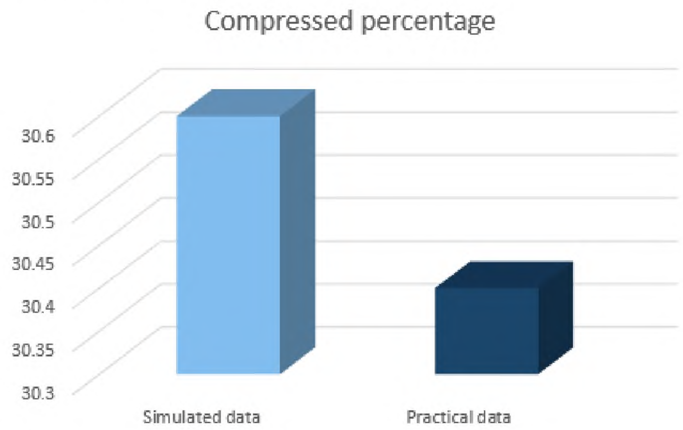

Fig. 4 Comparison between information without compression and information using the Huffman algorithm

The values obtained to determine the more efficiently the algorithm shown in Table 3.

Table 3. Numerical results obtained by the simulation of Huffman compression algorithm

Number of symbols uncompressed	3000 Symbols
Number of symbols compressed	2088 Symbols
Compression ratio data	0.6960
Percentage of compressed data	30.40%
BER	0 Bits

V Discussion of Results

To determine the efficiency of the algorithm, it was performed a quantitative evaluation of the results obtained in both simulations and experiments performed, see Table 4.

Table 4. Practical results obtained

	Information	Huffman Algorithm	
Test	**Without compression**	**Simulation**	**Experiments**
Compression ratio	0	0.694	0.69
Time of	15 min	1.5 min	1.5 min

Compression			
Entropy	4.2146	5.6	5.6
BER	0 Bits	0 Bits	0 Bits

It is noted that the compression achieved is 30%, the compression time is reduced in a significant way in consideration of the time required for transmission of uncompressed data, also shows that the values of subsequent entropy uncompressed and after the compression increased by 25% which results in improved channel use.

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

According to these results it can be concluded that the implementation of the compression algorithm Huffman favors the transmission of information from the three-dimensional scanning, because it reduces the transmission time of the information, the transmission channel is used in an efficient manner and there is no loss of information due to the compression or effect of the transmission channel.

Conclusions. This paper proposes a methodology for evaluating the performance of a compression algorithm, it is characterized by belonging to the algorithms of lossless compression, the parameters that were considered for this purpose, allow to determine the efficiency of a quantifiable way, because the results can be measured and verified by experiment.

The information considered for the development of this work comes from a system of mechatronic scanning whose purpose is the three-dimensional reconstruction of solid elements with a mechanical type precision, this system delivers digital samples of 10 bits, where each sample represents a pixel an image size of 10,000 pixels per side, for this reason that the design required a specialized dictionary such information.

According to the results, the methodology can be implemented in other demanding applications because it has been shown that it can be obtained compression algorithms more efficient, besides the information presented in this work, can be used as a tool to obtain better designs in the transfer and processing of data. Finally, future work can be recommended as resource optimization methodology.

The importance of this work is the proposal of a methodology for the study of compression algorithms to a deeper level because not only is studied the design and implementation, but the efficiency is determined from criteria for evaluating its performance in a complex communications system, allowing specific information on the use of compression algorithms. This information can be used to improve future design techniques in different applications, based on an approach not only functional but considering aspects of resource optimization.

REFERENCES

1. A. Huffman, «A Method for the Construction of Minimum-Redundancy Codes,» *Proceedings of the IRE,* vol. vol. 40, nº no. 9, pp. 1098-1101, Sept. 1952.

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

3. G. N. Y. N. a. A. O. T. Gagie, «Efficient and Compact Representations of Prefix Codes,,» *IEEE Transactions on Information Theory,* vol. 61, nº 9, pp. 4999-5011, Sept. 2015
4. R. Gupta, «Lossless Compression Technique for Real-Time Photoplethysmographic Measurements,» *IEEE Transactions on Instrumentation and Measurement,* vol. 64, nº 4, pp. 975-983, April 2015.
5. T. M. Cover, Elements of Information Theory, Segunda Edición ed., New Jersey: Wiley, 2006.
6. J. M. H. Madrid, Instrumentación y Automatización de un Sistema de Detección de Defectos Dentales, Puebla, México, 2012.
7. K. Y. L. S. Y. L. a. B. J. Q. Huo, «Compressed Relaying for Two-Way Relay Networks With Correlated Sources,» *IEEE Wireless Communications Letters,* vol. 4, nº 1, pp. 30-33, Feb. 2015.
8. 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. 1817-1826, Aug. 2015.
9. B. J. a. W. P. Burleson, «Real-time VLSI compression for high-speed wireless local area networks,» *Data Compression Conference,* pp. 431-442, 1995.
10. H. T. Y. I. a. S. M. T. Matsumoto, «Study on BER performance of SS system using a Huffman sequence against CW jamming over AWGN channels,» de *Information Theory and its Applications (ISITA)*, Melbourne, 2014.
11. D. L. a. W. Jia, «Optimal maximal and maximal prefix codes equivalent to Huffman codes,» *Acoustics, Speech, and Signal Processing (ICASSP),* vol. III, pp. 2525-2528., 2002.
12. 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.
13. 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.
14. 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.
15. D. L. a. W. Jia, «Optimal maximal encoding different from Huffman encoding,» de *Proceedings. International Conference*, Las Vegas, NV, 2001.
16. D. W. Jones, «Practical evaluation of a data compression algorithm,» de *Data Compression Conference,*, Snowbird, UT, 1991.