A CHINESE REMAINDER THEOREM BASED ENHANCEMENTS OF LEMPEL-ZIV-WELCH AND HUFFMAN CODING IMAGE COMPRESSION

;

ABSTRACT

Data size minimization is the focus of data compression procedures by altering representation and reducing redundancy of information to a more effective kind. In general, lossless approach is favoured by a number of compression methods for the purpose of maintaining the content integrity of the file afterwards. The benefits of compression include saving storage space, speed up of data transmission and high quality of data. This paper observes the effectiveness of Chinese Remainder Theorem (CRT) enhancement in the implementation of Lempel-Ziv-Welch (LZW) and Huffman coding algorithms for the purpose of compressing large size images. Ten images of Yale database was used for testing. The outcomes revealed that CRT-LZW compression saved more space and speedy compression (or redundancy removal) of original images to CRT-Huffman coding. Similarly, CRT-Huffman coding (28.13) offered better quality Peak-Signal-to-Noise-Ratio (PSNR) for the reconstructed images when compared to CRT-LZW (3.54) and (25.59) obtained in other investigated paper.

Keywords: LZW, Huffman coding, CRT, compression time, size reduction, image, compression.

1. INTRODUCTION

Data compression is the method of decreasing the size of information to be transmitted or stored by the process of eliminating redundancy in information without the loss or the ability to reconstruct the original data. There are several file formats that can be effectively compressed including text, image, video, and audio [1]. Image compression addresses the difficulty involved in decreasing the volume of data vital in denoting an image with no major loss of information. In recent times, Chinese Remainder Theorem (CRT) has developed in fields of coding theory, phase unwrapping, frequency estimation and distributed data storage. This is certainly due to the fact; CRT is profound for isolating errors in residue caused by noise. The traditional CRT reconstructs a single integer for error-free co-prime and residues [2]. Images have limited applications in real life situations such as medical, scientific, prepress and artistic applications, due to enormous sizes for broadcast or storage given low bandwidth [3]. In dealing with memory capacity insufficiency, compression schemes have been deployed; thereby offering the prospect of broadcasting images/video under scare bandwidth. The classical image compression scheme converts a spatial domain representation to frequency domain [3].

In classical JPEG codec, images are encoded independently. The surge in cloud data storage has thrown up issues of content duplication and large redundancy, which must be considered. Inter-coding is one of such solution for traditional video encoding of consecutive frames from previous frames. Another method is the inter-prediction tools of video codecs to encode comparable images as pseudo-video arrangements [4].

Again, lossy compression algorithms such as Huffman coding gives relatively good quality as well as compression rates with images but blocky look of reconstructed images [5]. The reverse is the case of LZW in which compressed image data quality is retained at the expense of little size decreases [6]. In this paper, CRT enhancement is proposed for independent implementation of Huffman coding and LZW algorithm for image compression procedures.

2. RELATED STUDIES

The goal of image compression in variety of application is to decrease the quantity of bytes in a graphic file but retaining the quality. In the study, [3] considered several approaches for compressing images especially in medicine. The forms of compression for image involve spatial to frequency domains. The main concept is that pictures are composed of neighbouring pixels; though related not without repetitive data. However, colour images require treatment of distinct colour segments.

A novel prediction scheme for performing cloud-based compression was initiated by [4]. This approach utilizes the semi-local geometric and photometric prediction scheme to compensate in a region-wise style the distortion between two images rather than inter-coding schemes (such as high efficiency video coding). This is most useful for highly correlated image content applications such as traditional video coding, cloud gaming streaming, photo albums compression. This minimizes the redundancy arising from similar content already available in cloud. Nevertheless, cloud multiple frames exploitations, and determining scalability of cloud-based image compression system.

Lossless compression schemes of LZW and Huffman coding were combined by [6]. The target was to enable medical imagery suitable for storage, quality retention and broadcast. Huffman coding offered massive size decreases in a speedy manner, but, poor quality of compressed image. Conversely, LZW algorithm produced finest quality with little size decreases. The combined compression schemes gave rise to high compression ratio and high PSNR values.

A fresh algorithm was proposed by [7] for encoding, decoding or regenerating the replica of the encoded data. The first step uses the forward difference scheme on Huffman. Then, the values are regenerated into fixed length code representation with twos complement for further new probabilities computation along Huffman's algorithm. There are improvements in compression factors for the new algorithm against the traditional Huffman encoding. Residue Number System is introduced in data encryption and decryption with Shannon Fano compression scheme by [16]. The outcomes revealed significant improvements in the security, speed and memory needed for existing information communication networks.

Typically, multimedia files (such as image data) are composed of redundancy and irrelevance limiting their usage on widespread basis. Since the advent of internetworks and communication infrastructure, there are possibilities of broadcasting or storing digital images seamlessly [4]. The sizes of the multimedia data make them inefficient for broadcast or storage purpose [6]. Majority of lossless compression methods are founded on probability or dictionary and entropy because they make use of the availability of the identical string or character within data in order to realize compression [6]. Researchers are focusing attention on removing redundancy and irrelevance in image data, which gave rise to the concept of data compression schemes such as Huffman coding and Lempel-Ziv-Welch algorithm [6]. In general, the performance of compression schemes is estimated with standard metrics such as effectiveness (compression ratio) and efficiency (speedup or throughput) [7]. In this paper, these compression algorithms are highlighted in certain details.

One common entropy encoding algorithm deployed for lossless image compression is Huffman coding [8]. The encoding strategy commences with calculation of each symbol probability in the image. Thereafter, these symbols probabilities are placed in a descending magnitude as to create leaf nodes of a tree. By individually coding the symbols, the Huffman code is built combining the lowermost probable symbols. These entire steps are continued until only two probabilities of two compound symbols are present. Eventually, a code tree is produced and the labelling of the code tree generates the Huffman code [9].

The Huffman codes for the symbols are realised by analysing the branch digits in succession from the root node to the respective terminal node or leaf using symbols 0 and 1. Huffman coding is the most deployed method for redundancy or relevance minimization [8]. The operational principle of Huffman code is based on these observations:

a) The more frequently occurring symbols are assigned shorter code words than less frequent occurring ones;
b) The two symbols occurring least frequently is assigned the similar length.

On the average, code length is determined as the average of the product of symbol probability and amount of encoding bits [10] [11]. The Huffman code efficiency is calculated as the ratio of entropy to the average length. The target of Huffman encoding creates the optimal code for a collection of symbols and probabilities whenever, the symbols are coded currently within the same time frame [8].

LZW algorithm is a popular lossless data compression scheme initiated by Abraham Lempel, Jacob Ziv, and Terry Welch. In 1984, as an improvement over the traditional LZ78 algorithm released 1978 by Lempel and Ziv, which is easy to deploy with the prospect of offering significantly high throughput in hardware applications. According [12], the algorithm encodes sequences of 8-bit data as fixed-length 12-bit codes. The codes from 0 to 255 depict 1-character sequences composed of the matching 8-bit character, and the codes 256 through 4095 are created in a dictionary for sequences contained in the data during the process of encoding [1] [12] [13].

2.1 The Concept of Chinese Remainder Theorem (CRT)

The basic operation of Chinese Remainder Theorem (CRT) is to generate a single integer through its residue modulo within moduli set [2]. CRT is an alternative to the Mixed Radix Conversion (MRC) in which large modulo M derivations are unnecessary. MRC accepts a low complexity of 0(n) unlike the CRT having computation complexity of order 0(n3). In CRT, arithmetic operations for modulo *M* are to be manually executed. CRT residue converters are much more complex. In contrast, the MRC procedure requires arithmetic operations for modulo m_i only, thereby simplifying all operations. In MRC method, a number *x* is expressed in mixed-radix system. Suppose for moduli set $(m_1, m_2, ..., m_n)$, RNS representation of a number *x* is given by $(x_1, x_2, ..., x_n)$. The number *x* can be expressed in mixed-radix form as:

$$|x|_{m_1} = a_1$$

$$X = a_n \prod_{i=1}^{n-1} m_i + ... a_3 m_1 m_2 + a_2 m_1 + a_1 \quad (1)$$

where, the a_i's are the mixed-radix coefficients. These a_i's are determined sequentially, starting with a_1, in the following manner:

Equation (1) is first taken in modulo m_1. Since, all terms except the last are multiples of m_1 to give

$$|x|_{m_1} = a_1$$

Hereafter, a_1 is just the first residue digit. To obtain a_2, first subtract a_1 from *x*. The quantity $x - a_1$ is divided by m_1, and doing modulo operation with respect to m_2, we have

$$\left| \frac{|x - a_1|}{m_1} \right|_{m_2} = a_2$$

Similarly, for a_3, $(a_2m_1+a_1)$ is subtracted from *x*. By dividing the quantity $(x-a_2m_1-a_1)$ by m_1m_2 and performing modulo operation with respect to m_3, we get

$$\left| \frac{|x - a_2 m_1 - a_1|}{m_1 m_2} \right|_{m_2} = a_3$$

In this way, by successive subtraction and division in residue notation, all the mixed-radix digits may be obtained.

Conversely, an RNS number $(x_1, x_2, x_3 ....x_k)$ for the moduli set $(m_1, m_2, m_3 ....m_k)$ whose the decimal equivalent is given by:

$$a_1 = x_1 \quad (3)$$

$$a_2 = \left| (x_2 - a_1) \left| m_1^{-1} \right|_{m_2} \right|_{m_2} \quad (4)$$

$$a_3 = \left| ((x_3 - a_1) \left| m_1^{-1} \right|_{m_3} - a_2 \left| m_2^{-1} \right|_{m_3} \right|_{m_3} \quad (5)$$

Therefore, a general expression is given by:

$$a_n = \left|(((x_n - a_1)\left|m_1^{-1}\right|_{m_n} - a_2)\left|m_2^{-1}\right|_{m_n} - \dots - a_{n-1})\left|m_{n-1}^{-1}\right|_{m_n}\right|_{m_n} \quad (6)$$

The mixed radix digit (MRD) of a_i, $0 \le a_i < m_i$, any positive number in the interval by *[0,* $\prod_{i-1}^{k} m_i - 1)$ can be uniquely represented. The major advantage of MRC, as can be seen from Equation (5) above is that the calculation of a_i; i = 1, k can be done only using arithmetic mod-m_i contrasting CRT, which entails arithmetic mod-M, M being the system dynamic range, a rather large constant. It can be noted that equations (5) and (6) are directly utilized, only if the moduli set $\{m_1 m_2 m_3 \ldots m_k\}$ are relatively prime and that Euclidean algorithm is the common way to verify this, i.e., if gcd $(m_i, m_j) = 1$, for i ≠ j.

The residue independence, carry-free operation and parallelism attributes of the RNS have been intensively used in variety of areas, such as digital signal processing (DSP), digital filtering, digital communications, cryptography, error detection and correction [14] [15]. The addition, subtraction and multiplication are dominant. And, division, comparison, overflow and sign detection are negligible. One key field of RNS-based applications is finite impulse response (FIR) filters. Likewise, digital image processing benefits from the RNS's features such as enhancing digital image processing applications [15].

3. METHODOLOGY

In this image compression process, the implementation process was coded from scratch using MATLAB R2015a programming language. The paper studied the performances of traditional compression schemes of LZW and Huffman coding with CRT. The purpose of the employing CRT is to enhance their individual effectiveness using image media lossless compression technique. The arrangement of the planned enhancements of compression approaches is illustrated in Figure 1.

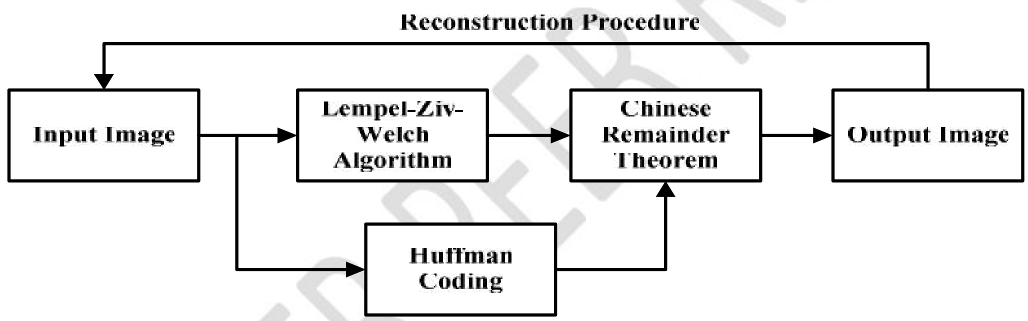

Fig. 1. The layout for the enhancement of compression schemes.

The input image is used to acquire the various formats of images for the complete data compression processes. These input images are composed of diverse degree of redundancy which is expected to be removed or minimized during planned compression processes. The data compression phase encompasses three distinct operations; firstly, the input image is received at Lempel-Ziv-Welch Algorithm block to commence the data compression. Similarly, the Huffman coding performs preliminary compression operation on the input images. Secondly, the complete compression of original image is achieved with CRT using the outcomes of LZW and Huffman coding schemes. Finally, the output image is realized from the last compression process of CRT computation, which enhanced the traditional image compressed format when compared to input images.

This paper considered four metrics in evaluating the effectiveness of the planned image compression schemes including:

1) *Compression Ratio* (CR) is expressed as the amount of uncompressed data size divided by compressed data size. This provides the relative size of compressed image data.
2) *Compression Time* (CT) calculates the time taken to compress bits in data in a second.
3) *Peak Signal-to-Noise Ratio* (PSNR) is used to estimate the amount of noise in the signals of compressed data relative to original data.
4) *Imperceptibility* calculates the rate of bits distributions of image data after complete compression. This infers on the appearance of compressed image data.

The operational algorithm of proposed LZW-CRT image compression scheme is presented below

Step 1- Extract first byte from input STRING.
Step 2- Extract the next byte from input CHARACTER.
Step 3- Lookup in table for the STRING and CHARACTER stored up.
Step 4- Generate code for the STRING and update the lookup table.
Step 5- Output STRING same as CHARACTER.
Step 6- STRING = STRING and CHARACTER.
Step 7- Apply CRT on the resulting STRING.
Step 8- The moduli set is chosen to obtain the best redundancy in data.
Step 9- The compressed image data is attained as final encoded values.
Step 10- The reconstructed image is obtained by applying decoding of LZW and CRT.
Step 11- Output is reconstructed image data.

Similarly, the steps for performing CRT-Huffman coding image compression scheme is presented in algorithm 2 below;

Step 1- INPUT original image
Step 2- Run Huffman coding functions
Step 3- Extract symbols of the pixels from input IMAGE.
Step 4-Create the probability of pixel symbols and organize in decreasing magnitude and smaller probabilities are combined.
Step 5- Concatenate the Huffman codeword ready for CRT
Step 6- Generate code for the STRING and update the lookup table.
Step 7- Apply CRT on the resulting STRING.
Step 8- The moduli set is chosen to obtain the best redundancy in data.
Step 9- The compressed image data is attained as final encoded values.
Step 10- The reconstructed image is obtained by applying decoding of LZW and CRT
Step 11- Output is reconstructed image data.

4. PRESENTATION OF RESULTS

The paper utilized 10 different image samples from Yale database for the purpose of validating the proposed concepts of CRT-LZW and CRT-Huffman coding compression. The outcomes of compression procedures on the sampled images using CRT and LZW are shown in Table 1.

Table 1. CRT-LZW based image compression

Image sample	Size before compression	Size after compression	Compression time	PSNR	Compression ratio
1	12282	3524	10.10	3.11	0.38
2	12906	3571	8.77	3.17	0.40
3	12353	3522	12.28	3.21	0.39
4	12762	3410	10.18	3.22	0.36
5	12872	3606	9.99	3.31	0.38
6	12357	3548	8.61	3.13	0.38
7	12150	4266	9.08	4.09	0.41
8	12243	3889	10.45	3.99	0.40
9	12530	3882	9.76	4.09	0.39
10	12319	3935	10.27	4.05	0.40
Total	124774	37153	99.49	35.37	3.89
Average	12477.4	3715.3	9.95	3.54	0.39

From Table 1, there is a significant decrease in the compressed images when compared to the original images at 12477.4 kb and 3715.3 kb on the average (that is 3:1) respectively. Consequently, 3 kb is used to represent 1 kb in the original image after performing the compression processes on the sampled images. In the same vain, the average values for reduced image size, compression time, PSNR and CR were 29.78%, 9.95 s, 3.54 and 0.39 respectively. The outlook of the compressed image is poor for Human Visual System (HVS) as depicted in Figure 2.

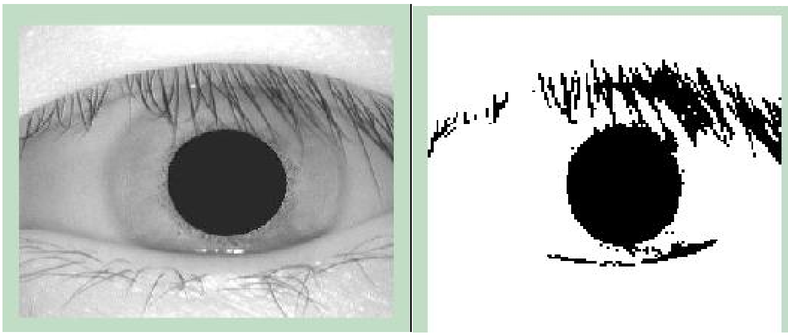

Fig. 2. Original image against CRT-LZW compressed image

In Figure 2, the image on the left hand side is the original image without compression operations. The image on the right hand side is realized after performing compression on the CRT and LZW. The compressed image looks washed out due to uneven distribution of bits composition to the HVS when matched with the reconstructed image. The outcomes of applying CRT to Huffman coding based compression using the sampled images are obtainable in Table 4.

Table 2. CRT-Huffman coding based image compression

Image sample	Size before compression	Size after compression	Compression time	PSNR	Compression ratio (%)
1	12282	1853	18.40	28.29	4.13
2	12906	1819	18.35	28.34	4.14
3	12353	1745	16.86	27.73	4.28
4	12762	1766	18.56	28.27	4.15
5	12872	1854	21.05	27.85	4.02
6	12357	1783	18.02	28.29	4.28
7	12150	1593	14.91	28.22	4.92
8	12243	1638	31.48	28.07	4.59
9	12530	1614	17.53	28.10	4.63
10	12319	1672	16.37	28.09	4.65
Total	124774	17337	191.53	281.25	43.79
Average	12477.4	1733.7	19.15	28.13	4.38

In Table 2, the data compression procedure of CRT and Huffman coding revealed substantial improvements in terms of the resultant image sizes, PSNR and compression ratio. On the average, CRT-Huffman based compressed scored 14.00%, 19.15 s, 28.13, and 4.38 for image size saved, time of compression, PSNR and CR. Again, the outlook of reconstructed image when compared to the original images is depicted in Figure 3.

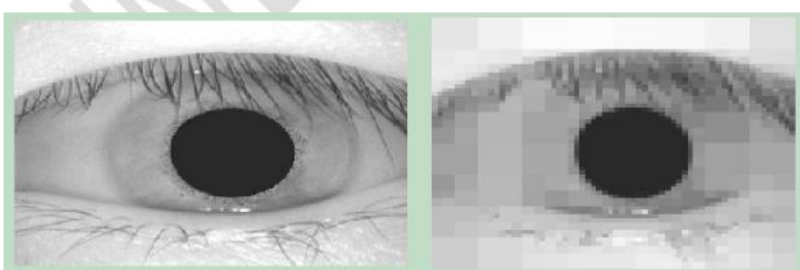

Fig. 3. Original image against CRT-Huffman compressed image

In Figure 3, the first image is the original sample image before applying the proposed compression algorithm. The second image is the output of compression procedure with CRT-Huffman coding. The two images showed large similarities, that is, the original and the reconstructed images, because of even distribution of bits compositions to the HVS.

The paper compared the performances of compression procedures of CRT-LZW and CRT-Huffman coding as shown in Table 3.

Table 3. Comparisons of CRT-LZW and CRT-Huffman coding based image compression

Evaluation Parameters	CRT-LZW	CRT-Huffman
Size Reduction (%)	29.78	14.00
Compression time (sec.)	9.95	19.15
PSNR	3.54	28.13
Compression ratio	0.39	4.38
Compressed Imperceptibility	Low	High

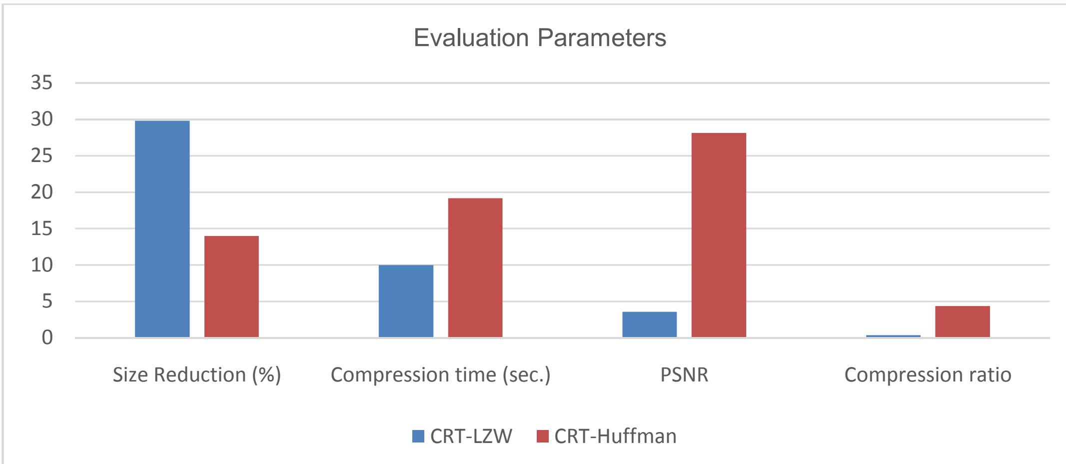

Fig. 4. Graphical view for parameters evaluation of LZW and Huffman coding

In Table 3 and figure 4, the introduction of CRT for LZW and Huffman coding based image compression showed significant performances with LZW saving more space and speedy compression (or redundancy removal) of original images. Conversely, Huffman coding offered better quality for the reconstructed images as against LZW [5].

5.CONCLUSION

The fundamental principle of data compression procedures ensure minimization of data redundancy (or resized data), better data compression time, improved or retention of data quality and high compression ratio. The conventional compression algorithms, such as LZW and Huffman coding, have shortcomings which limited their widespread implementations especially in image processing. One parameter for measuring the suitability compression procedure in image is the bits distribution because it reveals the similarity or otherwise of compressed image and original images to the Human Visual System (HVS). This paper implemented CRT in LZW and Huffman coding in order to improve their individual performances. The outcomes revealed that more space saving (or redundancy removal) and faster compression time were offered by CRT-LZW. But, CRT-Huffman coding (28.13) provided better quality (PSNR) for reconstructed images against CRT-LZW (3.54) and 25.59 in the study [5]. However, there is need for further implementation of these concepts in other media files such as text, videos and audio.

6.REFERENCES

1. Alhassan, A., Gbolagade, K. A. and Bankas, E. K. A novel and efficient LZW-RNS scheme for enhanced information compression and security. *International Journal of Advanced Research in Computer Engineering & Technology.* 2015;4(11):4015-4019.
2. Xiao, H., Huang, Y., Ye, Y. and Xiao, G. Robustness in Chinese Remainder Theorem for multiple numbers and remainder coding. *IEEE Transaction on Signal Processing.* 2018;1-16.
3. Thakur, U. and Rai, S. A study image compression techniques for the number of applications. *International Journal of Research and Innovation in Applied Science.* 2018;3(4):4-7.
4. Begaint, J., Thoreau, D., Guillotel, P. and Guillemot, C. Region-Based Prediction for Image Compression in the Cloud. *IEEE Transactions on Image Processing.* 2018;*27*(4):1835-1846.

5. Oswald, C. and Sivaselvan, B. Text and image compression based on data mining perspective. *Data Science Journal.* 2018;17(12):1-12.
6. Ajala, F. A., Adigun, A. A. and Oke, A. O. Development of hybrid compression algorithm for medical images using Lempel-Ziv-Welch and Huffman encoding. *International Journal of Recent Technology and Engineering.* 2018;7(4):1-5.
7. Garba, A. M. and Zirra, P. B. Analysing forward difference schemeon Huffman to encode and decode data losslessly. *The International Journal of Engineering and Science.* 2014;3(6):46-54.
8. Pujar, J. H. and L. M. Kadlaskar, L. M. A New Lossless Method of Image Compression and Decompression Using Huffman Coding Technique. *Journal of Theoretical and Applied Information Technology.* 2010;15(1):1-10.
9. Huffman, D. A. A method for the construction of minimum Redundancy codes. Proceedings of IRE, 40. 1952;1098-1101.
10. Gupta, K., Verma, R. L. and Alam S. Lossless Medical Image ... Communication Technology. 2013;1(2): 37-45.
11. Mishra, K., Verma, R.L., Alam, S. and Vikram, H. Minimum Entropy Analysis of Lossless Image Compression. International Journal of Research in Electronics and Communication Technology. 2013;1(2):67-75.
12. Welch, T. A. A Technique for High Performance Data Compression. IEEE, Sperry, Research Centre. 1984;8-19.
13. Jane H. and Trivedi, J. A Survey on Different Compression Techniques Algorithm for Data Compression. *International Journal of Advanced Research in Computer Science and Technology.* 2014:2(3);1-5.
14. Mohan, P. V. A. Residue Number Systems. The Springer International Series in Engineering and Computer Science, Springer US, 677. 2002;2-10.
15. Omondi, A. and B. Pumkumar, B. Residue Number Systems: Theory and Implementation. London WC2H 9HE: Imperial College Press. 2007;5-900.
16. Aremu, I. A. and Gbolagade, K. Z. RNS based on Shannon Fano Coding for Data Encoding and Decoding using $\{2^n-1, 2^n, 2^n+1\}$ Moduli Sets. Communications. 2018;6(1):25-29.