



Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

A survey on data compression techniques: From the perspective of data quality, coding schemes, data type and applications

Uthayakumar Jayasankar^{*}, Vengattaraman Thirumal, Dhavachelvan Ponnuramgam

Department of Computer Science, Pondicherry University, Puducherry, India

ARTICLE INFO

Article history:

Received 5 February 2018

Revised 8 May 2018

Accepted 16 May 2018

Available online 17 May 2018

Keywords:

Data compression

Data redundancy

Text compression

Image compression

Information theory

Entropy encoding

Transform coding

ABSTRACT

Explosive growth of data in digital world leads to the requirement of efficient technique to store and transmit data. Due to limited resources, data compression (DC) techniques are proposed to minimize the size of data being stored or communicated. As DC concepts results to effective utilization of available storage area and communication bandwidth, numerous approaches were developed in several aspects. In order to analyze how DC techniques and its applications have evolved, a detailed survey on many existing DC techniques is carried out to address the current requirements in terms of data quality, coding schemes, type of data and applications. A comparative analysis is also performed to identify the contribution of reviewed techniques in terms of their characteristics, underlying concepts, experimental factors and limitations. Finally, this paper insight to various open issues and research directions to explore the promising areas for future developments.

© 2018 The Authors. Production and hosting by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Contents

1. Introduction	120
1.1. Motivation	120
1.2. Contribution of this survey	121
1.3. Paper formulation	121
2. Related works	121
3. Compression quality metrics	122
4. Classification and overview of DC techniques	123
4.1. Based on data quality	124
4.2. Based on coding schemes	124
4.3. Based on the data type	126
4.3.1. Text compression	126
4.3.2. Image compression	128
4.3.3. Audio compression	130
4.3.4. Video compression	130
4.4. Based on application	131
4.4.1. Wireless sensor networks	131
4.4.2. Medical imaging	133
4.4.3. Specific applications	134

^{*} Corresponding author.

E-mail address: uthayresearchscholar@gmail.com (U. Jayasankar).

Peer review under responsibility of King Saud University.



Production and hosting by Elsevier

<https://doi.org/10.1016/j.jksuci.2018.05.006>

1319-1578/© 2018 The Authors. Production and hosting by Elsevier B.V. on behalf of King Saud University.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

5.	Research directions	137
5.1.	Characteristics of DC techniques	137
5.2.	Discrete tone images (DTI)	137
5.3.	Machine learning (ML) in DC	137
6.	Conclusions	137
	Appendix A	137
	Resources on data compression	137
	Web pages	137
	Books	137
	Scientific journals	137
	Conferences	137
	Software	138
	Dataset	138
	Popular press	138
	Mailing list	138
	References	138

1. Introduction

1.1. Motivation

Recent advancements in the field of information technology resulted to the generation of huge amount of data at each and every second. As a result, the storage and transmission of data is likely to increase to an enormous rate. According to Parkinson's First Law (Parkinson, 1957), the necessity of storage and transmission increases at least twice as storage and transmission capacities increases. Though optical fibers, Blu-ray, DVDs, Asymmetric Digital Subscriber Line (ADSL) and cable modems are available, the rate of growth of data is much higher than the rate of growth of technologies. So, it fails to address the above-mentioned issue of handling large data in terms of storage and transmission. In order to overcome this challenge, an alternative concept called data compression (DC) has been presented in the literature to compress the size of the data being stored or transmitted. It transforms the original data to its compact form by the recognition and utilization of patterns exists in the data. The evolution of DC starts with Morse code, introduced by Samuel Morse in the year 1838 to compress letters in telegraphs (Hamming, 1986). It uses smaller sequences to represent frequently occurring letters and thereby minimizing the message size and transmission time. This principle of Morse code is employed in the popular Huffman coding (Vitter Jeffrey, 1987). Nowadays, DC techniques are essential in most of the real time applications like satellite imagery, Geographical Information Systems (GIS), graphics, Wireless Sensor Networks (WSN), etc. Though the quality of data is tremendously increased by the development of recent technologies, it eventually increases the size of the data. For example, a moderate size color image of 512×512 pixels requires 0.75 MB of storage space. The uncompressed video for 1 s requires more than 20 MB of storage space. A micro tissue array image of 4096×4096 pixels consumes 48 MB of storage. A

remote sensing image of 8192×4096 pixels requires 96 MB of disk storage. However, the dedicated transmission of remote sensing data at the uplink and downlink speed of 64 kbps cost up to 1900 USD. The reduction of file size enables to store more information in the same storage space with lesser transmission time. So, without DC, it is very difficult and sometimes impossible to store or communicate a huge amount of data files.

In general, data can be compressed by eliminating data redundancy and irrelevancy. Modeling and coding are the two levels to compress data: In the first level, the data will be analyzed for any redundant information and extract it to develop a model. In the second level, the difference between the modeled and actual data called residual is computed and is coded by an encoding technique. There are several ways to characterize data and different characterization leads to the development of numerous DC approaches. Since numerous DC techniques have been developed, a need arises to review the existing methods which will be helpful for the upcoming researchers to approximately select the required algorithms to use in a particular situation. As an element of DC research, this paper surveys the development through the literature review, classification and application of research papers in the last two decades. In December 2017, a search was made on the important keywords related to DC in IEEE, Elsevier, Springer link, Wiley online, ACM Digital Library, Scopus and Google Scholar. As lossless compression, lossy compression, text compression, image compression, audio compression, video compression, bit reduction and data redundancy terms are the fundamental concepts involved in DC, a search was made using these keywords. The obtained search results in Table 1 reported that the number of articles related to image compression is much higher than text, audio and video compression. Next, the number of researches carried out in lossless compression is high when compared to lossy compression. In this study, some compression approaches are reviewed based on the citations, popularity, novelty, specific

Table 1
Search results for various keywords related to DC (1996–2017).

Keywords	IEEE Xplore	Elsevier	Springer link	Wiley online	ACM DL	Scopus	Google scholar
Data compression	37.6 k	413.7 k	34.6 k	156.4 k	144.1 k	79.0 k	1540 k
Lossy compression	2.48 k	3.83 k	1.96 k	1.69 k	5.04 k	4.17 k	41.1 k
Lossless compression	3.32 k	3.79 k	1.95 k	1.62 k	4.93 k	2.07 k	43 k
Text compression	1.15 k	42.9 k	17.34 k	41.2 k	30.4 k	0.15 k	1610 k
Image compression	25.9 k	179.6 k	9.38 k	121.0 k	31.8 k	0.13 k	154 k
Audio compression	2.26 k	8.14 k	2.91 k	5.11 k	10.6 k	0.033 k	502 k
Video compression	16.1 k	32.6 k	6.64 k	16.3 k	24.7 k	0.027 k	1100 k
Bit reduction	10.9 k	138.3 k	6.87 k	119.5 k	17.6 k	0.15 k	1790 k
Data redundancy	9.66 k	88.0 k	34.0 k	68.1 k	144.4 k	0.04 k	499 k

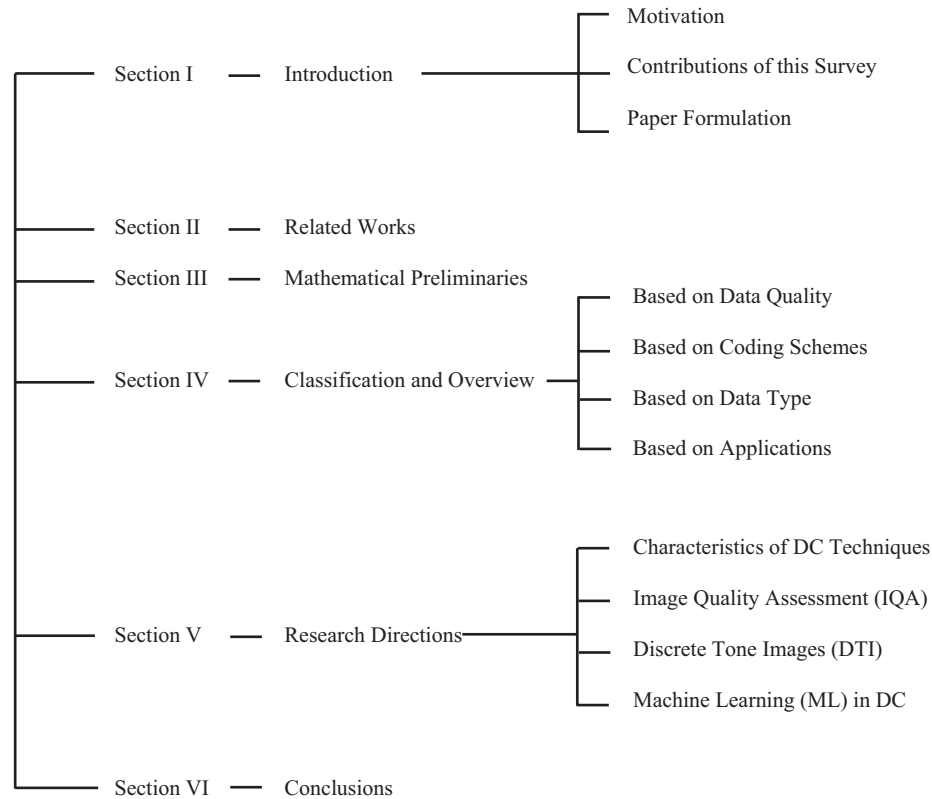


Fig. 1. Paper Formulation.

applications, etc. At the end of this paper, readers will aware of the existing DC techniques and research directions in different dimensions. The important resources which are highly useful in the field of DC is given in [Appendix](#).

1.2. Contribution of this survey

The contribution of this paper is four-folded, which is summarized as follows.

- We present an in-depth explanation of the basic concepts, its importance, mathematical formulation and performance measures in the context of DC techniques.
- We present a clear classification of various DC algorithms on the basis of data quality, coding schemes, data type and applications.
- We provide a comprehensive review on up-to-date DC techniques with their objectives, strength, weakness and performance evaluation.
- After discussing the state of art approaches, we suggest several open issues to explore the possible future trends in DC according to their recent advancements.

1.3. Paper formulation

The systematic organization of the paper is presented as follows. In Section II, we summarize the previous surveys on DC techniques with year wise classification. In Section III, we commence with mathematical preliminaries for the purpose of understanding the basic information theory and compression quality metrics. In Section IV, the comprehensive classification of DC techniques is done according to quality, coding schemes, data type and its applications. In Section V, we provide a wide range of issues, challenges

and possible future research directions in DC. Finally, in Section VI, we conclude and create an interestingness of research under DC. For better understanding and clarity, the formulation of this paper is shown in [Fig. 1](#).

2. Related works

In this section, existing survey papers presented in the field of DC are explained in detail. Initially, we reviewed some complete survey papers on DC techniques and then moved to the surveys focused on specific application criteria. One of the most important survey on DC has been presented in ([Smith, 2010](#)). In this paper, the author described DC techniques from mathematical level to coding level. This paper dealt with lossy compression and focused only on Fourier Transform and wavelet compression. A survey on lossless DC techniques is given in ([Holtz, 1999](#)). The author pointed out that “Theories are usually the starting point of any new technology”. The author theoretically explained some lossless compression methods namely Shannon’s theory, Huffman code, Lempel Ziv (LZ) code and Self-Learning Autopsy data trees. This paper missed numerous lossless DC techniques and reviewed only a few number of existing techniques. Another survey is presented in ([Hosseini, 2012](#)), which tried to explain many DC algorithms with its performance evaluation and applications. This paper has explained Huffman algorithm, Run Length encoding (RLE), LZ algorithm, Arithmetic coding, JPEG and MPEG with their applications in diverse fields. This paper is ended with today DC issues and research ideas towards energy efficiency.

In addition to these reviews, there are some survey papers which focused only on image compression techniques. A survey work is presented in ([Sudhakar et al., 2005](#)), where ten images compression techniques using wavelet coefficients are reviewed. This paper outlines the features, merits, demerits of reviewed tech-

niques and compared to one other. The author examined the performance of these techniques using Compression Ratio (CR) and Peak Signal to Noise Ratio (PSNR). Another attempt in this area is presented in (Chew and Ang, 2008), where some image compression algorithms employed in WSN are explained. The author reviewed eight major image compression methods and classified them into first generation and second generation image compression algorithms. A comparison is also done on several characteristics such as preprocessing, codebook, post-processing, memory, complexity, size and compression quality. DC algorithms become more and more popular in medical imaging and large numbers of researches are carried out in this field. Sridevi et al. (2012) presented a short survey on compression approaches for medical imaging. A research study on image compression algorithms are presented in (Rehman et al., 2014). This paper reviewed most of the image compression techniques and a comparison is also done based on underlying techniques, features, merits, demerits, applications and performance results. Another detailed survey has been given in (Tarek et al., 2016) which elaborates Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT) image compression techniques. This paper analyzed the performance using PSNR, CR, throughput, end to end (ETE) delay, battery life-time and average Media Access Control (MAC) layer delay. Another survey is presented in (Kavitha, 2016), which explained lossy as well as lossless compression approaches and is compared to one another. Rana and Thakur (2017) presented a survey of DC techniques developed especially for computer vision applications. The existing techniques are reviewed and the performance is analyzed based on entropy and CR. A review on 3D mesh compression technique is presented in (Peng et al., 2005) which focused on triangular, single-rate and progressive mesh compression approaches. Another survey on mesh compression is done in (Li et al., 2014), which reviews single rate, progressive and mesh compression methods with random accessibility. The reviewed techniques are compared with one another in terms of performance and complexity.

Kimura and Latifi (2005) presented a short survey on DC techniques especially developed for WSN such as coding by ordering, Pipelined In-Network compression, low complexity video compression and distributed compression. Another review work is presented in (Srisooksai et al., 2012), which reviewed and classified all DC algorithms in WSN into two major groups namely disturbed and local DC. The performance of the reviewed methods is evaluated and compared with respect to compression performance,

power saved from the transmission, power utilized for compression, net power saving, coding size, suitability of using DC classes and suitability of single or multiple data types. A survey of low power adaptive image compression algorithms for resource constrained WSN is presented in (Rekha and Samundiswary, 2015). This paper reviewed various image compression techniques and compared its performance using PSNR, Structural Similarity Index (SSIM), Mean Square Error (MSE) and CR. A survey of image compression algorithms involved in wireless multimedia sensor networks (WMSN) is given in (ZainEldin et al., 2015). This paper provides the analysis of related research directions and advanced image compression algorithms with its advantages and drawbacks. Another survey in this area is attempted in (Patel and Chaudhary, 2017), where the image compression techniques based on DCT, DWT and hybrid approaches in WMSN are reviewed. They are compared using ten parameters and observed that hybrid approaches are the efficient way to achieve energy efficiency in WMSN. Mittal and Vetter (2016) presented a survey of compression techniques employed in cache and main memory systems. It discussed different compression methods involved in CPUs, GPUs, non-volatile memory systems, 2D and 3D memory systems. An overview of various standards for remote sensing DC is provided in (Ian et al., 2014). It dealt with the recently approved standards by the Consultative Committee for Space Data Systems (CCSDS) along with ISO/IEC image coding standards. It also explained both mono band and multi band compression, lossless, lossy and near-lossless compression, along with a comparison against the state of art methods.

The existing survey works on DC techniques are tabulated in Table 2. After studying extensive survey papers in this field, we realized that there is no detailed survey which reviews all types of compression techniques in a clear and classified manner. This motivated us to perform this comprehensive survey that targets up to date DC techniques.

3. Compression quality metrics

Some of the concepts in information theory provides an outline for the design of lossless DC approaches and are explained here. In information theory, the entropy of a random variable is measured by a term called self-information (Shannon, 1938). Let us consider an event X, which is the outcome of a random experiment. When the probability of an event X is P(X) and the self-information related with X is represented in Eq. (1).

Table 2
List of previous surveys on DC techniques.

Reference	Year	Classification
Holtz (1999)	1997	Shannon entropy, Huffman code, LZ code and Self-Learning Autopsy data trees
Kimura and Latifi (2005)	2005	Compression techniques for WSN
Sudhakar et al. (2005)	2005	Ten wavelet coding techniques
Peng et al. (2005)	2005	3D mesh compression techniques
Chew and Ang (2008)	2008	Eight image compression algorithms
Smith (2010)	2010	Lossy compression
Sridevi et al. (2012)	2012	Short survey on medical image compression
Srisooksai et al. (2012)	2012	Distributed data compression and local data compression
Hosseini (2012)	2012	Huffman, RLE, LZ and Arithmetic coding with its applications
Rehman et al. (2014)	2014	Image compression techniques
Li et al. (2014)	2014	Mesh compression methods with random accessibility
Ian et al. (2014)	2014	Various standards for remote sensing data compression
ZainEldin et al. (2015)	2015	Image compression techniques in WMSN
Rekha and Samundiswary (2015)	2015	Low power adaptive image compression techniques in WSN
Kavitha (2016)	2016	RLE, LZW and Huffman coding, Transform coding, DCT and DWT
Tarek et al. (2016)	2016	DCT and DWT image compression techniques
Mittal and Vetter (2016)	2016	Compression in cache and main memory systems
Rana and Thakur (2017)	2017	Computer vision applications
Patel and Chaudhary (2017)	2017	DCT, DWT and hybrid approaches in WMSN

$$i(X) = -\log_b P(X) \quad (1)$$

For two events A and B, self-information related with X and Y is given as

$$i(X) = \log_b \frac{1}{P(X)} \quad (2)$$

when X and Y are independent, $P(XY) = P(X)P(Y)$, then the self-information related with X and Y is defined by,

$$i(XY) = \log_b \frac{1}{P(X)P(Y)} = \log_b \frac{1}{P(X)} + \log_b \frac{1}{P(Y)} = i(X) + i(Y) \quad (3)$$

The performance of DC algorithms can be analyzed in several aspects. We can measure the algorithm complexity, computational memory, speed, amount of compression and quality of reconstructed data. The most common measure to calculate the efficiency of a compression algorithm is CR. It is defined as the ratio of total number of bits required to store uncompressed data and total number of bits required to store compressed data.

$$CR = \frac{\text{No. of bits in uncompressed data}}{\text{No. of bits in compressed data}} \quad (4)$$

CR can also be termed as bit per bit (bpb). It defines the average number of bits required to store the compressed data. Likewise, bpb is referred as bits per pixel (bpp) in image compression techniques whereas the recent text compression methods use bits per character (bpc) which represents the number of bits, on average, needed to compress a character. Another measure called Space savings is also used, which defines the reduction in file size relative to the uncompressed size and is given in Eq. (5).

$$\text{Space savings} = 1 - \frac{\text{No. of bits in compressed data}}{\text{No. of bits in uncompressed data}} \quad (5)$$

For instance, when the original data is 10 MB and the compressed data is 2 MB, then space savings will be $(1 - 2/10) = 0.8$. A value of 0.8 indicates that 80% of storage space is saved due to compression. Next, the compression gain is equated in Eq. (6).

$$\text{Compression gain} = 100 \log_e \frac{\text{original data}}{\text{compressed data}} \quad (6)$$

The speed of compression can be calculated in terms of cycles per byte (CPB). It implies the number of cycles required, on average, to compress one byte. CR and CF are good enough to measure the performance of lossless compression techniques. Since in lossy compression, the reconstructed data actually varies from original data, it needs some additional performance measures to assess the level of distortion, fidelity and quality. The variation in reconstructed data from actual data in lossy compression is termed as distortion. Other terms such as fidelity and quality are also used to represent the difference from original and reconstructed data. A common metric employed for this purpose is PSNR, which is a dimensionless number used to identify the perceived errors noticeable by the human vision. PSNR is generally expressed in terms of logarithmic decibel scale (dB). For the original data X_i and the reconstructed data Y_i , PSNR is equated as

$$\text{PSNR} = 20 \log_{10} \frac{\max |X_i|}{\text{RMSE}} \quad (7)$$

where RMSE (Root mean square error) is the square root of Mean square error (MSE) and is formulated in Eq. (8).

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (X_i - Y_i)^2 \quad (8)$$

When the original and reconstructed data are exactly identical, then RMSE will be zero and PSNR will be infinity. For better similarity between the original and reconstructed data, the value of RMSE will be low and PSNR value will be high. Another related metric is

$$\text{Signal to Noise Ratio (SNR)} = 20 \log_{10} \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n X_i^2}}{\text{RMSE}} \quad (9)$$

SNR is used to measure the size of the error associated with the signal. The numerator in SNR is the root mean square of the original data. Though PSNR is easiest and commonly used measure, it eliminates the influence of image content by viewing conditions and may not correlate with the human perception. To handle this issue, several metrics based on human visual system (HVS) are proposed such as visual signal to noise ratio (VSNR) (Chandler and Hemami, 2007), weighted signal to noise ratio (WSNR) (Elshaikh et al., 2012) and noise quality measure (NQM) (Damera-Venkata et al., 2000). Next, universal image quality index (UQI) is introduced to compute the loss of structural information in the distorted image (Wang and Bovik, 2002). It represents any distorted image as an integration of correlation distortion, luminance distortion and contrast distortion. It can also be termed as a universal index because it is independent of distortion type, viewing distance and observers. SSIM and multi-scale SSIM (MS-SSIM) (Wang et al., 2003) are presented as a complementary measure of the conventional metrics like PSNR and MSE. SSIM is a HVS based measure to determine the structural similarity between the reference and distorted images (Wang et al., 2004). It is a process of modeling the image distortion by the combination of contrast sensitivity function (CSF), channel decomposition, loss in correlation and luminance. SSIM ranges between [0, 1], where value closer to 1 represents better resemblance between two images and the value nearer to 0 indicates lesser resemblance. Some other metrics are also devised based on SSIM such as FSIM (Zhang et al., 2011), GMSD (Liu et al., 2012), SR-SIM (Zhang and Li, 2012), HaarPSI (Reisenhofer et al., 2016) and so on. The measures using natural scene statistics (NSS) (Moorthy and Bovik, 2011) are also introduced namely information fidelity criterion (IFC) (Sheikh et al., 2005) and its extended version named visual information fidelity (VIF) (Wang et al., 2006). These measures evaluate the image quality by identifying the statistical features of real environment.

Rate distortion curve is also used to compute the compression performance of lossy compression techniques. It determines the minimum number of bits required to store a symbol, as measured by the rate R (Berger, 1971). In rate-distortion theory, the rate is defined as the number of bits per data sample to be stored or transmitted. In addition, the distortion is defined as the expected value of the square of the difference between input and output signal (i.e., the mean squared error).

Since numerous compression quality metrics have been developed, none of the metric assess the quality of all types of signals. For text compression, the quality is assessed by the use of CR, CF and bpc. The metrics used for image compression are CR, PSNR, MSE, RMSE, SSIM, MS-SSIM, bpp (bits per pixel), etc. are employed. For specific application-oriented compression techniques, the performance measures are based on the type of data involved in the application. For instance, when the application involves satellite or medical images, traditional metrics such as PSNR, MSE, and SSIM will be used.

4. Classification and overview of DC techniques

As numerous DC techniques have been developed, a need arises to review the techniques and approximately select the algorithm to use in a particular situation. At the same time, it is very hard to classify these techniques under some criteria. In this paper, we attempt to review and classify various DC techniques based on four categories namely reconstructed data quality, coding schemes, data type and application suitability. The classification hierarchy

is depicted in Fig. 2. The existing techniques are explained with their objectives, underlying technique, application and their performance evaluation.

4.1. Based on data quality

Generally, a DC technique influences the data quality to a certain extent based on the application criteria. When a DC technique is used for general purposes like messaging or browsing internet, the quality of reconstructed data is not highly considered. But, in text compression, the modification of a single character in a text document is not tolerable as it changes the entire meaning. Similarly, in medical imaging or remote sensing image compression, slight changes in pixel values are not desirable. The significance of the data quality of a DC technique is highly depends on the type of data or application involved. Depending upon the requirement of reconstructed data quality, DC techniques can be divided into lossless and lossy compression. As the name implies, lossless compression refers to no loss of information, i.e. the reconstructed data is identical to original data. It is used in applications where loss of information is undesirable like text, medical imaging, law forensics, military imagery, satellite imaging, etc. (Drost and Bourbakis, 2001). In some scenarios, lossy compression techniques are preferable where the reconstructed data is not perfectly matched with the original data and the approximation of original data is also acceptable. It leads to higher CR for lossy compression techniques when compared to lossless compression techniques. Near-Lossless compression techniques is an another type of compression technique where the difference between the original and reconstructed data is guaranteed to vary from the respectively values in the original data by no more than a user-specified amount called as maximum absolute distortion (MAD) (Ansari et al., 1998). It is used to compress medical images, hyper spectral images, videos and so on.

4.2. Based on coding schemes

In this section, the popular coding techniques such as Huffman coding, Arithmetic coding, LZ coding, Burrows-wheeler transform

(BWT) coding, RLE, transform coding, predictive coding, dictionary based methods, fractal compression, Scalar and vector quantization which are considered as the ancestors in the field of DC are reviewed. These techniques are shown in Fig. 2 and the comparison is tabulated in Table 3.

Huffman coding (Huffman, 1952) is the famous coding technique which effectively compress data in almost all file formats. It is a type of optimal prefix code which is widely employed in lossless DC. The basic idea is to assign variable length codes to input characters depending upon the frequency of occurrence. The output is the variable length code table for coding a source symbol. It is uniquely decodable and consists of two components such as constructing Huffman tree from input sequence and traversing the tree to assign codes to characters. Huffman coding is still popular because of its simpler implementation, faster compression and lack of patent coverage. There are several versions of Huffman coding includes Minimum variance Huffman code, Canonical Huffman code, Length-limited Huffman code, Non-binary Huffman code, Adaptive Huffman code, Golomb code, Rice code and Tunstall code. Several compression methods like Deflate, JPEG, MP3, etc. uses Huffman code as the back-end technique. A simpler form of lossless DC coding technique is Run Length Encoding (RLE) (Capon, 1959). It represents the sequence of symbols as runs and rest of them are termed as non-runs. A run consists of two parts namely data value and count instead of original run. Though RLE is simpler to implement and faster, it is ineffective for less redundant data and leads to increase in compressed file size greater than original file size. It is used in line drawings, fax, palette-based images like textures, animation and graphic icons.

DC techniques can be classified to predictive and transform coding techniques. In predictive coding, the present information is exploited to predict the upcoming data, and the actual difference is encoded. It is simple, easier to implement and can be adjustable to diverse local image features. Next, transform coding, converts an input data from one kind of representation to other kind of representation and the transformed values (coefficients) are encoded by compression techniques. Transform coding performs well than predictive coding with the cost of high computational complexity. Generally, a compression model under transform coding consists of

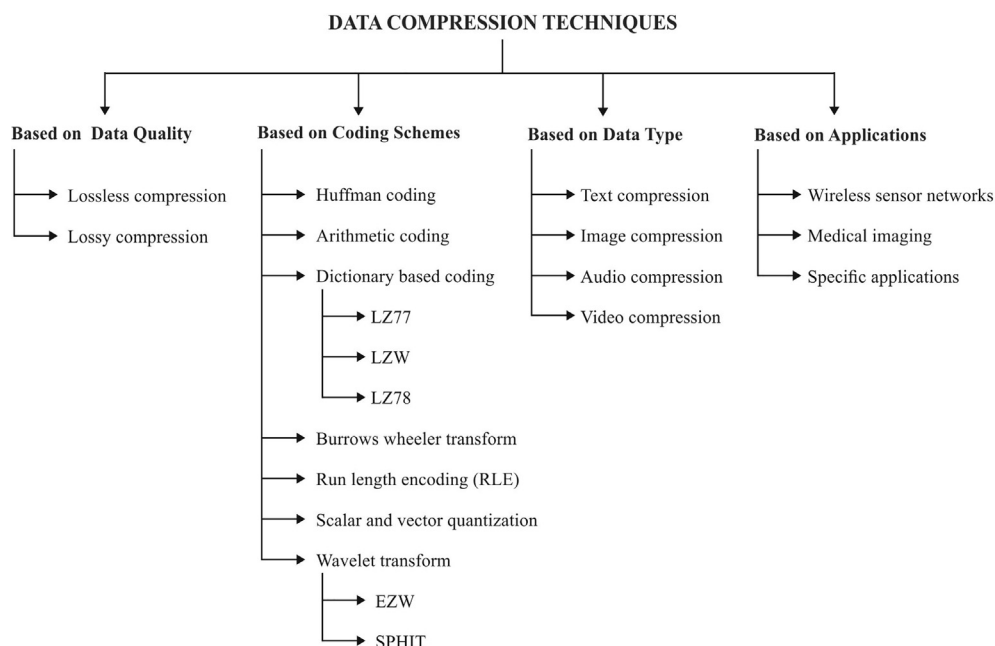


Fig. 2. Classification of DC techniques.

Table 3

Comparison of various coding schemes.

Reference	Coding	Feature	Compression Type	Versions	Advantages	Applications
Huffman (1952)	Huffman coding	Entropy based	Lossless	Minimum variance Huffman code, Length Limited Huffman code, Adaptive non-binary, Golomb-rice coding, Tunstall code	Effective in all file formats	ZIP, ARG, JPEG, MPEG, PKZIP
Langdon (1984)	Arithmetic coding	Entropy based	Lossy and Lossless	Adaptive arithmetic coding	Flexibility	JPEG, multimedia applications
Ziv and Lempel (1977)	LZ coding	Dictionary based coding	Lossless	Binary arithmetic coding LZ77, LZ78, LZW	Compress all kinds of data	TIFF, GIF, PDF, Gzip, ZIP, V.42, Deflate and PNG
Saupe and Hamzaoui (1994)	Fractal compression	Block based coding	Lossy	–	Suitable for textures and natural images	Live video broadcasting
Burrows and Wheeler (1994)	BWT	Block sorting compression	Lossless	–	No need to store additional data for compression	Bzip2
Capon (1959)	RLE	Employs in high redundant data	Lossless	–	Faster	TIFF, BMP, PDF and fax
Sayood (2006)	Scalar and Vector Quantization	Represents larger set of values to a smaller set	Lossless and Lossy	–	Less complexity	–

three components namely transformer, quantizer and encoder. The resultant coefficients from transformer are quantized and symbol encoding methods are used to attain output bit stream, which represents the compressed data. During decompression, at the decoder size, the reversible operation takes place. A better transform coding technique has the ability to compress data using less number of coefficients. DCT and DWT are the most widely used transform coding techniques (Narasimha and Peterson, 1978).

In (Langdon and Rissanen, 1981), a black and white lossless image compression technique is developed using source models with no alphabet extension. The proposed method is based on efficient binary arithmetic code and it involves two levels: modeling and coding. In the modeling phase, the structure is chosen which identifies the way the events are to be conditioned, and then the relative frequencies of the conditioned events are collected. The frequencies are coded by the coding unit to create the code string. To validate the compression performance, the test documents were provided by French PTT to the CCITT Study Group XIV. The results reported that the proposed method is superior to existing methods. (Todd et al., 1985) introduces a lossless compression of gray-scale images which follows the previous black/white image compression techniques. An universal data compression algorithm is also proposed in (Rissanen, 1983), which has the capability to compress long strings created by a “finitely generated” source, with near optimum per symbol length without any knowledge of the source. Here, an arithmetic binary code is applied to encode every symbol $u = x(t)$ by the use of conditional probability $P(u/z^*(t))$. (Langdon and Rissanen, 1983) introduced a one-pass double-adaptive file compression (DAFC) method which uses an adaptive variable-order Markov structure and adaptive statistics unit. A one-pass compression method is used which assumes that no statistical properties exists in the data needs to be compressed. The model structure adaptively chooses a subset of first-order Markov contexts, based on an estimation of the candidate context's popularity. In (Langdon and Rissanen, 1983), a general framework for sequential modeling of gray-scale images is presented and employed to lossless compression. This model uses stochastic complexity considerations and is implemented with a tree structure. It is effectively computed by a modification of the universal Algorithm Context. Several variants of the algorithm are discussed. The sequential, lossless compression methods attained when the con-

text modeler is employed with an arithmetic coder are validated with a representative set of gray-scale images. The results of the proposed method are compared with existing methods in terms of CR.

Arithmetic coding (Langdon, 1984) is an another important coding technique to generate variable length codes. It is superior to Huffman coding and is highly useful in situations where the source contains small alphabets with skewed probabilities. A benefit of arithmetic coding over Huffman coding is the capability to segregate the modeling and coding features of the compression technique. When a string is encoded using arithmetic coding, frequently occurring symbols are coded with less number of bits than rarely occurring symbols. It is not easier to implement when compared to other methods. There are two versions of arithmetic coding namely Adaptive Arithmetic Coding and Binary Arithmetic Coding.

Dictionary based coding approaches find useful in situations where the original data contains more repeated patterns. When a pattern comes in the input sequence, they are coded with an index to the dictionary. When the pattern is not available in the dictionary, it is coded with any less efficient approaches. It can be divided into two classes such as frequently and infrequently occurring patterns. This method will be efficient by considering shorter codeword for frequently occurring patterns and vice versa. The two types of available dictionaries are static and dynamic dictionary. Static dictionary is useful when the prior knowledge of source output is available. When the prior information of the original data is not available, dynamic dictionary will be used. Lempel–Ziv algorithm (LZ) is a dictionary based coding technique commonly used in lossless file compression. This is widely used because of its adaptability to various file formats. It looks for frequently occurring patterns and replaces them by a single symbol. It maintains a dictionary of these patterns and the length of dictionary is set to a particular value. This method is much effective for larger files and less effective for smaller files. In the year 1977 and 1978, two versions of LZ were developed by Ziv and Lempel named as LZ77 (Ziv and Lempel, 1977) and LZ78 (Ziv and Lempel, 1978). These algorithms vary significantly in means of searching and finding matches. LZ77 algorithm basically uses a sliding window concept and searches for matches in a window within a predetermined distance back from the present position. LZ78 algorithm follows a

more conservative approach of appending strings to the dictionary. Lempel–Ziv–Welch (LZW) is an enhanced version of LZ77 and LZ78 which is developed by Terry Welch in 1984 (Welch, 1984). The encoder constructs an adaptive dictionary to characterize the variable-length strings with no prior knowledge of the input. The decoder also constructs the similar dictionary as encoder based on the received code dynamically. UNIX compress, GIF images, PNG images and others file formats use LZW coding whereas LZ77 is used in Gzip and ZIP.

Fractal compression is one of the lossy compression technique used to compress digital images using fractals. It is highly suitable for textures and natural images, based on the fact that parts of an image often resemble with other parts of the same image. Fractal algorithms transform the similar parts to mathematical data known as fractal codes which are employed to generate the encoded image. Fractal image representation can be defined mathematically as an iterated function system (IFS). Fractal image compression algorithms involve two stages: image segmentation and encoding. In the first stage, the image is segmented to parts and the parts are individually coded by IFS probabilities. The idea behind this is the use of collage theorem which provides criteria for selecting transformations in IFS to optimize the overall result. At last, the decoded parts are arranged to generate the entire decoded IFS representation. As the decoding process is automated, human interaction is needed in the encoding or segmentation process. Some of the fractal image compression techniques are found in the literature (Saupe and Hamzaoui, 1994).

Burrows–Wheeler Transform (BWT) (Burrows and Wheeler, 1994) is a block sorting compression technique which rearranges the character string into runs of identical characters. It uses two techniques to compress data include move-to-front transform and RLE. It compresses data easily in situations where the string consists of runs of repeated characters. It is employed to compress data in bzip2. Quantization is an easiest way to represent larger set of values to a smaller set. The quantizer input can be either scalar or vector. Scalar quantization is a simpler and commonly used lossy compression technique. It is a process of mapping an input

value to a number of output values. Vector quantization (VQ) is a different type of quantization, which is typically implemented by choosing a set of representatives from the input space and then mapping all other points in the space to the closest representative.

4.3. Based on the data type

Generally, DC has been employed to compress text, image, audio, video and some application specific data. This section reviews most of the important and recent compression techniques under each datatype and the classification hierarchy is shown in Fig. 3 and the reviewed methods are compared in Table 4.

4.3.1. Text compression

Basically, textual data undergoes lossless compression where loss of information is not allowable. (Abel and Teahan, 2005) developed a universal pre-processing method to compress textual data. It integrates five algorithms include capital letter conversion, end of line (EOL) coding, word replacement, phrase replacement and recording of alphabets. It has no dependency on languages and does not need any dictionaries. But, it is observed that the pre-processing cost is found to be high. A new compression technique based on evolutionary programming approach is developed (Ullah, 2012). In this paper, an evolutionary approach is defined by extracting some features from Genetic Programming (GP) and Cartesian Genetic Programming (CGP). The compression model uses an evolutionary algorithm and the input text is converted into number of chunks depending upon the number of nodes. Then, evolution begins on various metrics including mutation rates, number of offspring and types of functions. Once the textual data is compressed, it will be sent to the decoder along with the genotype. The genotype serves as a key and is very useful in the decompression process. The proposed method is tested with textual data copied from several portions of the poem, Alexander Pope’s “Eloisa to Abelard”. The proposed method achieves better compression when the number of chunks increases.

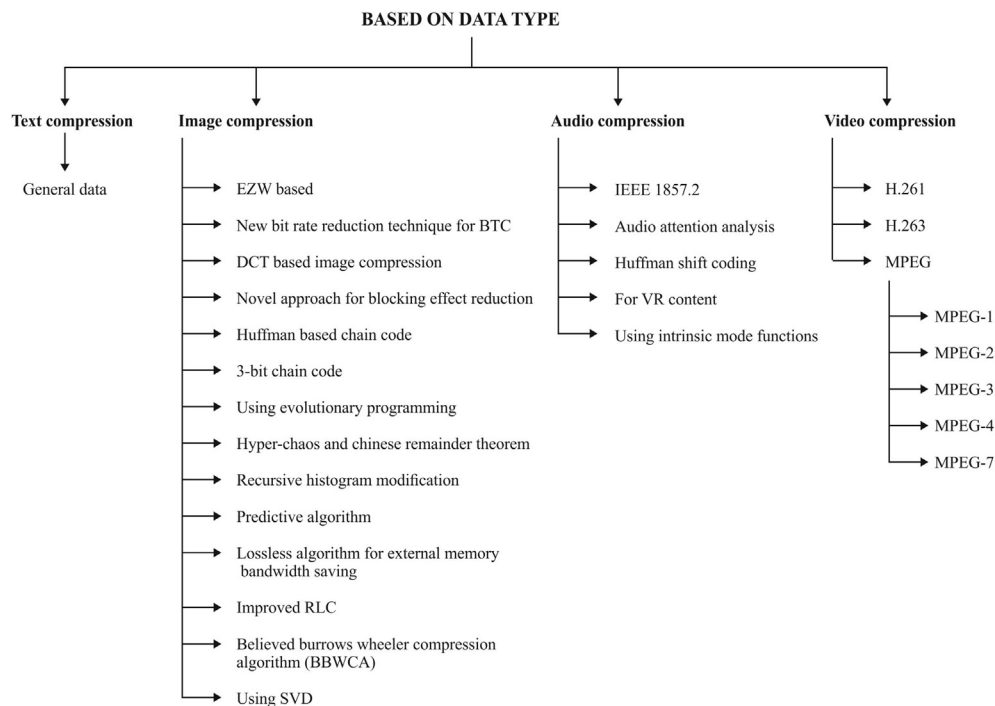


Fig. 3. Classification of compression techniques based on data type.

Table 4
Comparison of various compression techniques based on data type.

Reference	Objective	Compression Type	Data type	Methodology	Performance metrics	Compared with	Applications
Mahmud (2012)	Represent data by its half bits	Lossless	General data	Logical truth table	CR, CF	–	Wired and wireless networks
Abel and Teahan (2005)	Develop a universal pre-processing method to compress textual data	–	Textual data	capital letter conversion, EOL coding, word replacement, phrase replacement and recording of alphabets	Compression gain	–	–
NoPlatoš et al. (2008)	To compress smaller text files	lossless	Textual data	BWT and Boolean minimization	CR	LZW, Huffman, bzip, gxp2	–
Kalajdzic et al. (2015)	To compress short messages	Lossless	Textual data	b64 pack	CR, CT	Compress, gzip, bzip2	–
Robert and Nadarajan (2009)	To introduce a pre-processing method for traditional coding schemes	Lossless	Textual data	genetic reversible transformation	–	–	–
De Agostino (2015)	To develop static compress technique	Lossless	Textual data	greedy approach	CR	Suffix closed and prefix closed dictionaries	distributed systems
Che et al. (2015)	To compress sentimental sentences	Lossless	Textual data	Sent comp	–	–	Sentimental analysis
Oswald et al. (2015)	To utilize data mining tools in DC	Lossless	Textual data	FPHuffman	CR	Huffman	–
Oswald et al. (2017)	to mine sequence of characters involved in the compression process	Lossless	Textual data	Graph based method	CR, CT	–	–
Oswald and Sivaselvan (2017)	To introduce a novel FIM based Huffman coding techniques	Lossless	Textual data	FPH2 method	CR, CT	LZW, LZSS, LZ77, FPH1	–
Jerome (1993)	Bit streams are created in the order of importance	Lossless	Image	DWT, Successive- Approximation Quantization and Adaptive Arithmetic coding	CR, PSNR, MSE, Significant coefficient	JPEG	Browsing images, multimedia applications
Rao and Eswaran (1996)	Reduction of bit rate for BTC	Lossless	Image	One-bit adaptive quantizer	Bit rate, MSE, quality	BTC, MBTC, HVO, HVB	Low bit rate applications
Luo et al. (1996)	Artifact reduction in DCT	Lossless	Image	Huber-Markov random field model, ICM	Bit rate, PSNR	DCT	Document images, resolution chart
Wang and Zhang (1998)	To reduce blocking effect	Lossless	Image	Post-processing technique at decoder	PSNR, Visual quality	POCS, CLS, Crouse method	–
Hermilo and Rodríguez-Dagnino (2005)	To compress bi-level images	Lossless	Image	3-bit chain code	CR	Freeman code, ZIP and JBIG	Bi-level images
Ullah (2012)	Compression based on Evolutionary programming	Lossless	Image	Genetic Programming, Cartesian Genetic Programming	CR	LZW, Huffman, RLE	–
Zhu et al. (2013)	Simultaneous encryption and compression	Lossless	Image	Chinese Remainder theorem-based compression	CR	DET, DES	–
Rad et al. (2013)	Design an efficient highly adaptive predictor	Lossless	Image	GBTA	MAE, visual quality	MED, GAP, GBSW, Graham, DARC	Multimedia
Zhang et al. (2013)	Establishing equivalency between RDH and lossless compression	Lossless	Image	Recursive Histogram Modification	PSNR	Tailored reversible watermarking reversible watermarking	medical images, military imagery and law forensics
Yin and Hu (2014)	To compress the image pixels prior to writing in SDRAM	Lossless	Image	Residue coding Huffman based VLC tables	CR	High profile video decoder Significant Bit Truncation for HD Video Coding	–
Alzahir and Borici (2015)	To compress discrete color images	Lossless	Image	Codebook and row column reduction coding	CR	Maps, GIS, binary images	JBIG
Anantha Babu et al. (2016)	Overcome the drawback of RLC	Lossless	Image	Matrix based mapped pair	RMSE, SNR, PSNR, CR, CT, bit rate, compressed file size	Location-based, Huffman, DCT	TIFF, GIF, JPEG and textual files

(continued on next page)

Table 4 (continued)

Reference	Objective	Compression Type	Data type	Methodology	Performance metrics	Compared with	Applications
Khan et al. (2016)	To achieve robustness and compression efficiency	Lossless	Image	Reversible Color Transform (RCT), BWT	CR	HEVC, LZ4X	Color images, EEG data, raster images
Kumar and Vaish (2017)	To achieve better compression and security	Lossless	Image	SVD	CR, PSNR	JPEG	Secured image transmission
Auristin (2016)	Optimizing bandwidth and storage	Lossless	Audio	Huffman coding IEEE1857.2	CR, Compressed file size	Mp4, wav, ogg, flac	Music audio and speech files
Hang et al. (2016)	To adapt bandwidth variations	Lossless	Audio	Audio attention analysis	Bit rate, quality	-	-
Brettle and Skoglund (2016)	To transform the nature of lossless audio data to lossy	Lossy	Audio	Huffman Shift Coding	CT, CR	-	Multimedia
Jain et al. (2017)	To produce original audio quality	Lossy	Audio	LZ compression	CR, SNR, PSNR, NRMSE	DCT, DWT	Audio-video conferencing, VoIP, VoWLAN

NoPlatoš et al. (2008) devised a novel compression method based on Boolean minimization of binary data along with BWT coding to compress smaller text files. It is used in various applications where smaller texts needs to stored or transmitted like Facebook, SMS, instant messaging, etc. The performance of this method is tested using medium as well as small text files and compared with versions of Huffman coding, LZW and BWT coding in terms of CR and CS. (Kalajdzic et al., 2015) presented an efficient technique called b64pack for short messages, it is an efficient and lightweight algorithm which is easy to deploy and interoperable. b64pack operates in two phases such as conversion of original data to a compressible format and transformation. It is found to be efficient and faster than other methods like compress, gzip and bzip2. Robert and Nadarajan (2009) introduced a pre-processing method for Huffman coding, arithmetic coding, versions of LZ and BWT coding. The author uses a genetic reversible transformation which converts a text file to other formats with lower file size. It helps to progress the performance of the backend compression algorithm. A novel DC technique is proposed to compress general data using a logical truth table (Mahmud, 2012) and thereby two bits of the data are represented by only one bit. Using this method, the data bits can be represented by its half bits only, i.e.1 GB data can be represented by 512 MB. Traditionally, there are two levels to represent digital data but the proposed method uses four levels and is validated by the use of CR and CF. The author fails to compare the proposed method with the existing methods and it is employed in wired and wireless scenarios. De Agostino (2015) presented a greedy approach for static text compression to relax the prefix property of the dictionary. It involves a finite state machine (FSM) realization of greedy based compression with arbitrary dictionary to attain high speed in distributed systems. A new compression model called sent comp is developed in (Che et al., 2015), for the compression of sentiment sentences in aspect based sentimental analysis. Sent comp intends to eliminate the unwanted information by compression the complex sentiment sentences to short and easier to parse sentiments. Oswald et al. (2015) utilized data mining tools in the domain of text compression. Here, Huffman coding is enhanced by the combination of frequent itemset mining (FIM). It is based on the idea assigning shorter codewords to repeatedly occurring patterns. (Oswald et al., 2017) used a graph-based method to mine sequence of characters involved in the compression process. This method constructs a graph in once pass of the text and mine all patterns which are mandatory for compression in one pass of the graph. (Oswald and Sivaselvan, 2017) introduced another novel FIM based Huffman coding techniques using hash table (FPH2) to compress text in the process of frequent pattern counting. Optimal set of patterns is used in FPH2 while character based approach is involved in traditional Huffman coding. This method is tested against a set of 19 benchmark dataset and the results are compared with gzip, LZW, LZSS and FP Huffman in terms of CR and CT.

4.3.2. Image compression

As the number of image transmissions are increasing day by day, large numbers of researches are carried out to effectively store or transmit images. This section discusses the existing image compression techniques with the comparison against one another.

Embedded Image Coding using Zerotree of Wavelet Coefficient (EZW) (Jerome, 1993) is a simplest and efficient image compression technique where stream of bits are created in the order of importance, thus producing a completely embedded code. An embedded code defines a string of binary decisions that differentiates an image from a null or gray image. In EZW, the encoding and decoding process can be terminated when the target rate is achieved. The target can be a bit count, when the required bit count is reached, then the encoding process will be terminated. Likewise,

the decoding process also be discontinued at a certain point and reconstruct the image according to the lower rate encoding. The absence of prior training and precision rate control is the major advantage of this method. The performance of EZW is compared to JPEG in terms of CR, PSNR, MSE and significant coefficient. It is used in various fields like progressive transmission, browsing of images, multimedia applications, etc.

Another image compression technique is proposed in (Rao and Eswaran, 1996), where two simple algorithms for block truncation coding (BTC) are presented. BTC is a simpler and faster way to implement the image coding algorithm. These algorithms lead to a lower bit rate of 2 bits/pixel with no loss in image quality. The algorithm 1 uses one-bit adaptive quantizer and algorithm 2 uses linear quantization technique which results in the additional reduction in the bit rate. It performs well in terms of bit rate, MSE and quality of reconstructed image compared to BTC and Modified BTC (MBTC). The proposed algorithm produces comparable results with less complexity than hybrid algorithms.

Luo et al. (1996) proposed a method for artifact reduction of low bit rate DCT compressed images. To overcome artifact reduction problem, the proposed method uses a two-step process namely DC and iterative conditional mode (ICM). This method is evaluated on gray scale images, document images and resolution charts. The performance metrics used to investigate the proposed method are bit rate and PSNR.

The artifact problem is addressed by the use of post-processing technique at the decoding side (Wang and Zhang, 1998). This approach enhances the visual quality of the decompressed image with no additional bits or any changes in the encoder or decoder. The test image is first coded with JPEG standard and then blocking effect algorithm is applied. This method is compared with some existing methods include Projection onto Convex Sets (POCS), constrained least squares (CLS) and Crouse method with respect to PSNR and visual quality of the image.

A new chain code algorithm is introduced with an eight direction Freeman code (Liu & Zalik, 2005) in which every individual in the chain is encoded as the relative angle difference between the present and past elements. Statistical analysis reveals that the probability of occurrence of the chain code significantly varies and Huffman coding is utilized to compress it. This method is compared to Freeman chain code in eight directions (FCCED), Freeman chain code in four directions (FECFD), Derivative of the Freeman chain (DFC) and Vertex Chain Code (VCC). The comparison results show that proposed method achieves least bits per code and has shortest code length.

Another chain based lossless compression method for bi-level images is proposed in (Hermilo and Rodríguez-Dagnino, 2005). The chain code is a way of contour coding with three symbols. This technique is useful to compress 2-D binary object shapes and is composed of representing the orthogonal direction variations of the discrete contour. This method is highly applicable to represent bi-level images. It is compared with Freeman code, ZIP and JBIG (Joint Bi-level Image Experts Group) compressors in terms of compression rate. This method produces 25% and 29% better results than Freeman chain code and JBIG.

A new image encryption method is incorporated with compression in (Zhu et al., 2013). The encryption method is based on a 2D hyper-chaos discrete non-linear dynamic system. The 2D hyper-chaos is used to create two hyper-chaos sequences and they are used to mix up the position of the pixels in the original image. Using this method, the image is confused and a shuffled image is produced. Additionally, Chinese remainder theorem is utilized to perform image compression. This scheme encrypts and compresses the images simultaneously with no extra operation. Several analyses are done to validate the efficiency of the proposed method. The proposed method has very low correlation coefficient, stable and

secure against entropy attacks. This method passes NIST SP800-22 tests and implies that the encrypted image is stochastic and results to better encryption quality when compared to A5/1 and W7. The proposed method is much faster than classic DES algorithm and it can be used in multimedia applications.

A new technique is presented in (Zhang et al., 2013), which achieves equivalency between Reversible Data Hiding (RDH) and lossless DC. RDH is employed in medical images, military imaging and forensic sciences. In this paper, a histogram modification of RDH is done which embeds the message iteratively in the compression and decompression process of an entropy encoder. The proposed method is highly efficient for lossless DC (LDC), which basically achieves equivalency between RDH and LDC. The proposed method is compared with reversible watermarking and Tailored reversible watermarking schemes. PSNR is used to analyze the performance of this method.

Predictive data coding is the popular lossless DC where the future values are predicted by the help of past and present values. An effective and highly adaptive predictor is proposed in (Rad et al., 2013) which is a new gradient-based tracking and adaptive method (GBTa) that follow edges in many directions accurately. By examining the neighboring pixels of an unknown pixel, maximum intensity change level is determined to estimate edges. This method utilizes 11 directions to find out the pixel value. The proposed method is validated with 6 set of images using MAE and is widely used in multimedia files. This method produces less prediction error and higher quality than median adaptive prediction (MED), Gradient adjusted prediction (GAP), Gradient based selective weighting (GBSW), Graham and DARC.

A lossless hardware oriented image compression method is proposed in (Yin and Hu, 2014). It is used to compress the image pixels prior to writing in Synchronous Dynamic Random Access Memory (SDRAM) and decompress it after reading out from SDRAM. Huffman based variable length encoding (VLC) tables are designed to encode the prediction mode syntax elements. This method gives better tradeoff between prediction accuracy and control overhead cost. It is compared with two existing methods using 35 standard test images and it reduces the average number of bits significantly than existing methods.

A new lossless compression technique is developed for discrete-color images like maps, graphics, GIS and binary images (Alzahir and Borici, 2015). It contains two elements: fixed-size codebook and row-column reduction coding. The proposed method is tested on two kinds of images: images with a predefined number of discrete colors and binary images. The obtained results prove that the proposed approach achieves better performance than JBIG.

Improved RLC (IRLC) method for gray scale images are proposed in (Anantha Babu et al., 2016) to overcome the drawbacks of Run Length Coding (RLC). RLC works well for binary images but it does not achieve better CR for non-repetitive pixels. In order to achieve higher compression performance, IRLC algorithm is introduced which decomposes the original image to a number of smaller size non-overlapping blocks and the sub-blocks are organized in a matrix format. The block matrix elements are rearranged in a row-wise sequential number. It eliminates the redundancy by using a simple mapped pair which can be represented as follows:

$$\text{Mapped pair} = (C_i, L_i), (C_{i+1}, L_{i+1}), \dots, (C_n, L_n) \quad (10)$$

where $i = 1, 2, \dots, n$, C_i represents the intensity of pixels and L_i indicates the number of consecutive pixels of same intensity. The proposed approach is implemented for binary gray scale images and is compared with well-known compression techniques include Location based, Huffman coding and DCT. The performance metrics used for analysis are RMSE, SNR, PSNR, CR, bit rate, compressed file size and computation time. Because of optimal and compact code,

IRLC can be used in TIFF, GIF, JPEG and Textual files. But, the high computational complexity is the drawback of IRLC.

This paper proposes a Bi-level burrows wheeler compression algorithm (BWCA) to achieve robustness and compression efficiency for various types of images (Khan et al., 2016). The increased inter-pixel redundancies from two pass BWT stage and reversible color transform (RCT) are used. Initially, the proposed method (RCT-BWCA) was tested on PNG format images from Kodak and it achieves better compression than JPEG-2000 and BWCA. This method is also tested with color images, color filter array (CFA) images, Electroencephalography (EEG) data and Raster images. The proposed method produces 18.8% better compression when compared to High Efficiency Video Coding (HEVC) and 21.2% more effective than LZ4X compressor for Kodak color images.

A lossless encryption then compression (ETC) technique is developed to achieve higher compression performance while ensuring the security of the images (Kumar and Vaish, 2017). Initially, the original image is segmented into significant and less significant parts using 2D DWT. This part results to DWT approximation (LL) and detailed wavelet sub-bands (LH, HL, HH) where L represents low and H represents high. After decomposition, LL is encrypted using pseudo random numbers which are generated by Pseudo Random Number Generator (PRNG) series. After encrypting LL, the detailed sub-band LH, HL HH is encrypted and compressed efficiently. The coefficients of approximation sub-bands are termed as significant and detailed sub-bands are treated as less significant. LL is encrypted using addition modulo 256 and detailed sub-bands are encrypted using coefficient permutation. After encryption, these bands are arranged as LL, LH, HL and HH. The channel provides an effective technique to compress the bands individually. First, the encrypted sub-bands are quantized using SVD and lossless compression is attained using Huffman coding. The performance is evaluated using 8-bit gray scale images and it is compared to JPEG in terms of CR and PSNR.

4.3.3. Audio compression

IEEE 1857.2 is the recently developed standard for advanced audio coding (AAC) (Auristin, 2016). It is an efficient lossless audio compression technique specifically designed to achieve better audio quality with optimized bandwidth and higher speed. It involves a collection of tools to attain specific audio coding functions. In this method, encoding is done by Linear Prediction Coding (LPC) followed by preprocessing and entropy coding block. The performance is analyzed in both music and speech audio file. It produces better compression and faster rate of encoding and decoding process.

Another audio compression method based on scalable variable bit rate encoding method is proposed in (Hang et al., 2016) which is adaptable to bandwidth variations. An audio attention model is used to determine the attention of every audio sample which results in the determination of attention presence. A new method is proposed in (Jain et al., 2017) to produce original audio signals after the compression process. The proposed method uses LZ compression over Intrinsic Mode Functions (IMF) produced by Empirical Mode Decomposition (EMD). The proposed method produces better results than DCT and DWT in terms of CR, SNR, PSNR and Normalized RMSE (NRMSE). This method is highly useful for audio–video conferencing, Voice over IP (VoIP) services, Voice over LAN (VoWLAN), etc.

4.3.4. Video compression

A video is also an essential part of multimedia applications. Generally, video files consume more resources for communication, processing, and storage purposes. So, compression is much needed for video files to store, process or transmit it. Several techniques were developed to efficiently compress video files to avoid large

amount of data being transmitted or stored. In this section, existing video compression techniques are discussed.

H.261 standard has been proposed to transmit video at a rate of 64kbps and its multiples. The frames of H.261 can be classified into two types namely Intra coded frames (I-frames) and Predicted-coded (P-frames). In I-frames, the frames are encoded with no dependency on the previous frames whereas the frames are coded using past frames in P-frames. H.261 is identical to JPEG compression standard and it employs motion-compensated temporal prediction. It is widely used for video calling and video conferencing.

H.263 is same as H.261 but it is specially designed for lower bitrates. The image is partitioned into a number of macro-blocks. Macro block has 16×16 luminance blocks and 8×8 chrominance blocks. The macro blocks are encoded as intra or inter blocks. Spatial redundancy is utilized by DCT coding and temporal redundancy by motion compensation. H.263 contains motion compensation with half-pixel accuracy and bi-directional coded macro blocks. As H.263 is developed for low bit rate applications, these characteristics are not utilized in MPEG-1 and MPEG-2.

Moving Pictures Experts Group (MPEG) is an ISO/IEC working group which develops compression techniques, representation of motion pictures and audio in international standards (Watkinson, 2004). MPEG is a layered, DCT based video compression standards leads to VHS quality compressed video stream with a bit rate of nearly 1.5Mbps at a resolution of nearly 352×240 . MPEG video sequences comprised of different layers which can randomly access a video sequence and protects it against corrupted data. MPEG frames can be coded in three ways: I-frames, P-frames and Bi-directionally-predictive-coded or B-frames. I-frames are coded as discrete frames which are not dependent on the previous frames. P-frames are coded to smaller frame size than I-frames. B-frames need a previous as well as the future frame (P-frames and I-frames) for decoding purposes. MPEG-2 is mainly developed for various applications which provide a bit rate of 100Mbps. It is used in digital high-definition television (HDTV), interactive storage media (ISM), cable TV (CATV), etc. Decoding in MPEG-2 is very expensive; bit stream scalability provides flexibility in the required processing power for decoding. Video sequence layers are same as MPEG with some improvements. MPEG-3 has been developed for HDTV and Enhanced definition television (EDTV) for higher bit rates and it later combined with MPEG-2. MPEG-4 enables the user to interact with the objects inside the boundaries. It takes multimedia to lower bit rate networks and utilizes media objects to specify audio as well as visual content. Media objects are correlated to produce compound media objects. Prior to transmission, MPEG-4 synchronizes and multiplexes the media objects to attain QoS and at receiver's machine, it permits interaction with the constructed scene. MPEG-4 arranges the media objects in a hierarchical form, where lower level has primary media objects to indicate 2 or 3-dimensional media objects. Recovery tools are employed to retrieve the lost data, after the synchronization of data. Error concealment tools are utilized to cover up the lost data. Effective resynchronization is an important component for better data recovery and error concealment. MPEG-7 is mainly used to represent a collection of descriptors to explain different forms of multimedia. Also, it regulates other descriptors, structures for the descriptors and relationship between them. This information will associate with the content to enable quicker and effective search.

H.264 is a new video compression standard aimed to attain higher quality video at a lower bit rate developed by the ITU-T Video Coding Experts Group (VCEG) together with the ISO/IEC JTC1 Moving Picture Experts Group (MPEG). It is a block-oriented motion-compensation-based video compression standard. It offers significantly lower bit rates than existing standards like MPEG-2, H.263, or MPEG-4 Part 2. An additional goal is to provide flexibility to various applications on a wide variety of networks and systems,

includes low and high bit rates, low and high resolution video, broadcast, DVD storage, RTP/IP packet networks, and ITU-T multimedia telephony systems. It is basically used for lossy compression, although it is also possible to generate truly lossless-coded regions inside lossy-coded pictures or to maintain rare use cases for which the whole encoding is lossless. H.264 provides same Digital Satellite TV quality as present MPEG-2 implementations with less than half the bitrate, with present MPEG-2 implementations working at around 3.5 Mbit/s and H.264 at only 1.5 Mbit/s.

High Efficiency Video Coding (HEVC), also called as H.265 and MPEG-H Part 2, is a video compression standard potential successors to the widely used AVC (H.264 or MPEG-4 Part 10). HEVC is designed with the goal of providing video content with higher coding efficiency at low bit rate. On comparing with AVC, HEVC offers double the CR at the equivalent level of video quality, or considerably enhanced video quality at the same bit rate. It supports resolutions up to a maximum of 8192×4320 . The aim of HEVC is to provide the similar level of picture quality as AVC, but with better CR when compared to H.264/MPEG-4 AVC HP.

4.4. Based on application

Although some of the common compression techniques can be applicable to preferred applications, the concentration has been on the technique rather than the application. However, there are certain techniques where it is impossible to separate the technique from the application (Sayood, 2000). This is due to the fact that several techniques depend upon the nature of data involved in the application. In this section, the compression techniques developed for specific applications are discussed, the classification hierarchy is shown in Fig. 4 and the reviewed approaches are tabulated in Table 5.

4.4.1. Wireless sensor networks

As sensor nodes are energy constrained in WSN, compression techniques play a major role to minimize the energy consumption. DC techniques are used to minimize the amount of data and thereby reduce the number of data transmission to preserve

energy. In this section, DC techniques developed for WSN is reviewed.

A simple data compression scheme particularly suitable for reduced storage space and low computational resource of WSN is proposed (Marcelloni and Vecchio, 2008). The compression algorithm utilizes the high correlation between successive samples measured by sensor node. Entropy compression enables to produce compressed data of every sensed value, with the help of smaller dictionary, whose size is estimated by the resolution of the ADC converter. Huffman coding is employed to encode the data and CR assessed the effectiveness of the proposed method. It achieves compression performance of 66.99% and 67.33% for temperature and relative humidity values collected by the sensor node. It is compared with Sensor-LZW (S-LZW), gzip and bzip2.

Next, an Adaptive Lossless Data Compression scheme (ALDC) for WSN is proposed (Kolo et al., 2012). ALDC is a light weight lossless compression technique with multiple code options. It performs compression based on two code options namely 2 Huffman table ALDC and 3 Huffman table ALDC. These tables are designed after observing various real time WSN datasets with various levels of correlation. ALDC uses predictive coding for better capturing of underlying temporal correlations present in the sampled data of surveillance applications. In prediction coding, linear or non-linear prediction models are employed followed by a number of coding approaches. The proposed system can be applicable for real time as well as delay insensitive situations. It achieves 74.02% better compression for real-world datasets when compared with LEC and S-LZW.

LZW coding is employed to decrease the energy consumption and maximizing the network lifetime (Ruxanayasmin and Krishna, 2013). LZW compresses the file 1/3 of its original size. The performance of LZW is compared with Huffman and RLE in terms of CR where LZW attains a maximum CR for different file formats such as text, speech and images. It is widely used in mobile adhoc networks.

A new lossless DC technique named Sequential Lossless Entropy Compression (S-LEC) to achieve robustness in WSN is proposed (Liang and Li, 2014). S-LEC is the extension of LEC by adding sequential coding technique. To assess the performance of S-LEC,

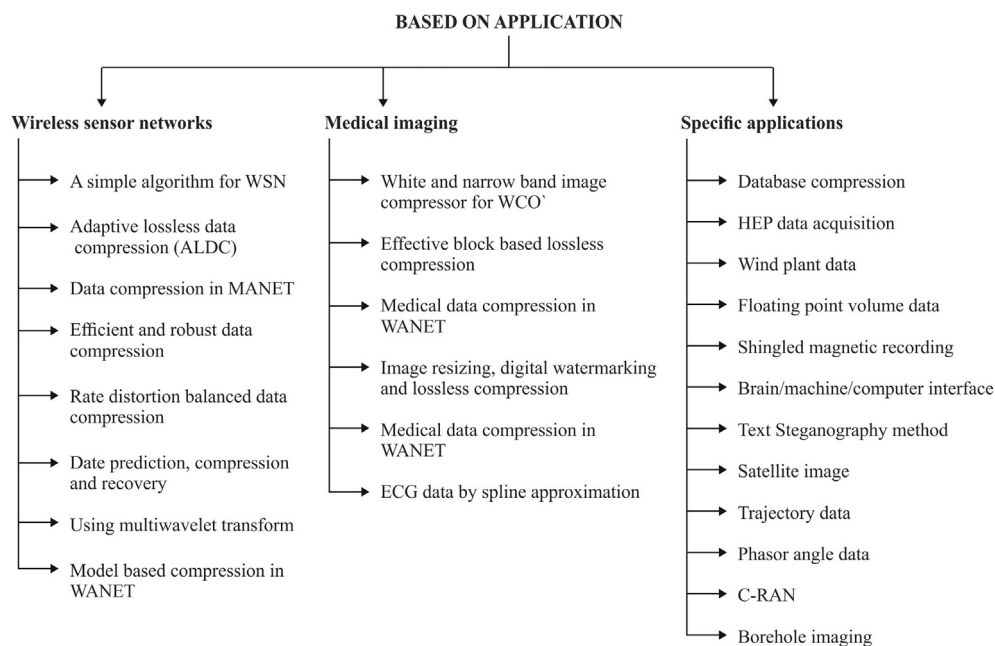


Fig. 4. Classification of compression techniques based on application.

Table 5
Comparison of various compression techniques based on applications.

Reference	Objective	Compression Type	Application domain	Application	Data type	Methodology	Performance metrics	Compared with
Marcelloni and Vecchio (2008)	To develop a simple data compression algorithm	Lossless	WSN	Collect signals from various calorimeters, TPC, detectors	ALICE TPC dataset	Huffman coding	Compression factor, compression error	S-LZW, gzip, bzip2
Kolo et al. (2012)	Lightweight adaptive data compression algorithm	Lossless	WSN	Real-time and delay tolerant applications	Real world data	2 Huffman table ALEC, 3 Huffman table ALEC	CR	LEC, S-LZW
Ruxanayasmin and Krishna (2013)	To reduce energy consumption and maximize network lifetime	Lossless	WSN	Mobile Adhoc Networks	Text, speech and image	LZW	CR	Huffman, RLE
Incebacak et al. (2015)	To provide contextual privacy with lower energy consumption in WSN	–	WSN	Privacy protection in WSN	–	Mathematical programming framework	Packet error probability, normalized lifetime	OC, OSLC, LC and AC
Abu Alsheikh et al. (2016)	To provide error bound guarantee	Lossy	WSN	WSN	Metrological data	Neural networks	CR, RMSE, R^2	PCA, DCT, FFT, CS
Wu et al. (2016)	Compression using data prediction with guaranteed delay	Lossless	WSN	Continuous monitoring application of WSN	Sensor data	LMS data prediction PCA technique	CR	AZTEC, peak pitching method, linear prediction method
Rajakumar and Arivoli (2016)	Integer Multi-Wavelet Transform (IMWT)	Lossy	Wireless networks	Remote sensing images, forensic images	Images	Magnitude set coding RLE	PSNR, bpp, visual quality	JPEG, AIC, JPEG 2000
Way (2003)	To reduce the size of the routing table	Lossless	WANET	GPS	Trajectory data	–	–	–
Khan and Wahid (2014)	To develop a lossless image compressor for endoscopic images	Lossless	Medical imaging	WBI and NBI endoscopic images, video sequences	Endoscopic images	Lossless JPEG based predictor Golomb-rice coding	CR, PSNR, VIF, SSIM, VSNR	–
Venugopal et al. (2016)	To achieve better compression performance	Lossless	Medical imaging	Telemedicine applications	Natural images and medical images	Hadamard transform and Huffman coding	CR	JPEG2000
Dutta (2015)	To reduce the size of the medical data of the patients	Lossless	Medical imaging	Transmitting medical images in WANET	Endoscopic images, ECG, magnetic resonance	DCT, AGR encoder, adaptive edge based fuzzy filtering	CR, Visual quality	JPEG, JPEG 2000
Amri et al. (2017)	To propose a compression technique named WREPro. TIFF and WREPro.JLS	Lossless	Medical imaging	Telemedicine applications	Medical images	TIFF or JLS	PSNR, SSIM	JPEG
Venugopal et al. (2016)	To reduce the number of bits transmitted from neural implant	Lossless	Medical imaging	Brain Machine/Computer interface	Neural signal	Hadamard transform RLE	SNR, RMS, CR	EMBC, TBioCAS
Nielsen et al. (2006b)	To minimize the distortion rate of the compressed signal	Lossless	Medical imaging	Bio medical signal compression	electromyographic signals	wavelet based signal compression technique and EZW coding	distortion rate, CR	–
Nielsen et al. (2006a)	To present an efficient signal compression technique using signal-dependent wavelets	Lossless	Medical imaging	Bio medical signal compression	electromyographic signals	wavelet based signal compression technique and EZW coding	distortion rate, CR	–
Brechet et al. (2007)	To propose a signal compression using DWPT	Lossless	Medical imaging	Bio medical signal compression	electromyographic and ECG signals	Modified EZW coding	CR	–
Ranjeet et al. (2011)	To compare the performance of transform coding technique	Lossless	Medical imaging	Bio medical signal compression	ECG signals	DCT, DWT AND FFT	PRD, ME, MSE, PSNR, CR	–
Vadori et al. (2016)	To present SAM technique for physiological quasi-periodic signals	Lossless	Medical imaging	Bio medical signal compression	PPG, ECG and RESP signals.	subject-adaptive dictionary	CR, RMSE	–
Lee et al. (2006)	To compress database	Lossless	Application oriented	Database compression	Database	Association rule mining	CR	Apriori and TDC
Patauner et al. (2011)	To compress data from TPC of the ALICE	Lossless	Application oriented	HEP data acquisition	Data from pulse digitizing electronics	Delta calculation Huffman coding	–	–
Louie and Miguel (2012)	To compress wind plant monitoring data	Lossless	Application oriented	Wind plant data compression	Wind turbine data	Shannon entropy	CR	BZIP2, Deflate, LMZA
Fout and Ma (2012)	To compress floating point data	Lossless	Application oriented	Scientific and medical floating-point data	Floating point data	APE, ACE	CR, PSNR	JPEG

Table 5 (continued)

Reference	Objective	Compression Type	Application domain	Application	Data type	Methodology	Performance metrics	Compared with
Venkataraman, et al. (2013)	To reduce ITI read latency	Lossless	Application oriented	Shingled recording drives	-	LZSS	CR, SNR, read latency and average read latency	-
Satir and Isik (2014)	To improve security and capacity of text stenography	Lossless	Application oriented	Secured stenography	text	Huffman coding	Capacity	Textor, Mimic, Listega
Muthukumar and Ravi (2014)	To provide precision and less bandwidth	Lossless	Application oriented	Remote sensing image compression	Remote sensing images	Wavelet based coding	-	-
Mahmood et al. (2014)	To modify LZ algorithms using DNA behavior	Lossless	Application oriented	-	Text	Bi-directional LZ77, LZ88 and LZW84	CR	-
Nibali and He (2015)	To compress trajectory data	Lossless	Application oriented	GPS based tracking data	Trajectory data	Residual encoding scheme	CR, compression time	TD-SED line generating TD-SED + delta
Cheng et al. (2015)	To compress DNA sequences	Lossless	Application oriented	Medical	DNA sequences	Intra and inter sequence similarities	CR, bpb	RLCSA, PLZ-opt, GDC, FRESKO, iDoComp
You (2016)	To achieve near lossless codec	Lossless	Application oriented	Front haul C-RAN	Wireless communication signal	Removal of LSB CR Constraint	CR, latency, complexity	-
Uthayakumar et al. (2017a)	To compress seismic data	Lossless	Application oriented	WSN	Seismic data	BWT coding	CR, CF and bpc	LZMA, Huffman, LZW and AC
Uthayakumar et al. (2017b)	To compress wind plant data	Lossless	Application oriented	WSSN	Wind plant data	LZMA	CR, CF and bpc	BWT, Huffman and AC
Fan et al. (2017)	To compress acoustic signal in borehole imaging	Lossless	Application oriented	Borehole imaging	Acoustic signal	SOE-MP	CR, CT	SOE-SAGE SOE-CSD

it is compared with LEC and S-LZW using real word dataset from sensor scope and volcanic monitoring. The performance metrics used to analyze the results are CR and energy consumption of the sensor data. S-LEC achieves reduced energy consumption for dynamic volcano dataset when compared to LEC.

Several DC techniques are studied in (Incebacak et al., 2015) to enhance the lifetime of WSN working in stealth mode. A mathematical programming framework is generated to analyze the advantages of DC approaches to maximize the network lifetime. This study revealed that the DC technique can save the energy consumption to provide contextual privacy in WSN.

A low cost, lossy compression approach with error bound guarantee for WSN is proposed in (Abu Alsheikh et al., 2016). The proposed method reduces data congestion and minimizes energy dissipation. This is achieved by the utilization of spatio-temporal correlation between data samples. It is based on neural networks, a machine learning algorithm to automatically predict human activities and environmental conditions. This method eliminates the power and bandwidth constraints while gathering data in the tolerable error margin. The algorithm is tested with meteorological datasets and produces better results than Principal Component Analysis (PCA), DCT and Fast Fourier Transform (FFT) in terms of CR, RMSE and Coefficient of determination (R^2).

In (Wu et al., 2016), data prediction, compression and recovery techniques are integrated to develop a novel energy efficient framework for clustered WSN. The major intention is to minimize the amount of data transmission using data prediction and compression with guaranteed delay. Least Mean Square (LMS) data prediction algorithm is used with optimal step size by decreasing the Mean Square Deviation (MSD). PCA technique is used for compression and recovery of predicted data. In clustered WSN, the sensor nodes are organized into clusters and parallel dual prediction process is implemented at sensor nodes and cluster heads. Then, CHs filters the principal component using PCA technique to avoid redundant data. Finally, data is completely recovered at the base station. The proposed method is tested with public sensor dataset. The results prove that the proposed method is found to be efficient for continuous monitoring applications of WSN. Prediction algorithm uses prediction accuracy, convergence speed and communication reduction to evaluate the performance. Compression algorithm uses communication cost as the performance metric and it improves both accuracy and convergence rate.

Integer Multiwavelet Transform (IMWT) based lossy compression technique is proposed (Rajakumar and Arivoli, 2016). The image is transformed using IMWT whereas compression is achieved by magnitude set coding and RLE. IMWT is a block transform, effectively implemented with bit shift and additional operations. IMWT gives coefficients of both positive and negative magnitudes. These coefficients have to be coded in a proper way to produce better CR.

4.4.2. Medical imaging

A huge amount of medical image sequences and patient data are maintained in various hospitals and medical organizations. It occupies more storage space and also consumes more time for transmission. The compression technique is employed to compress medical data and is very useful in telemedicine. In this section, various compression techniques developed for medical images are reviewed.

Wireless capsule endoscopy (WCO) or video capsule endoscopy (VCO) is used to capture the image of the human intestine for medical diagnosis. Once the data is captured, the images are wirelessly transmitted to the recorder in the external world. The recorder forwards the data to a personal computer (PC) where the images are reconstructed and displayed for diagnosis. This paper presents a new image compressor for WCO that supports both WBI (Wide

Band Imaging) and NBI (Narrow Band Imaging) modes (Khan and Wahid, 2014). So, it is called as “Dual Band Imaging”. The proposed method involves two components: Lossless JPEG based predictor and Golomb-rice (GR) coding for the encoding process. The proposed image compressor can be directly interfaced to any digital-video-port (DVP) compatible commercial RGB image sensor which produces output pixels in a raster scan manner, avoiding the requirements of buffer memory and temporary storage. The reconstructed images are verified by medical doctors for acceptability. To evaluate this method, 100 WBI images, 15 NBI images and 1698 frames of 7 different endoscopic video sequences are used. It produces low complexity, negligible memory, hardware cost, latency and power consumption. CR, PSNR, VIF (Visual Information Fidelity), SSIM and VSNR (visual SNR) are used as performance metrics. It achieves a CR of 80.4% (WBI), 79.2% (NBI) and PSNR of 43.7 dB.

Another image compression method is developed to achieve better CR for medical images (Venugopal et al., 2016). It is a block-based method which uses Transform and Huffman coding. The input image is divided into LL, LH, HL and HH sub-bands with the help of Integer Wavelet Transform (IWT). Then, lossless Hadamard Transformation (LHT) is employed to remove the correlation within the block. DC Prediction (DCP) eliminates the correlation between the adjacent blocks. The non-LL sub-bands are evaluated for Non-Transformed Block (NTB) using a threshold value. After the process of validation, the non-truncated non-LL bands are encoded directly. LL and validated non-LL bands are truncated and coded by lossless Huffman compression. It produces better compression when compared to the existing methods for various sizes and different resolutions. The results are validated by different medical officers include surgical, general medicine and pediatrics. They confirmed that the results are within the acceptable standards. This method is very effective for storing medical data and telemedicine applications.

A novel medical data compression (MDC) method is employed for the reduction in size of the patient's medical data (MDP) (Dutta, 2015). The compression and decompression are done at Primary Health Care (PHC) and community care centre (CC). The compression is divided into two levels: Image transformation and encoding of coefficients. The compressed data is transmitted from PHC to CC center using WANET. Because of transformation, quantization and transmission errors, blocking artifacts, additive noises and multiplicative noise may present. To eliminate the noise during decompression, an adaptive edge-based fuzzy filtering technique is used. A Fuzzy based Route Selection (FRS) is also proposed to find the optimal path from PHC to CC. Fuzzy logic uses input parameters as energy consumption, residual energy and routing delay. The probability of selecting a routing path is the output parameter of fuzzy logic. The path with higher probability has more chance of becoming the routing path. Once a path is established, the compressed image is transmitted from PHC to CC center efficiently. The proposed method is compared with JPEG, JPEG2000, Turcra and Duplaga in terms of CR visual quality of medical data of patients and robustness against signal transmission errors. This method is used to compress endoscopic images, ECG images, magnetic resonance imaging, etc.

Amri et al. (2017) proposed a compression technique named Watermarked Reduction/Expansion Protocol integrated to TIFF format (WREPro.TIFF) and Watermarked Reduction/Expansion Protocol combined with JPEG.LS format (WREPro.JLS) for medical images. In this approach, four techniques such as expansion technique, digital watermarking and lossless compression standards like JPEG-Ls and TIFF format are employed. When compared with original JPEG, proposed method achieves better compression in terms of PSNR and SSIM.

Computerized electrocardiogram (ECG), electroencephalogram (EEG), electromyographic (EMG) and magnetoencephalogram (MEG) processing systems are commonly employed in hospitals to record and process biomedical signals. These systems lead to the generation of larger size signal databases for successive investigation and comparisons. It also makes harder to transmit the biomedical information possible over telecommunication networks in real time. Biomedical signal compression methods are developed to reduce the storage space with no loss of clinically significant information, which is attained by the elimination of redundant data in the signal at a reasonable rate.

Nielsen et al. (2006b) proposed a wavelet based signal compression technique and employed to surface EMG signals recorded from ten subjects. It introduces a signal-based optimization of the mother wavelet to minimize the distortion rate of the compressed signal given a target compression rate. An unconstrained parameterization of the wavelet is used for wavelet optimization. The wavelet coefficients are encoded by EZW coding algorithm. The results of the proposed method is investigated by the use of distortion rate and CR. Nielsen et al. (2006a) presented an efficient signal compression technique using signal-dependent wavelets. To adapt the mother wavelet to the signal for the purpose of compression, it is essential to represent a family of wavelets which depends upon the set of parameters and quality criteria for wavelet selection (i.e., wavelet parameter optimization). The orthogonal wavelets are parameterized by a scaling filter, with optimization criterion using the minimization of signal distortion rate given the desired compression rate. The wavelet coefficients are encoded by EZW coding algorithm. The proposed method is employed to EMG signals to validate its performance. The result of the proposed method is investigated by the use of distortion rate and CR. Brechet et al. (2007) introduced a signal compression using discrete wavelet packet transform (DWPT) decomposition. The mother wavelet and the basis of wavelet packets are optimized, and then the wavelet coefficients are encoded by modified EZW coding algorithm. This signal dependent compression scheme was designed by a two-step process: best basis selection and selection of the mother wavelet. This method is validated by two sets of ten EMG and ECG signals which are compressed with CRs of 50%–90%. The proposed method leads to significant enhancement in signal compression based on DWT and random selection of the mother wavelet. The method provides an adaptive approach for optimal signal representation for compression and can employed to any type of biomedical signal. In (Ranjeet et al., 2011), the author compared the performance of transform based coding techniques while compressing ECG signal. This paper employed DCT, DWT and FFT techniques to analyze its performance based on CR, MSE, SNR, Percent root mean square difference (PRD) and Maximum error (ME). The experimental results demonstrate the efficiency of these transforms in biomedical signal compression. The results revealed that DCT and FFT attain better CR. In addition, DWT produces good fidelity parameters with comparable CR. Vadori et al. (2016) presented a subject-Adaptive (lossy) compression (SAM) technique for physiological quasi-periodic signals. It leads to significant minimization of data reduction and allows effective archival or communication. SAM is based on a subject-adaptive dictionary, which is learned and refined upon execution to exploit the time-adaptive self-organizing map (TASOM) unsupervised learning algorithm. The simulation results reported that the proposed method is superior to existing methods in terms of CR and RMSE.

4.4.3. Specific applications

A new database compression method is proposed to compress large databases using association rule mining (Lee et al., 2006). Association rule mining can identify the frequent itemsets in the

database relations. The extracted association rules represent the repeated itemsets in the database and they are compressed to decrease the file size. The proposed method is compared with Apriori algorithm and it achieves better CR than existing methods. Though the present compression approaches are limited by languages, the proposed method based on data mining techniques are not limited by language. The computational cost can also be improved by the use of advanced data mining technologies.

A lossless DC algorithm is proposed to compress the data from pulse digitizing electronics (Patauner et al., 2011). This is used to compress real-time data from time projection chamber (TPC) of the ALICE project (A large Ion Collider experiment). This algorithm consists of three steps namely vector quantization, delta calculation and Huffman encoding. For TPC data, the sample values are correlated based on shape of the pulses. Vector quantization is used to correlate the input data using the pulse shape. To achieve lossless compression, delta computation is done where the difference between the samples in the original input and selected reference vector are computed. As the delta values are smaller numbers, it is compressed by lossless Huffman coding technique. The higher occurrence of delta values is coded with least bits and lower occurrences are coded with many bits. This method is compared with Huffman and arithmetic coding in terms of CR and compression error. The real time application ALICE-TPC dataset with 10,000 pulses are used to evaluate the efficiency of the proposed method.

A lossless compression algorithm for point and gridded wind plant data is developed and analyzed in (Louie and Miguel, 2012). This algorithm is useful for data involved in wind plant monitoring and operation. The algorithm uses wind speed-to-wind power relationship, temporal and spatial correlations in the data. The zero and first order Shannon entropy of wind power data sets were calculated to gain uncertainty of wind power data. A preprocessing algorithm is also proposed for gridded and point dataset with respect to wind plants of different capacities with various levels of precision. The preprocessing algorithm provides better CR up to 50% off the shelf compression of raw data.

Floating point compression is very difficult than integer compression as the Least Significant Bit (LSB) of the massive data are less correlated. An adaptive prediction based lossless compression method for floating point data is proposed (Fout and Ma, 2012). This is done by the use of switched predictor which chooses the best predictor from a smaller set of candidate predictor for encoding at a predefined time. This method is flexible for varying data

loss to a particular level, free storage space will become available to store Elliptic Curve Cryptography (ECC) redundancy. This extra space can be used to store a stronger ECC than normal ECC for the current sector. This leads to decrease the probability of reading one or multiple neighboring tracks for explicit ITI compensation. Lempel–Ziv–Storcs–Symanski (LZSS) algorithm is used for compressing several types of data stored in hard disks. This method uses intra-sector DC method to minimize the read latency is shingled recording. Here, lossless DC technique is employed on multiple physically consecutive sectors on the same track to increase the compression efficiency. It results in the reduction of read latency overhead by enabling a stronger ECC and is termed as virtual sector compression. Because of overlapping writing in shingled recording, update-in-place feature fails to work. This makes the virtual sector compression practically possible. Two scenarios (fixed size and context aware virtual sector) are used to investigate the results of virtual sector compression. Intra-sector lossless compression results in the decrease in disk read latency by 40% with the conventional design. Virtual sector compression achieves the reduction in read latency by 39%.

Walsh-Hadamard transform (WHT) is used for DC in brain-machine/computer interfaces. Hosseini-nejad et al. (2014) designed a DC method to decrease the number of bits transmitted from the neural implant to the real world using WHT. The performance metrics used to analyze the results are SNR, RMSE and CR. This method is compared with EMBC, TBioCAS and the proposed method is implemented as a 128-channel WHT processor.

To improve the security and capacity of text steganography, DC algorithm is proposed (Satir and Isik, 2014). Lossless Huffman coding is used to compress the text data. It reads the source data and generates a coding tree based on the produced probability model. For improving the security, two classes of stenography keys are used based on the requirement. One key is useful in the embedding phase of the proposed method. The second key is the global stegno key which shared between sender and receiver in advance. Combinatory based coding is also used to improve security because it provides desired randomness. However, Huffman coding also contributes to security in addition to increase in capacity. Email is chosen as the communication channel between two parties. The proposed method is evaluated using the term 'capacity'. Bit rate or capacity is the ratio of size of secret message to size of stegno cover and is represented in Eq. (11). The average capacity is computed as 7.96% for the secret message with 360 characters (300:8).

$$\text{Capacity [character of the secret message]}_c = \frac{\text{Number of bits in secret message}}{\text{Number of bits in stegno cover}} \quad (11)$$

statistics by choosing the best predictor which is suitable for the local characteristic of data. This method can be extended and combined with existing or upcoming predictor without changing the basic framework. It supports progressive precision and provides data access up to three levels for achieving lossy, lower resolution option suitable for visualization preview. A faster entropy encoding is done by the residual leading zeros with the help of rank table and university codes. The proposed method produces better CR than existing methods namely Lovenzo, FCM/DFCM and Adaptive Polynomial Encoder (APE). This method is robust and faster, but the computational complexity is increased.

To reduce the Inter-Track Interference (ITI) read latency of Shingled Magnetic Recording (SMR), two lossless compression mechanisms are used (Venkataraman et al., 2013). It is a known fact that most of the files stored in disks can be compressed in a lossless manner. When a sector of user data can be compressed with no

Transmission of remote sensing images requires increased precision and minimum bandwidth. To compress remote sensing images, a wavelet based algorithm is proposed (Muthukumaran and Ravi, 2014). This method intends to achieve greater CR and better visual quality which is implemented by sub-band coding and decoding algorithm. The proposed method achieves better compression performance when compared to Aztec method, Peak pinching method, Linear prediction method, Artificial neural network and Fourier transform.

Mahmood et al. (2014) altered the traditional LZ77, LZ78 and LZW84 approaches for bi-directional data reading and investigated their performance among them. This idea is inspired from the nature of DNA, which reads enzymes in forward and reverse direction on opposite strands leading to diverse proteins. They analyzed the versions of LZ algorithms for variable length windows and variable number of bits are used to address the dictionaries. The attained

results depicted that better compression is possible when symmetries exist in the data needs to be compressed.

Cheng et al. (2015) proposed a compression algorithm to use intra-sequence and inter-sequence similarity to compress multiple DNA sequences. The subsequence matches are searched so that the proposed method is useful for sequences with various levels of similarities like partial or nearly identical similarities. The proposed method is compared with existing methods in terms of CT, decompression time and bits per base.

Trajic is an efficient compression technique to reduce the amount of memory needed to store trajectory data (Nibali and He, 2015). It achieves better CR with two core elements namely predictor and a residual encoding technique. The predictor computes the value of next data point and creates smaller residuals. The residual encoding scheme tries to employ fewer bits to encode the number of leading zeros in the residuals. It achieves lossy compression by eliminating a number of LSB. Trajic runs in linear time and makes it suitable for the database application. It produces better CR than delta and line generalization for maximum error bounds.

Phasor measurement units (PMU) are used to monitor and regulate power grids where devices record globally synchronized measurements of voltage and current signals. It provides the respective phasor magnitude and angles at a typical rate of 30 frames/s. To compress the phasor angle data, a new preprocessor is proposed (Tate, 2016). An adaptive high-performance entropy encoder called GR coding is used and two stages are also involved. In the first stage, a new preprocessor based on frequency compensated differential encoding is proposed. In the second stage, GR encoding is used. It is a lossless compression technique where the data transmission from space vessels takes place and higher throughput is achieved with limited memory and computational resources. Golomb encoding is designed to encode non-negative, integer valued signals in which probability value n decreases exponentially. It is very effective than existing coding techniques like Huffman coding and Arithmetic coding. Next, mapping of values to GR codes does not require any lookup tables or trees. GR coding compress 8 million phasor angles/s and it is very useful for applications requires less delay. The data rate, compression rate and Error Propagation Rate (EPR) are used to evaluate the performance of this technique.

For cloud- Radio Access Network (C-RAN), near lossless compression and decompression algorithm for digital data transported through front haul in C-RAN is developed (You, 2016). The compression is attained by the elimination of redundant data in wireless communication signals. Two near-lossless compression/decompression methods suitable for front haul digital data are presented. The proposed method uses two lossless compression methods namely simple compression by eliminating LSB and compression/decompression algorithm with a CR constraint. The processing time of this method is low and makes it more suitable for implementing in real time applications. CR, latency and complexity are used to analyze the performance of the proposed method. Through the proposed method, it can be easily extended to several applications like devices which capture or store sampled signals in a fixed amount of physical memory.

To efficiently and robustly compress Ultrasonic Borehole Imaging (UBI), a new compression method is proposed (Fan et al., 2017). In this method, Sum of Exponentials (SOE) is used to model the waveforms and Matrix Pencil (MP) algorithm is used for compression and recovery. The SOE-MP method is used to achieve the tradeoff between compression, computation and accuracy. This is due to the parameters of SOE model which can be efficiently and reliably computed using MP algorithm. The borehole or parallel scanning data consists of redundant waveforms. To eliminate redundancy and exploit the correlation between consecutive

waveforms, Angle Based Adjacent Basis Grouping (ABBG) is proposed. The SOE-MP method is evaluated and compared with SOE-SAGE and SOE-CSD. It achieves better accuracy in the decompressed image with higher speed and compact storage space. The online ABBG method utilizes the correlation among consecutive waveforms results to enhancement in computational complexity and compression performance. The performance metrics used to analyze the results are CR, RMSE and CT.

An energy efficient architecture is developed for on-the-fly DC and decompression whose field of operation is cache to memory path (Benini et al., 2004). The uncompressed cache lines undergo compression prior to writing in main memory and decompression takes place during cache refill. Two classes of table-based compression schemes are proposed. The first method is offline data profiling which is appropriate for the embedded system. The second method is adaptive data profiling where decisions are taken to select whether data words should be compressed or not. It depends upon the data statistics of the program being executed. The experimental results prove that the energy consumption is significantly reduced (39% for profile-driven and 26% for adaptive approach).

A new binary compression algorithm using Hamming code (HCDC) is introduced (Hussein, 2008). In HCDC, the binary sequence needs to be compressed is segmented into blocks of n -bits length. For using Hamming codes, the block is treated as Hamming codeword with 'p' parity bits and 'd' data bits ($n = p + d$). Then, each codeword is validated to determine whether the codeword is a valid Hamming codeword or not. In case of valid blocks, only d bits preceded by 1 are written to compressed file. For non-valid blocks, all n -bits are preceded by 0 are used to differentiate valid and non-valid blocks during the decompression process. For optimum CR, the length of each block is ' n ' ($n = 2^p - 1$). For valid codewords, parity bits are eliminated, then the data bits are extracted and written into a temporary file during compression. In decompression, the parity bits are recovered by Hamming codes. For non-valid blocks, no changes were done and blocks are written to the temporary file. A new analytical formula for CR is computed based on the block size and the fraction of valid data blocks in the sequence.

$$CR = \frac{2^p - 1}{2^p - rp} \quad (12)$$

CR value is based on the value of n and r . For maximum CR, all blocks should have valid code words and r is equal to one. When all blocks appear as non-valid code words, then the CR will be minimum.

A high throughput memory efficient pipelining architecture for fast and efficient Set Partitioning in Hierarchical Trees (SPIHT) image compression system is presented in (Muthukumaran and Ravi, 2016). The spatial oriented tree approach in the SPIHT algorithm is used for compression. IWT is employed to encode and decode data in SPIHT algorithm. The metrics used to measure the performance of proposed technique are CR, MSE and PSNR value.

A novel lossless DC device is proposed which expands the enterprise network to branch offices by combining various communication technologies (Mehboob et al., 2010). A high-speed architecture which implements LZ77 coding technique on FPGA providing multi-gigabit throughput of data is presented. A comparison is carried out between the proposed method (super unfolded and pipelined architectures) and titan-R optimized architecture. The proposed method performs well than existing methods and throughput can be increased by increasing the replication factor. The higher throughput allows interfacing of various high-speed communication technologies to preserve bandwidth for various applications.

Another paper is proposed in (Uthayakumar et al., 2017a), employed BWT algorithm to compress seismic data efficiently.

The results of BWT coding are compared with LZMA, LZW and AC. The obtained results show that BWT attained superior performance than other methods in terms of CR, CF and bpc. Uthayakumar et al. (2017b) applied Lempel Ziv Markov-chain Algorithm (LZMA) to effectively compress the wind plant data in WSN. LZMA is a lossless DC algorithm which is highly applicable for real time applications. The sensor node in WSN collects the data from the physical environment and executes LZMA algorithm to compress it. The achieved results are compared with BWT, Huffman and arithmetic coding in terms of CR, CF and bpc.

5. Research directions

An efficient DC technique should achieve an optimal tradeoff between performance and computational complexity. Though numerous compression techniques were developed, still significant issues and challenges need to be addressed. In this section, some categories of open issues and research directions are discussed.

5.1. Characteristics of DC techniques

The characteristics of any DC techniques includes compatibility, efficient representation, computational complexity, and memory management. Recently, amount of data generation is tremendously increasing day by day and also various file formats are introduced. Therefore, compression techniques should be designed in a compatible way to achieve effective CR irrespective of file size and formats. Next, during compression, the appropriate representation of data also influences the performance of a compression technique, various ways of representation exist such as matrix format, tree-structure, chain code, quantization, etc. An inappropriate representation of data may result to larger compressed file than the original file. So, the proper way of representing data may give better CR and it should be concentrated more in this area.

On the other side, computational complexity and available memory management also plays a major role in the design of DC techniques. Most of the existing techniques suffer from the problem of high computational complexity, which is particularly a serious issue in real time applications like WSN, medical imaging, satellite imaging, etc. As the applications are energy constrained and delay sensitive, computationally inexpensive techniques need to be developed. Likewise, compression technique increases the complexity in memory management. When the amount of memory needs to execute the compression technique exceeds the available system memory, effective compression cannot be achieved. Even though some compression techniques achieve better CR, it does not manage the available memory efficiently. Therefore, memory management in compression technique is another open research area.

5.2. Discrete tone images (DTI)

Most of the image compression techniques are developed only for continuous tone images and very few compression techniques are proposed for discrete tone images. The importance of discrete tone images is increasing day by day by the machine generated images like digital maps, GIS, logos, etc. So, the development of compression techniques for discrete tone images should be concentrated.

5.3. Machine learning (ML) in DC

The application of compression algorithms in machine learning (ML) tasks such as clustering and classification has appeared in a

variety of fields, sometimes with the promise of reducing problems of explicit feature selection. In contrary, the nature of ML can be utilized to predict the original input sequence from the past sequence which will be very useful in the decompression process. Since ML approaches find useful to attain optimal solutions in various scenarios, the inclusion of ML techniques in the field of DC should be encouraged to enhance the compression performance.

6. Conclusions

DC techniques plays a significant role to handle massive amount of data generated in various forms in digital world. No universal DC approaches has been proposed to effectively compress different types of data in diverse applications. Several DC approaches are proposed to compress various forms of data like text, audio, video, images and so on. This paper outlines and surveys the state-of-the-art DC concepts in several aspects. With the goal of understanding further intricacies of the DC, we have broadly divided the DC techniques based on data quality, coding schemes, data type and applications. The traditional and recently developed DC techniques are reviewed and compared by highlighting their objective, methodology used, performance metrics and suitability to various applications. Nevertheless, given the relative infancy of the field, there are still quite a number of outstanding problems that need further investigation from the perspective of proposing key techniques and advanced solutions. At the end of paper, it results to a useful guideline to select appropriate technique for intended applications or designing new DC algorithms based on the requirement.

Appendix A

Resources on data compression

Web pages

- www.mattmahoney.net/dc.
- www.datacompression.info.
- www.compression.ca.

Books

- David Salomon, Data Compression: The Complete Reference. Springer-Verlag, London, 2007
- Khalid Sayood, Introduction to Data Compression. Elsevier, 2012
- Mark Nelson, The Data Compression Book, M&T Press, 1991
- Ida Mengyi Pu, Fundamentals of Data Compression, Elsevier, 2005.

Scientific journals

Scientific articles on DC are published in many journals, including “Real-Time Image Processing”, “Multimedia Tools and Applications”, “Multimedia Systems”, “Journal of Computer Vision” (Springer). “Image and Vision Computing” (Elsevier). “IEEE Transactions on Image Processing”, “IEEE Transactions on Multimedia”, “IEEE Transactions on Information Theory” (IEEE).

Conferences

- International Conference on Data Compression (IEEE)
- International Conference on Image Processing (IEEE)
- European conference on Machine Learning (Springer)
- <http://www.cs.brandeis.edu/~dcc/index.html>.

Software

- <http://www.techradar.com/news/the-best-file-compression-software>.
- <https://en.wikipedia.org/wiki/7-Zip>.
- <https://en.wikipedia.org/wiki/BetterZip>.
- <https://en.wikipedia.org/wiki/PKZIP>.
- <https://en.wikipedia.org/wiki/Bzip2>.
- <https://en.wikipedia.org/wiki/Compress>.
- https://en.wikipedia.org/wiki/Stufit_Expander.
- https://en.wikipedia.org/wiki/Windows_Media_Encoder.
- <https://en.wikipedia.org/wiki/WavPack>.
- <https://en.wikipedia.org/wiki/Blu-code>.
- <https://en.wikipedia.org/wiki/Huffyuv>.

Dataset

- <http://corpus.canterbury.ac.nz/descriptions/>.
- <http://corpus.canterbury.ac.nz/descriptions/#cantrbry>.
- <http://corpus.canterbury.ac.nz/descriptions/#calgary>.
- http://www.imageprocessingplace.com/root_files_V3/image_databases.htm.
- <https://www.cs.cmu.edu/~cil/v-images.html>.
- <https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/>.
- <http://www2.irccyn.ec-nantes.fr/ivcdb/>.
- <http://live.ece.utexas.edu/research/quality/subjective.htm>.
- <http://vision.eng.shizuoka.ac.jp/mod/page/view.php?id=26>.
- <https://iee-dataport.org/documents/videoset>.

Popular press

- Springer/Chapman & Hall/crc/Elsevier/LAP Lambert Academic Publishing

Mailing list

- Data Compression: The Complete Reference – David.salomom@csun.edu
- To Join DC Community – <http://www.internz.com/compression-pointer.html>.

References

- Abel, J., Teahan, W., 2005. Universal text preprocessing for data compression. *IEEE Trans. Comput.* 54, 497–507.
- Abu Alsheikh, M., Lin, S., Niyato, D., Tan, H.P., 2016. Rate-distortion balanced data compression for wireless sensor networks. *IEEE Sens. J.* 16, 5072–5083. <https://doi.org/10.1109/JSEN.2016.2550599>.
- Amri, H., Khalfallah, A., Gargouri, M., Nebhani, N., Lapayre, J.-C., Bouhlel, M.-S., 2017. Medical image compression approach based on image resizing, digital watermarking and lossless compression. *J. Signal Process. Syst.* 87, 203–214. <https://doi.org/10.1007/s11265-016-1150-5>.
- Anantha Babu, S., Eswaran, P., Senthil Kumar, C., 2016. Lossless compression algorithm using improved RLC for grayscale image. *Arab. J. Sci. Eng.* 41, 3061–3070. <https://doi.org/10.1007/s13369-016-2082-x>.
- Ansari, R., Memon, N., Ceran, E., 1998. Near-lossless image compression techniques. *J. Electron. Imaging Syst.* 7, 486–494.
- Auristin, F.N., 2016. Advanced audio compression for lossless audio coding using IEEE 1857.2. *Int. J. Eng. Comput. Sci.* 5, 18124–18127. <https://doi.org/10.18535/ijecs/v5i9.57>.
- Benini, L., Bruni, D., Macii, A., Macii, E., 2004. Memory energy minimization by data compression: algorithms, architectures and implementation. *IEEE Trans. Very Large Scale Integr. Syst.* 12, 255–268. <https://doi.org/10.1109/TVLSI.2004.824304>.
- Berger, T., 1971. Rate Distortion Theory: A Mathematical Basis for Data Compression. Prentice Hall, London.
- Brechet, L., Lucas, M.-F., Doncarli, C., Farina, D., 2007. Compression of biomedical signals with mother wavelet optimization and best-basis wavelet packet selection. *IEEE Trans. Biomed. Eng.* 54, 2186–2192.
- Brettell, J., Skoglund, J., 2016. Open-Source Spatial Audio Compression for VR Content, 2nd ed. In: SMPTE 2016 Annual Technical Conference and Exhibition. California technical publishing. <https://doi.org/10.5594/M001712>.
- Burrows, M., Wheeler, D., 1994. A block-sorting lossless data compression algorithm. *Algorithm. Data Compression* 18. <https://doi.org/10.1.1.37.6774>.
- C. Oswald, I.A.K., Avinash, J., Sivaselvan, B., 2017. A Graph-Based Frequent Sequence Mining Approach to Text Compression, in: In: International Conference on Mining Intelligence and Knowledge Exploration. pp. 371–380.
- Capon, J., 1959. A probabilistic model for run-length coding of pictures. *IRE Trans. Inf. Theory* 100, 157–163.
- Shannon, C.E., 1938. A symbolic analysis of relay and switching circuits. *IEEE Trans. Commun.* 57, 713–723.
- Chandler, D.M., Hemami, S.S., 2007. VSNR: a wavelet-based visual signal-to-noise ratio for natural images. *IEEE Trans. Image Process.* 16, 2284–2298.
- Che, W., Zhao, Y., Guo, H., Su, Z., Liu, T., 2015. Sentence compression for aspect-based sentiment analysis. *IEEE/ACM Trans. Audio Speech Lang. Process.* 23, 2111–2124.
- Cheng, K.-O., Wu, P., Law, N.-F., Siu, W.-C., 2015. Compression of multiple DNA sequences using intra-sequence and inter-sequence similarities. *IEEE/ACM Trans. Comput. Biol. Bioinf.* 12, 1322–1332.
- Chew, L., Ang, L., 2008. Survey of image compression algorithms in wireless sensor networks. In: 2008 Int. Symp. Inf. Technol. 1–9. <https://doi.org/10.1109/ITSIM.2008.4631875>.
- Damera-Venkata, N., Kite, T.D., Geisler, W.S., Evans, B.L., Bovik, A.C., 2000. Image quality assessment based on degradation model. *IEEE Trans. Image Process.* 9, 636–650.
- De Agostino, S., 2015. The greedy approach to dictionary-based static text compression on a distributed system. *J. Discret. Algorithms* 34, 54–61.
- Drost, S.W., Bourbakis, N., 2001. A hybrid system for real-time lossless image compression. *Microprocess. Microsyst.* 25, 19–31. [https://doi.org/10.1016/S0141-9331\(00\)00102-2](https://doi.org/10.1016/S0141-9331(00)00102-2).
- Dutta, T., 2015. Medical data compression and transmission in wireless Ad Hoc networks. *Sens. J., IEEE* 15, 778–786. <https://doi.org/10.1109/JSEN.2014.2354394>.
- Elshaikh, M., Fadzil, M.F.M., Kamel, N., Isa, C.M.N.C., 2012. Weighted Signal-to-noise ratio average routing metric for Dynamic Sequence Distance Vector routing protocol in mobile ad-hoc networks. In: IEEE 8th International Colloquium on Signal Processing and Its Applications. pp. 329–334. <https://doi.org/10.1109/CSPA.2012.6194744>.
- Fan, B., Pedrycz, S., Adam, A., Valero, H.-P., 2017. On acoustic signal compression for ultrasonic borehole imaging. *IEEE Trans. Comput. Imaging* 3, 330–343.
- Fout, N., Ma, K.L., 2012. An adaptive prediction-based approach to lossless compression of floating-point volume data. *IEEE Trans. Vis. Comput. Graph.* 18, 2295–2304. <https://doi.org/10.1109/TVCG.2012.194>.
- Hamming, R.W., 1986. Coding and information theory.
- Hang, B., Wang, Y., Kang, C., 2016. A scalable variable bit rate audio codec based on audio attention analysis. *Rev. Tec. Fac. Ing. Univ. Zulia* 39, 114–120. <https://doi.org/10.21311/001.39.6.14>.
- Hermilo, S.-C., Rodriguez-Dagnino, R.M., 2005. Compressing bilevel images by means of a three-bit chain code. *Opt. Eng.* 44, 97004. <https://doi.org/10.1117/1.2052793>.
- Holtz, K., 1999. The evolution of lossless data compression. *Techniques*, 140–145.
- Hosseini-nejad, H., Jannesari, A., Sodagar, A.M., 2014. Data compression in brain-machine/computer interfaces based on the Walsh – Hadamard Transform. *IEEE Trans. Biomed. Circuits Syst.* 8, 129–137. <https://doi.org/10.1109/TBCAS.2013.2258669>.
- Hosseini, M., 2012. A survey of data compression algorithms and their applications. *Appl. Adv. Algorithms*, 1–14. <https://doi.org/10.13140/2.1.4360.9924>.
- Huffman, D.A., 1952. A Method for the Construction of Minimum-Redundancy Codes. *A Method Constr. Minimum-Redundancy Codes* 1098–1102.
- Hussein, A.-B., 2008. A novel lossless data compression scheme based on the error correcting Hamming codes. *Comput. Math. Appl.* 56, 143–150. <https://doi.org/10.1016/j.camwa.2007.11.043>.
- Ian, B., Magli, E., Serra-Sagrista, J., 2014. A Tutorial on image compression for optical space imaging systems. *IEEE Geosci. Remote Sens. Mag.* 2, 8–26.
- Incebacak, D., Zilan, R., Tavli, B., Barcelo-Ordinas, J.M., Garcia-Vidal, J., 2015. Optimal data compression for lifetime maximization in wireless sensor networks operating in stealth mode. *Ad Hoc Networks* 24, 134–147.
- Ziv, J., Lempel, A., 1977. A universal algorithm for data compression. *IEEE Trans. Inf. Theory* 23, 337–343.
- Jain, D., Tech, M., Engineering, C., Engineering, C., 2017. A Modified Method for Sound Compression Using Intrinsic Mode Functions 97–102.
- Jerome, S., 1993. 119.pdf.
- Ranjeet, K., Kumar, A., Pandey, R.K., 2011. ECG signal compression using different techniques. *Advances in Computing, Communication and Control. Communications in Computer and Information Science*.
- Kalajdzic, K., Ali, S.H., Patel, A., 2015. Rapid lossless compression of short text messages. *Comput. Stand. Interfaces* 37, 53–59.
- Kavitha, P., 2016. A Survey on Lossless and Lossy Data Compression Methods 7, 110–114.
- Khan, A., Wahid, K., 2014. White and narrow band image compressor based on a new color space for capsule endoscopy. *Signal Process. Image* 29, 345–360.
- Khan, A., Khan, A., Khan, M., Uzair, M., 2016. Lossless image compression: application of bi-level burrows wheeler compression algorithm (BBWCA) to 2-D data. *Multimed. Tools Appl.* <https://doi.org/10.1007/s11042-016-3629-2>.

- Kimura, N., Latifi, S., 2005. A Survey on Data Compression in Wireless Sensor Networks. In: Proc. Int. Conf. Inf. Technol. coding Comput. pp. 16–21.
- Kolo, J.G., Shanmugam, S.A., Lim, D.W.G., Ang, L.-M., Seng, K.P., 2012. An adaptive lossless data compression scheme for wireless sensor networks. *J. Sens.* <https://doi.org/10.1155/2012/539638>.
- Kumar, M., Vaish, A., 2017. An efficient encryption-then-compression technique for encrypted images using SVD. *Digit. Signal Process. A Rev. J.* 60, 81–89. <https://doi.org/10.1016/j.dsp.2016.08.011>.
- Langdon, G.G., 1984. An introduction to arithmetic coding. *IBM J. Res. Dev.* 28, 135–149. <https://doi.org/10.1147/rd.282.0135>.
- Langdon, G.G., Rissanen, J., 1981. Compression of black-white images with arithmetic coding. *IEEE Trans. Commun. COM-29*, 858–867.
- Langdon, G.G., Rissanen, J.J., 1983. A double-adaptive file compression algorithm. *IEEE Trans. Commun. COM-31*, 1253–1255.
- Lee, C.F., Changchien, S.W., Wang, W.T., Shen, J.J., 2006. A data mining approach to database compression. *Inf. Syst. Front.* 8, 147–161. <https://doi.org/10.1007/s10796-006-8777-x>.
- Li, H., Meng, W., Zhang, X., 2014. A Survey on Recent Approaches of Mesh Compressions. In: 2014 International Conference on Virtual Reality and Visualization, Shenyang. pp. 50–57.
- Liang, Y., Li, Y., 2014. An efficient and robust data compression algorithm in wireless sensor networks. *IEEE Commun. Lett.* 18, 439–442. <https://doi.org/10.1109/LCOMM.2014.011214.132319>.
- Liu, A., Lin, W., Narwaria, M., 2012. Image quality assessment based on gradient similarity. *IEEE Trans. Image Process.* 21, 1500–1512.
- Liu, Y.K., Zalik, B., 2005. An efficient chain code with Huffman coding. *Pattern Recognit.* 38, 553–557. <https://doi.org/10.1016/j.patcog.2004.08.017>.
- Louie, H., Miguel, A., 2012. Lossless compression of wind plant data. *IEEE Trans. Sustain. Energy* 3, 598–606. <https://doi.org/10.1109/TSTE.2012.2195039>.
- Luo, J., Chen, C.W., Parker, K.J., Huang, T.S., 1996. Artifact reduction in low bit rate DCT-based image compression. *IEEE Trans. Image Process.* 5, 1363–1368. <https://doi.org/10.1109/83.535848>.
- Mahmood, A., Islam, N., Nigatu, D., Henkel, W., 2014. DNA inspired bi-directional Lempel-Ziv-like compression algorithms. In: 2014 8th International Symposium on Turbo Codes and Iterative Information Processing (ITC), Bremen. pp. 162–166.
- Mahmud, S., 2012. An Improved Data Compression Method for General Data 3, 1–4.
- Marcelloni, F., Vecchio, M., 2008. A simple algorithm for data compression in wireless sensor networks. *IEEE Commun. Lett.* 12, 411–413. <https://doi.org/10.1109/LCOMM.2008.080300>.
- Mehboob, R., Khan, S.A., Ahmed, Z., Jamal, H., Shahbaz, M., 2010. Multigig lossless data compression device. *IEEE Trans. Consum. Electron.* 56, 1927–1932. <https://doi.org/10.1109/TCE.2010.5606348>.
- Mittal, S., Vetter, J.S., 2016. A survey of architectural approaches for data compression in cache and main memory systems. *IEEE Trans. Parallel Distrib. Syst.* 27, 1524–1536.
- Moorthy, A.K., Bovik, A.C., 2011. Blind image quality assessment: from natural scene statistics to perceptual quality. *IEEE Trans. Image Process.* 20, 3350–3364. <https://doi.org/10.1109/TIP.2011.2147325>.
- Muthukumar, N., Ravi, R., 2016. Hardware implementation of architecture techniques for fast efficient lossless image compression system. *Wireless Pers. Commun.* 90, 1291–1315. <https://doi.org/10.1007/s11277-016-3391-9>.
- Muthukumar, N., Ravi, R., 2014. The performances analysis of fast efficient lossless satellite image compression and decompression for wavelet based algorithm. *Wireless Pers. Commun.* 81, 839–859. <https://doi.org/10.1007/s11277-014-2160-x>.
- Narasimha, M., Peterson, A., 1978. On the computation of the discrete cosine transform. *IEEE Trans. Commun.* 26, 934–936.
- Nibali, A., He, Z., 2015. Trajic: an effective compression system for trajectory data. *IEEE Trans. Knowl. Data Eng.* 27, 3138–3151. <https://doi.org/10.1109/TKDE.2015.2436932>.
- Nielsen, M., Kamavuako, N., Andersen, M.M., Lucas, M., 2006a. Biomedical signal compression with optimized wavelets. In: InAcoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on. pp. 1044–1047.
- Nielsen, M., Nlandu, E., Anderson, M.M., Farina, D., 2006b. Optimal wavelets for biomedical signal compression. *Med. Biol. Eng. Comput.* 44, 561–568. <https://doi.org/10.1007/s11517-006-0062-0>.
- NoPlatoš, J., Šnášel, V., El-Qawasmeh, E., 2008. Compression of small text files. *Adv. Eng. Inf.* 22, 410–417.
- Oswald, C., Ghosh, A.I., Sivaselvan, B., 2015. Knowledge engineering perspective of text compression. In: In India Conference (INDICON), 2015 Annual IEEE. pp. 1–6.
- Oswald, C., Sivaselvan, B., 2017. An optimal text compression algorithm based on frequent pattern mining. *J. Ambient Intell. Hum. Comput.* 1–20.
- Parkinson, C.N., 1957. Park Parkinson's First Law: "Work expands so as to fill the time available." In: Parkinson's Law and Other Studies in Administration. Ballantine Books, New York.
- Patauner, C., Marchioro, A., Bonacini, S., Rehman, A.U., Pribyl, W., 2011. A lossless data compression system for a real-time application in HEP data acquisition. *IEEE Trans. Nucl. Sci.* 58, 1738–1744. <https://doi.org/10.1109/TNS.2011.2142193>.
- Patel, N., Chaudhary, J., 2017. Energy efficient WMSN using image compression: A survey. In: International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Bangalore. pp. 124–128.
- Peng, J., Kim, C.S., Kuo, C.C.J., 2005. Technologies for 3D mesh compression: a survey. *J. Vis. Commun. Image Represent.* 16, 688–733. <https://doi.org/10.1016/j.jvcir.2005.03.001>.
- Rad, R.M., Attar, A., Shahbahrani, A., 2013. A predictive algorithm for multimedia data compression. *Multimed. Syst.* 19, 103–115. <https://doi.org/10.1007/s00530-012-0282-0>.
- Rajakumar, K., Arivoli, T., 2016. Lossy image compression using multiwavelet transform for wireless transmission. *Wireless Pers. Commun.* 87, 315–333. <https://doi.org/10.1007/s11277-015-2637-2>.
- Rana, K., Thakur, S., 2017. Data compression algorithm for computer vision applications: A survey. In: 017 International Conference on Computing, Communication and Automation (ICCCA), Greater Noida, India. pp. 1214–1219.
- Rao, R.Y., Eswaran, C., 1996. New bit rate reduction techniques for block truncation coding. *IEEE Trans. Commun.* 44, 1247–1250.
- Rehman, M., Sharif, M., Raza, M., 2014. Image compression: a survey. *Res. J. Appl. Sci. Eng. Technol.* 7, 656–672.
- Reisenhofer, R., Bosse, S., Kutyniok, G., Wiegand, T., 2016. A Haar wavelet-based perceptual similarity index for image quality assessment. *Comput. Vis. Pattern Recognit.* <https://doi.org/10.1016/j.image.2017.11.001>.
- Rekha, H., Samundiswary, P., 2015. Survey on low power adaptive image compression techniques for WSN. In: 2015 International Conference on Communications and Signal Processing (ICCS), Melmaruvathur. pp. 1128–1132.
- Rissanen, J., 1983. A universal data compression system. *IEEE Trans. Inform. Theory* IT-29, 656–664.
- Robert, L., Nadarajan, R., 2009. Simple lossless preprocessing algorithms for text compression. *IET Software* 3, 37–45.
- Ruxanayasin, B., Krishna, B.A., 2013. Implementation of data compression techniques in mobile Ad hoc Networks 80, 8–12.
- Alzahir, S., Borici, A., 2015. An innovative lossless compression method for discrete-color images. *IEEE Trans. Image Process.* 24, 44–56.
- Sridevi, S., Vijayakumar, R., Anuja, R., 2012. A survey on various compression methods for medical images. *Int. J. Intell. Syst. Appl.* 4, 13–19. <https://doi.org/10.5815/ijisa.2012.03.02>.
- Satir, E., Isik, H., 2014. A Huffman compression based text steganography method. *Multimed. Tools Appl.* 70, 2085–2110. <https://doi.org/10.1007/s11042-012-1223-9>.
- Saupe, D., Hamzaoui, R., 1994. A review of the fractal image compression literature. *Comput. Graph (ACM)* 28, 268–276.
- Sayood, K., 2006. Introduction to data compression. *Managing.* <https://doi.org/10.1159/000207355>.
- Sayood, K., 2000. Introduction to Data Compression, second ed. San Francisco, CA.
- Sheikh, H.R., Bovik, A.C., de Veciana, G., 2005. An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Trans. Image Process.* 14, 2117–2128. <https://doi.org/10.1109/TIP.2005.859389>.
- Smith, C.A., 2010. A survey of various data compression techniques. *Int. J. Recent Technol. Eng.* 2, 1–20.
- Srisooksai, T., Keamrungrasri, K., Lamsrichan, P., Araki, K., 2012. Practical data compression in wireless sensor networks: a survey. *J. Netw. Comput. Appl.* 35, 37–59. <https://doi.org/10.1016/j.jnca.2011.03.001>.
- Sudhakar, R., Karthiga, M.R., Jayaraman, S., 2005. Image compression using coding of wavelet coefficients – a survey. *Icgst-Gvip* 5, 25–38.
- Tarek, S., Musaddiq, M., Elhadi, S., 2016. Data compression techniques in Wireless Sensor Networks. *Future Gener. Comput. Syst.* 64, 151–162. <https://doi.org/10.1016/j.future.2016.01.015>.
- Tate, J.E., 2016. Preprocessing and Golomb-Rice encoding for lossless compression of phasor angle data. *IEEE Trans. Smart Grid* 7, 718–729. <https://doi.org/10.1109/TSG.2015.2495290>.
- Todd, S., Langdon, G.G., Rissanen, J., 1985. Parameter reduction and context selection for compression of gray-scale images. *IBM J. Res. Dev.* 29, 188–193.
- Ullah, F., 2012. A new data compression technique using an evolutionary programming approach. *Emerging Trends Appl. Inf.*, 524–531.
- Uthayakumar, Vengattaraman, Amudhavel, 2017a. A simple lossless compression algorithm in wireless sensor networks: an application of seismic data. *IIOAB J.*
- Uthayakumar, Vengattaraman, Amudhavel, 2017b. A simple lossless compression algorithm in wireless sensor networks: an application of wind plant data. *IIOAB J.*, 281–288.
- Vadori, V., Grisan, E., Rossi, M., 2016. Biomedical signal compression with time- and subject-adaptive dictionary for wearable devices for signal compression. In: 2016 IEEE INTERNATIONAL WORKSHOP ON MACHINE LEARNING FOR SIGNAL PROCESSING. pp. 1–6.
- Venkataraman, K.S., Dong, G., Xie, N., Zhang, T., 2013. Reducing Read Latency of Shingled Magnetic Recording With Severe Intertrack Interference Using Transparent Lossless Data Compression 49, 4761–4767.
- Venugopal, D., Mohan, S., Raja, S., 2016. An efficient block based lossless compression of medical images, second ed. Optik. California technical publishing. <https://doi.org/10.1016/j.jijleo.2015.10.154>.
- Vitter Jeffrey S, 1987. Design and Analysis of Dynamic Huffman Coding 34, 825–845.
- Wang, Z., Bovik, A.C., 2002. A universal image quality index. *IEEE Signal Process. Lett.* 9, 81–84. <https://doi.org/10.1109/97.995823>.
- Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* 13, 600–612. <https://doi.org/10.1109/TIP.2003.819861>.
- Wang, Z., Simoncelli, E.P., Bovik, A.C., 2003. Multiscale structural similarity for image quality assessment. In: The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers. pp. 1398–1402. <https://doi.org/10.1109/ACSSC.2003.1292216>.

- Wang, Z., Wu, G., Sheikh, H., Simoncelli, E., Yang, E.-H., Bovik, A., 2006. Quality-aware images. *IEEE Trans. Image Process.* 15, 1680–1689. <https://doi.org/10.1109/TIP.2005.864165>.
- Wang, Z., Zhang, D., 1998. A novel approach for reduction of blocking effects in low-bit-rate image compression. *IEEE Trans. Commun.* 46, 732–734. <https://doi.org/10.1109/26.681401>.
- Watkinson, J., 2004. *The MPEG Handbook: MPEG-1, MPEG-2, MPEG-4*. Taylor & Francis.
- Way, O.M., 2003. Model-Based Compression in Wireless Ad Hoc Networks Categories and Subject Descriptors. pp. 231–242.
- Welch, T.A., 1984. *lzw 1984.pdf*.
- Wu, M., Tan, L., Xiong, N., 2016. Data prediction, compression, and recovery in clustered wireless sensor networks for environmental monitoring applications. *Inf. Sci. (Ny)* 329, 800–818. <https://doi.org/10.1016/j.ins.2015.10.004>.
- Yin, H., Hu, H., 2014. An efficient lossless image compression algorithm for external memory bandwidth saving. In: *Data Compression Conf. Proc.* 46, 435. <https://doi.org/10.1109/DCC.2014.88>.
- You, C., 2016. Near-lossless compression/decompression algorithms for digital data transmitted over fronthaul in C-RAN. *Wireless Networks*. <https://doi.org/10.1007/s11276-016-1352-6>.
- ZainEldin, H., Elhosseini, M.A., Ali, H.A., 2015. Image compression algorithms in wireless multimedia sensor networks: a survey. *Ain Shams Eng. J.* 6, 481–490.
- Zhang, L., Li, H., 2012. SR-SIM: A fast and high performance IQA index based on spectral residual. In: *Image Processing (ICIP), 2012 19th IEEE International Conference on*. pp. 1473–1476.
- Zhang, L., Zhang, L., Mou, X., Zhang, D., 2011. FSIM: a feature similarity index for image quality assessment. *IEEE Trans. Image Process.* 20, 2378–2386. <https://doi.org/10.1109/TIP.2011.2109730>.
- Zhang, W., Hu, X., Li, X., Yu, N., 2013. Recursive histogram modification: establishing equivalency between reversible data hiding and lossless data compression. *IEEE Trans. Image Process.* 22, 2775–2785. <https://doi.org/10.1109/TIP.2013.2257814>.
- Zhu, H., Zhao, C., Zhang, X., 2013. A novel image encryption-compression scheme using hyper-chaos and Chinese remainder theorem. *Signal Process. Image Commun.* 28, 670–680. <https://doi.org/10.1016/j.image.2013.02.004>.
- Ziv, J., Lempel, A., 1978. Compression of individual sequences via variable-rate coding. *IEEE Trans. Inf. Theory IT-24*, 530–536.