

**ADAPTIVE LOSSLESS DICTIONARY
BASED COMPRESSION**

**SUMMARY
of
THESIS**

**SUBMITTED TO THE
BABASAHEB BHIMRAO AMBEDKAR UNIVERSITY
LUCKNOW
FOR THE AWARD OF THE DEGREE OF**

Doctor of Philosophy

IN

COMPUTER SCIENCE

BY

SEEMA GUPTA

Under the Supervision of

DR. DEEPA RAJ

Asst. Professor

**Department of Computer Science
School of Information Science & Technology**

**BABASAHEB
BHIMRAO
AMBEDKAR
UNIVERSITY**



LUCKNOW
प्रज्ञा शील करुणा
ESTABLISHED 1996

**DEPARTMENT OF COMPUTER SCIENCE
BABASAHEB BHIMRAO AMBEDKAR UNIVERSITY
(A CENTRAL UNIVERSITY)
VIDYA VIHAR, RAEBARELI ROAD
LUCKNOW-226025 (U.P.) INDIA**

2017

SUMMARY

The development of multimedia and digital imaging (image data says thousands of words than text files) has led to high quantity of data required to represent modern imaginary. This requires large disk space for storage, long time for transmission over computer networks, as these two are relatively expensive and very difficult to operate. These factors arise the need for less data or compression of data for transmission. As in this digital world less data option is not applicable and only viable option is compression of data. Compression addresses the problem of reducing the amount of space required to represent a digital text data/ image data yielding a compact representation of a text/image, and thereby reducing the text/image storage and transmission time requirements. The key idea of this research is to remove redundancy of data presented within a text/image to reduce its size without affecting the essential information. Different types of media used in multimedia applications - text, audio, video, fax, images and speech are represented in digital form and has different types of redundancy. If the representation of a media captures content that is not perceivable by humans, then removing such content will not affect the quality of the content. For example, capturing audio frequencies outside the human hearing range can be avoided without any harm to the audio's quality. In most cases, the bandwidths derived were greater than those that are available with the communication networks over which the related services are provided. In addition, when using a public network in which call charges are based on the duration of a call, considerable cost savings can be made if the volume of information to be transmitted is reduced. In almost all multimedia applications, therefore, a technique known as compression can be applied to get the

desired result. Compression is first applied to the source information prior to its transmission. This is done either to reduce the volume of information to be transmitted-text and images-or to reduce the bandwidth that is required for its transmission-speech, audio and video. In digital mastering of telemedicine and medical images, lossless compression method reconstructs the same image hence better diagnosis can be done.

CHAPTER I

This chapter described fundamental aspects of compression and information theory related to probability. A compression model is a measurement of the probability distribution of inputs to a compressor and represented as a sequence of predictions of symbols like bits, bytes or words. A model can be static and dynamic compression. Static Compression is of fixed length code hence the best compressor uses dynamic model. There are many types of compression e.g. speech compression, audio compression, video compression, text compression and image compression. Easy and less time consuming transmission of high quality images/text requires compression-decompression technique to be as simple as possible and completely lossless. This very chapter also describes fundamental basic of compression, basic information theory, different types of compression techniques, various types of redundancies, different parameters and applications of compression in details.

CHAPTER II

The second chapter is based upon the literature survey related to the present work. This chapter contains brief description of the research papers collected from the year 2000 onwards. From the literature, it is observed that a very little information about the dictionary based lossless compression, towards the hardware and software

architectures is available. The various dictionary based compression method designed in the present research work are efficient, effective and less time consuming in terms of performance. The main goal of this chapter is to survey the different research papers for the lossless compression of image file/text file by which gives the way to move forward for better compression. The content of this chapter have been presented and published in the proceedings of **National Seminar on Novel Paradigms of Software Engineering and Database Technologies by Delhi Institute of Advanced Studies, Delhi.**

CHAPTER III

This chapter describes about Huffman coding and LZW coding. This chapter compares between Huffman coding and Lempel Ziv Welsh coding and find major parameter i.e. compression ratio metrics. Huffman is the most efficient but requires two passes over the data while Lempel-Ziv-welsh uses just one pass. This feature of a single pass is obviously important when saving to a hard disk drive or when encoding and decoding data in real-time communications. Huffman coding is a technique used to compress files for transmission using statistical coding. Huffman coding contains more frequently used symbols have shorter code words. It works well for text and fax transmissions and its application that uses several data structures. LZW take only one pass at the time of encoding and dictionary is also required to send at the time of transition it takes more processing time. When LZW and Huffman are used to compress a binary file (all of its contents either 1 or 0), LZW gives a better compression ratio than Huffman. The content of this chapter have been presented and published in **International Journal on Information and Computation Technology, India.**

CHAPTER IV

The chapter describes adaptive text and image compression where performance is measured using compression ratio factor for different text string and image data. Text compression is very necessary for sending data from one end to another end and for storing in the secondary storage devices. Compression reduces the bandwidth of network and utilization of memory which increases the speed of processing. Lossless text compression, compresses the data which matches with the original data and no information is lost where each data is very important for analysis. Image compression means image data reduction. The principle approach in image compression is the reduction of the amount of image data (bits) while preserving information (image details). While hearing that image data is reduced, one could expect that automatically also the image quality will be reduced also. A loss of information is totally avoided in lossless compression, where image data are reduced while image information is totally preserved. Image compression address the problem of reducing the amount of data required to represent a digital image with no significant loss of information. Adapting compression refers compression which can be applied dynamically and the environment is subject to change with time. The adaptive Huffman coder assumes an initial distribution of characters in the block and then changes the coding of individual symbols using Huffman tree based on the actual content of the symbols in the data file as they are being processed. The advantage of the adaptive code is that the tree is not passed to the decompression device but it is generated receiver side sequentially. The main problem with the adaptive stage is that it takes a certain amount of data to be processed. The contents of this chapter have been published in **Proceeding of National Conference on Information Security and Software Engineering, Lucknow,**

Second paper is published in Far East Journal of Electronic and Communication, Allahabad and Third paper is published in International Conference on Advances in Computer Engineering and Applications, Ghaziabad.

CHAPTER V

This chapter describes location based compression based on location of the pixel value of image. Lossless image compression retains the same image after compression. The proposed method is adaptive, not taking complete image but accessing row major order of the block of the image to reduce the size of the pixel and transmit the pixel data at receiver side, decode the pixel and display the same pixel. This chapter tells the concept of dynamic allocation for generating variable length of each pixel after subtracting from the maximum intensity pixel in the current block. In the proposed method, one can adaptively select either row wise or column wise scan within a block of a matrix. This method can be applied on grey scale images. The content of this chapter have been published in **Asian Journal of Technology and Management Research, India.**

CHAPTER VI

This chapter discusses dictionary based lossless compression with block wise data. In this method, coming data is used for making dictionary for each block. Previous block's dictionary is used for the next block for generating binary code for each generated string, using dynamic data coming online simultaneously. Adaptive Huffman tree is in use while data is not matched with the dictionary, then variable length code is generated with the tree. The content of this chapter have been published in **International Conference on Computing for Sustainable Global Development, Delhi.**

CHAPTER VII

This chapter describes adaptive lossless dictionary based double image compression. There are two techniques in dictionary based image compression i.e. bitmask and Bose, Choudhary and Hocquendham (BCH) technique. First level compression is bitmask technique and second level compression is BCH technique so it is called double image compression. Bitmask can be used for handling mismatch pattern and repeat pattern aggressively. Bitmask converts 8 bit data into either 7 bit or 3 bit data. BCH method is based on error detecting or error correcting code in data transmission in block coding. BCH method again compresses 7 bit (received by bitmask) data into 4 bit data. Number of bits more reduced in BCH method. The main motive is to provide good compression ratio which minimize the time and space complexity. The content of this chapter have been published in **International Journal of Graphics and Image processing, India.**

CHAPTER VIII

This last chapter describes conclusion and future scope of work. Chapters above summarized different methods of compression and their outcome and established that number of bits are reduced using these methods by the technique described in this summary which also saves energy and improve the performance of the system involved. In Future, the work may be done in a way to develop an object oriented design of some lossless compression technique, compare the technique which will be the best and try to develop a model and frame work for compression of image using prediction by partial matching. It may also include to apply technique in other domains such as data compression for manufacturing testing. Future work may also apply this technique for compression a wide variety of input patterns including

application programs, FPGA configuration bitstreams, as well as NISC control words. The present work is done on grey scale which gives way to used the technique in future on color images also.

LIST OF PUBLICATIONS

1. Raj, Deepa and **Gupta, Seema**, *Survey Paper of Lossless Image Compression*, **Published in the Proceedings of National Seminar on Novel Paradigms of Software Engineering and Database Technologies** organized by Delhi Institute of Advanced Studies, Delhi, ISBN: 978-81-927406-3-8, pp. 1-8, March (2014) .
2. Raj, Deepa and **Gupta, Seema**, *Performance of Lossless Image Compression*, **Published in the International Journal on Information and Computation Technology**, ISSN: 0974-2239, Volume 4, Number 9, pp. 903-910, India (2014).
3. Raj, Deepa and **Gupta, Seema**, *Comparative Study of Adaptive Huffman Coding and Huffman Coding*, **Proceedings in National Conference on Information Security and Software Engineering** organized by Lucknow University, Lucknow, pp.132-136, March (2014).
4. Raj, Deepa and **Gupta, Seema**, *Adaptive Huffman Coding: overview*, **Published in the International Conference on Modeling and Computation** organized by Babasaheb Bhimrao Ambedkar University, Lucknow, July (2014).
5. Raj, Deepa and **Gupta, Seema**, *Adaptive Lossless Forward Move Dictionary Based Compression*, **Published in International Conference on Computing for sustainable Global Development organized by BVICAM, Delhi**, pp.1859-1864, ISBN: 978-9-3805-4415-1, IEEE Publisher, March (2015).
6. Raj, Deepa and **Gupta, Seema**, *Framework of Adaptive Lossless Compression with Block wise Data*, **Published in International Conference on Advances in Computer Engineering and Applications** organized by **IMS Engineering College**, Ghaziabad, ISBN: 978-1-4673-6910-7, IEEE Publisher, pp.103-107, March (2015).

7. Raj, Deepa and **Gupta, Seema**, *Segmentation and Reassembly of Images using Bitplane Slicing in Adaptive Lossless Dictionary Based Compression*, **Published in International Journal of Computer Application**, Number1, pp.1-4, ISSN: 0975-8887, U.S.A., August (2016).
8. Raj, Deepa and **Gupta, Seema**, *Location Based Adaptive Block wise Lossless Image Compression*, **Published in Asian Journal of Technology and Management Research**, ISSN: 2249-0892, Volume: 05, Issue: 02, December (2015).
9. Raj, Deepa and **Gupta, Seema**, *Framework of Adaptive Block Wise Lossless Image Compression* **Published in Far East Journal of Electronic and Communication**, ISSN: 0973-7006, Issue: 04, pp. 493-503, 2016.
10. Raj, Deepa and **Gupta, Seema**, *Adaptive Lossless Dictionary Based Double Image Compression*, **Published in International Journal of Graphics and Image Processing**, ISSN: 2249-5452, Volume: 5, Issue: 3, November (2015).
11. Raj, Deepa and **Gupta, Seema**, *Adaptive Lossless Dictionary Based Compression with Block Wise Data communicated in International Journal of Synthetic Emotions on data analytics*, U.S.A.