Synopsis
A Study Towards Design and Implementation of Data/Image Compression Techniques

For the registration of the degree
Of
Doctor of Philosophy
In
Computer Science and Engineering
By
Utpal Nandi

Under the supervision
of
Prof. (Dr.) J. K. Mandal

Department of Computer Science and Engineering
Faculty of Engineering, Technology & Management
University of Kalyani
1. Introduction

To store huge amount of data requires large storage space and to transfer the same over network needs huge time. The huge data also cause of traffic congestion. To reduce these problems, the data compression techniques are introduced that reduce the amount of data required to represent a source of information to reduce the storage space requirement and the time of data transmission over network. There are two major class of data compression techniques i.e. loss-less and lossy. The loss-less techniques generate exact duplicate of the original data after compress/expand cycle. This is the type of compression used when storing database records, spreadsheets, or word processing files. In these applications, the loss of even a single bit could be catastrophic. But, the lossy techniques concede a certain loss of accuracy. Lossy compression proves effective when applied to graphics images and digitized voice. Over the years several compression techniques are proposed. One of the loss-less data compression technique is Huffman coding. The technique produces minimum length code for maximum frequency symbol. One of the limitation of the technique is that it does not provide minimum length code to region-wise maximum frequency element. The limitation is eliminated by RBH and MRBH coding proposed by the candidate. The variants of same is also proposed where region formation is adaptive based on ASCII value difference such as SARBH, SARBHI and SARBHS which provide better rate of compression than the previous techniques. One problem of these techniques is that they need to attach the frequency table with the compressed file. The adaptive version of this method i.e. Adaptive Huffman coding solves this limitation. Two adaptive technique is proposed by the candidate are WHDS and WHMW which uses a window to store most recently used elements and provides much better result than Huffman, RBH and its variants and adaptive Huffman. So far, the compression methods we have looked at used a statistical model to encode single symbols. They achieve compression by encoding symbols into bit strings that use fewer bits than the original symbols. But dictionary-based compression algorithms use a completely different method to compress data. It encodes variable-length strings of symbols as single tokens. The tokens form an index to a phrase dictionary. If the tokens are smaller than the phrases they replace, compression occurs. LZ77 compression uses previously seen text as a dictionary. It replaces variable-length phrases in the input text with fixed-size pointers into the dictionary to achieve compression. The amount of compression depends on how long the dictionary phrases are, how large the window into previously seen text is, and the entropy of the source text with respect to the LZ77 model. LZSS improved on LZ77 compression by eliminating the requirement that each token output a phrase and a character. LZ78 is similar to LZ77 in some ways. LZ77 outputs a series of tokens. Each token has three components: a phrase location, the phrase length, and a character that follows the phrase. LZ78 also outputs a series of tokens with essentially the same meanings. Each LZ78 token consists of a code that selects a given phrase and a single character that follows the phrase. Unlike LZ77, the phrase length is not passed since the decoder knows it. Unlike
LZ77, LZ78 does not have a ready-made window full of text to use as a
dictionary. It creates a new phrase each time a token is output, and it adds that
phrase to the dictionary. After the phrase is added, it will be available to the
encoder at any time in the future, not just for the next few thousand characters.
LZW improved on LZ78 compression by eliminating the requirement that each
token output a phrase and a character. In fact, under LZW, the compressor never
outputs single characters, only phrases. To do this, the major change in LZW is
to preload the phrase dictionary with single-symbol phrases equal to the number
of symbols in the alphabet. Thus, there is no symbol that cannot be immediately
encoded even if it has not already appeared in the input stream. One dictionary
based technique proposed by the candidate is OLZW that optimize the LZW
code by starting encoding process with empty dictionary. The technique offers
better rate of compression than LZW for particularly small size files. All the
techniques discussed are loss-less. Several lossy techniques were also proposed.
One well-known technique is DCT-based JPEG compression that is quite
effective at low or moderate compression ratios, up to ratios of 20 or 25 to 1.
Beyond this point, the image becomes very “blocky” as the compression
increases and the image quality becomes too poor for practical use. Another
drawback of JPEG compression is its resolution dependence. In order to “zoom-
in” on a portion of an image and to enlarge it, it is necessary to replicate pixels.
The enlarged image will exhibit a certain level of “blockiness” which soon
becomes unacceptable as the expansion factor increases. Because of this
problem, it is sometimes necessary to store the same image at different
resolutions, thus wasting storage space. So, although JPEG is now a well-
established standard for lossy image compression, it has its limits and alternative
compression methods must be considered. Wavelet-based methods are gaining
popularity. They are similar in spirit to the DCT methods but do not suffer from
some of its shortcomings. Another technique that is become very popular is
fractal image compression. It really shines for high compression ratios, or when
zooming on a portion of an image or enlarging a complete image. Two fractal
based lossy image compression techniques i.e. FCI-HV and FLCD-HV are
proposed by the candidate that are better than quadtree partitioning in terms of
compression ratio and faster than HV partitioning scheme. The algorithms have
been executed on the machine configuration as Intel Core2duo processor @ 2.56
GHz CPU and 2 GB RAM with Turbo C environment.

1.1 Performance Metric

The performance of the proposed research work is measured quantitatively using
compression ratio, compression time for both loss-less and lossy compression and
additionally peak signal-to-noise ratio (PSNR) for lossy image compression. The aim is
to maximize the value of compression ratio and minimize the compression time. And
additionally for lossy image compression, maximization of PSNR is required.
2. Proposed works

The research works towards the compression of data and image done by the candidate are followed and the candidate will research the limitations of the proposed work in future.

**Region Based Huffman (RBH) compression Techniques:** RBH divides the input file into a number of region and interchanges maximum frequency element code of each region with the same of entire file before encoding that region. MRBH is the modified form of RBH where number of region is selected by RSA algorithm. The techniques offer better result than Huffman coding for most of the files.

**Region Based Huffman Compression with region wise multiple interchanging of codes:** divides also the input file into a number of region and interchanges not only maximum frequency element code of each region with the same of entire file before encoding that region but also the other high frequency elements. The technique offer better result than Huffman coding and RBH coding for most of the files.

**Size Adaptive Region Based Huffman Compression Technique (SARBH):** SARBH is the adaptive version of RBH coding that creates regions of variable size using the ASCII value differences before compression. One variant of the same is SARBHI creates not only regions of variable size using the ASCII value differences but also interchange code between the maximum frequency element of a region and maximum frequency element of entire file is done before symbols of that region are compressed. Another variation SARBS where region wise interchanging of code is done based on an additional condition.

**WHDS and WHMW techniques:** The adaptive Huffman coding with a window of limited distinct symbols i.e. WHDS uses a window buffer to store a specified number of distinct symbols most recently processed. The total number of symbols within the window may vary, but number of distinct symbols does not exceed a specified value. The adaptive Huffman tree is constructed based on the probability distribution of symbols within the window. Then, a variant of the proposed method is WHMW. The proposed variant uses two windows. A small primary window buffer is used to store the most recently processed symbols. A comparatively large secondary window buffer is used to store more past processed symbols. The first proposed technique comparatively offers better
results than its counterpart for most of the file type. The performance of the second proposed technique is also close to the other techniques.

**A Compression Technique Based On Optimality Of LZW Code (OLZW):** optimizes the LZW codes by starting the encoding process with empty dictionary which is quite effective for small size files.

**FCI-HV and FLCD-HV Fractal Image Compression:** FCI-HV partitions middle of range either horizontally or vertically to create to sub-ranges if the range is not covered well by any domain. The decision to select one of the two possibilities i.e. horizontal or vertical partitioning is done only by a simple checking which side of the range is larger than the other. One variant of the same FLCD-HV where the decision to select one of the two sides of range is done by computing the pixel value differences of the middle vertical lines and the middle horizontal lines and determining which is greater than other. The fractal image compression for grey scale image with partitioning schemes offer better compression rates than the quardtree partitioning scheme maintaining almost same compression times with improved PSNRs. Though the compression rates are not as well as offered by HV partitioning scheme, the schemes are much faster than the same.

3. **Conclusion**

The work done yet by the candidate stated in the synopsis primarily focuses on data compression techniques. The techniques of image compression have also been devised. A loss-less data compression technique is considered in the earlier part of the research by the candidate. That is Region based Huffman (RBH). The modification of the same is also considered known as Modified Region based Huffman (MRBH). Then, adaptive region formation algorithm is introduced and Size Adaptive Region Based Huffman Compression Technique (SARBH) and its two versions SARBHI and SARBHS are proposed. The Adaptive Region Based techniques offer better rate of compression than its earlier versions. Further, Windowed Huffman Coding with limited distinct symbols (WHDS) and Windowed Huffman Coding multiple window (WHMW) techniques are proposed. The techniques are adaptive eliminating the need of transmitting the frequency table with encoded data. The performances of both the techniques with respect to compression are better than Huffman, Adaptive Huffman and Region based Huffman. After that a dictionary based data compression technique is proposed i.e. Compression Technique Based On Optimality Of LZW Code (OLZW). It optimizes the LZW codes by starting the encoding process with empty dictionary which is quite effective for small size files. Then, two image compression
techniques based on fractal are proposed that are FCI-HV and FLCD-HV offering better compression rates than the quadtree partitioning scheme maintaining almost same compression times with improved PSNRs. As it has been stated that the existing techniques as well as proposed techniques have some limitations in terms of either computational times or compression rates or both, investigation may be carried out in this direction to develop more efficient techniques offering better compression rates with reduced time of compression.

---

Prof.(Dr.) J. K. Mandal  
Supervisor

Utpal Nandi  
Candidate

4. Dissemination of Work

4.1 List of papers published:

A. International journal:


B. International conference:

4.2 List of papers accepted:

A. International journal:
   i. Nandi, U., Mandal, J. K., “Region Based Huffman Compression with region wise multiple interchanging of codes”, Advancement of modelling & simulation techniques in enterprises (AMSE), France.

B. International conference:

4.3 List of papers submitted:

A. International conference:

Utpal Nandi (Candidate)
Dept. of Computer Science & Engineering
University of Kalyani, Kalyani.