Lossless Compression of Medical Images using Huffman Algorithm with 3D Predictors and Masking

Susheel George Joseph M.C.A., M.Tech(CSE), M.Phil(CS), UGC-NET(CS)[1]
Associate Professor, Department of Computer Applications, Kristu Jyoti College of Management and Technology
Changanassery, Kottayam

***Abstract*— Lossless compression is a class of data compression algorithms that allows the original data to be perfectly reconstructed from the compressed data.Lossless data compression is used in many applications.Lossless compression is used in cases where it is important that the original and the decompressed data be identical, or where deviations from the original data would be unfavourable. Typical examples are executable programs, text documents, and source code..**

***Keywords*— Image Compression, Lossless, medical image, predictors, masking**

I. Introduction

Image compression is a type of an application for data/image compression in which the basic image gets encoded with the limited bits. To lower the irrelevance and the redundancy of image data is the major target of the image compression is to enable them to get saved or transmit the data in the better form. Image compression is the lowering of the image data size, also with maintaining the required details. The basic objective of the image compression is to show an image in small quantity of bits also the needed content of information is not lost within the actual image. Compression techniques are developed rapidly to compress huge files of data like images. By the rapid growth of the technology a large quantity of image data should be managed to store those images in the proper manner by the use of effective techniques normally results in the compressing images.

The use of medical imaging systems for diagnosis purposes is a standard in current healthcare practices. With the advance of scanning technologies and digital systems in the last decades, these systems have become increasingly important, producing more accurate images with improved quality, using higher spatial resolutions and bit depths. Such improvements increase the amount of information that needs to be processed, transmitted and stored. This is particularly relevant when using volumetric scanning technologies, such as Computed Tomography (CT) or Magnetic Resonance Imaging (MRI), which produce multiple image slices.
These facts, associated to the increasing number of archiving databases of medical images deployed by healthcare providers, motivate the research of efficient compression methods for these signals. For this purpose, it is important to use lossless compression algorithms to avoid introducing errors in the data outputted by the scanning system.

Some algorithms are there which are used to perform these. types of compression in various manners like the lossless and the lossy. The image which is needs to be compressed is a gray scale having the pixel values ranges from 0 to 255. Compression addressed to the decreasing the number of data to be used to show content of an image, file or the video without decreasing excessively quality of actual data. Also it lowers the quantity of the bits that is needed to save and send the digital media [8].To do compression of something implies that there is a piece of data whose size is forced to decrease. For this the JPEG is a best selection for the digitized-photographs. The system called Joint Photographic Expert Group (JPEG) which is dependent on Discrete Cosine Transform (DCT), have been a very largely used method for compression [9]. Image compression is one of various known techniques in the image processing. This technique may have various implementations and performs a vital role in the effective storage and transmission of the images. The image compression targeted at decreasing the redundancy in the image data to record or sends only few numbers of the samples and by this also a good accession can be reconstruct for the actual image corresponding with the perception of human visual. To compress images is the major target of this paper by decreasing the number of the bits on the basis of per pixel which is needed to show it and also to lower the time of transmission for the transmission of images and for reconstructing again by the Huffman encoding algorithm.

II. Image Compression

Decreasing the irrelevance or redundancy of an image is the fundamental aim of the image compression techniques to provide the facility for storing and transmitting the data in an effective manner .The initial step in this technique is to convert the image from the representation of their spatial domain into a separate type of the representation by the use of few already known conversions and then encodes the converted values i.e., coefficients.

This technique allows the huge compression of data as compared to the predictive techniques, though at the cost of the huge computational needs. Compression is obtained by eliminating any of one or more of the below three fundamental data redundancies:

1. Coding redundancy: This is presented when the less than best (that is the smallest length) code words were used.

2. Inter-pixel redundancy: This results from the correlations between the pixels of an image.

3. Psycho-visual redundancy: This is because of the data which is neglected by human visual-system (that is, visually not required information).

An image is a 2-D signal that is processed by human visual systems. [2]These signals that are representing an image are commonly in the form of analog. Although for the storage, processing and the transmission through the computer applications, these signals needed to be converted from the analog form to their digital form. An image or a digital image is usually a 2-Dimensional array of the pixels. In the raw form, the images may cover a huge amount of the memory in the RAM and in the storage, both. Image compression is for reducing the redundancy and irrelevance of image/data to allow them to either store or transmit the data in a better way.

Many assumptions have been made during the development of document image compression systems; about image resolution, whether the image is skewed or not, how many pages are compressed at a time, and the quality of the scanned image. Moreover, these systems do not discriminate between compressing areas of text or areas of graphics. This thesis examines these assumptions for lossless document image compression in finer detail. Lossless compression reproduces the original image exactly, unlike lossy compression which trades off higher compression with slight reproduction errors.

The objectives of are:

- to investigate methods for decreasing the compressed file size during lossless document image compression;
- to explore methods for compressing document images that minimise the number of parameter choices and assumptions that are imposed;
- to design a robust method for detecting skew in document images, and the use of skew in reading order determination algorithms;
- to use machine learning techniques to combine template matching methods to achieve better results than are possible using a single method;

III. Related Work

The digital world becomes more and more important in our lives, therefore it also becomes more and more important that our transactions in that world are secure such that nobody can manipulate our data, eavesdrop on our communications etc. Mathematical algorithms called compression primitives form the basis of the size. Any trapdoor in these primitives threatens the security of the application. It is important to analyse them carefully before they are employed in practice.

This part reviews and focuses on the various studies carried out for providing necessary background to the compression techniques to protect the user's sensitive data in real-time. The literature survey provides an overview of the compression algorithms with the study of basic operations with secret components used in its designs and describes about full encryption approaches and selective lossless compression approaches for protecting the digital content. The survey from various literature collections give a good background about confusion and diffusion components for developing new compressed image designs with greater security.

Sub-Trees Modification of Huffman Coding for Stuffing Bits Reduction and Efficient NRZI Data Transmission

In this paper , proposed that a compression technique by the use of the twoloss-less technologies Huffman coding and the Lempel-Ziv-Welch coding for image compression. At first stage of this, a image is get compressed with the Huffman coding that is resulting in the Huffman-tree and a generated Huffman Code-words.In this paper, a technique has been proposed which is called the "Sub-TreesModification of Huffman Coding for Stuffing Bits Reduction and Efficient NRZI Data Transmission".

They basically targeted on transmission of the data and the multi-media compression and also treating this issue as encoding of the compression and the transmission to generate a low bit rate of transmission model which is dependent on the Huffman encoding algorithm. The suggested technique can balance the "0" bit and "1" bitby measuring the chances of mismatch found in the traditional Huffman-tree. Additionally, the suggested techniques also get modified with the transitional-tree within the same ratio of compression [10].

In recent decades, image and video compression was widely used on network access. However, there are few researches focused on the behavior between data transmission and multimedia compression. Therefore, this paper considers this problem between the encoding of compression and transmission to develop a low bit rate transmission scheme based on Huffman encoding. The proposed method can balance "0" and "1" bits to save the issue by analyzing the probability of the miss match in the typical Huffman tree. Moreover, the proposed method also can modify the transitional tree under the same compression ratio. Experimental results show that the proposed method can reduce the stuffing bits to 51.13% of standard JPEG compression. Besides, the file size after the proposed encoding is the same with the original one. It is observed that the proposed method provides a way to reduce the transmitted bits under the same compression ratio.
Lakhani and Ayyagari designed the End-Of-Block (EOB) marker to reduce the net increase of marker's code length of image data. Chang et al. used the Embedded Block Coding with Optimized Truncation (EBCOT) tier-2 feedback control to terminate redundant computation of the Embedded Block Coding (EBC) . Yeung et al. proposed a scheme based on priority scanning to encode the coding passes in a different order from high to low priority and reduce the computation by 52 percent at 0.25 bits per pixel (bpp) . Lin et al. made the balanced bit probabilities to solve the problem of Huffman codes a Reversible Variable-Length Codes (RVLCs) with balanced zeros and ones in encoded bit streams. Lakhani used an optimized custom code table to develop an optimal Huffman coding for each AC coefficient position of Discrete Cosine Transform (DCT) blocks. Besides, the minor advantage of this work is that no EOB marker is needed to represent the end of block. Although these Huffman coding methods can be very close to the minima compression rate, increasing the transmission speed by reducing the compression rate is still a very difficult task. Hence, recently, many researchers have been devoted to reduce the overhead of hardware implementation for speed up. Chang et al. indicated that coding methods can be classified into two categories: tree arrangement and table arrangement [9]. In the tree arrangement methods, Huffman tree is represented to design the finite state machine in terms of decoding the codewords.

An Adaptive Vector Quantization Method for Image Compression

In paper An Adaptive Vector Quantization Method for Image Compression, it is presented a vector quantization depend ent technique of image compression . In this paper they also adjusted the encoding for the difference in map between the actual images and then after that it got restored in the VQ compressed variation. It is the experimental results that show there model which is required toenable the additional data, it may considerably enhances the quality of the VQ images compression and also again be compromised based on the difference in map from technique of lossy compression to the technique of lossless compression.

Image compression is to reduce redundancy of the image data in order to store or transmit data in an efficient form. Compression is carried out for the following reasons about reduce, the storage requirement, processing time and transmission duration. The most powerful and quantization technique used for the image compression is vector quantization (VQ). The Existing methods Linde-Buzo-Gray (LBG) and Fast Back Propagation (FBP) algorithm are presented. In existing methods, the compression ratio is decreased. The proposed method adaptive vector quantization is used to analyze for image vector quantization (VQ). The performance of proposed work is analyzed using the factors SNR, MSE, PSNR and CR. The experimental work using MatLab shows that the proposed scheme is efficient and produced expected result.

Digital images have turn out to be popular for transferring, sharing store and image in sequence and hence high speed compression technique are need with many advantages of image compression, and the most important one is to reduce the time for the transmission of images. Fundamentally these compression techniques can be categorized into the Lossy compression techniques and lossless compression techniques. The lossy compression technique produces imperceptible difference that may be call visually lossless.

The aim of compression is to decrease the number of bits that are not required to represent data and to decrease the transmission time. Achieve compression by encoding data and the data is decompressed to its original form by decoding. A common compressed file extension is .sit, .tar, .zip; which indicates different types of software used to compress files. The compressed file is initially decompressed and at that time use. There are large numbers of software used to decompress and it depends upon which variety of file is compressed. For example WinZip software is used to decompress .zip file.

Vector Quantization is an [3] well-organized method of image compression. VQ compression system contains two components are VQ encoder and decoder. The VQ encoder finds a closest match codeword for each image block in the codebook or directory and the key of the codeword is transmit to VQ decoder. The next phase is decoding phase in which VQ decoder replaces the key values with the relevant codeword from the codebook and produces the quantized image that is called as reconstructed image. A vector quantization is generally defined as a block of pixel values. Vector Quantization is also known as “Block Quantization" or "Pattern Matching Quantization”. This process is commonly used in lossy compression methods. It works by programming values from a multidimensional vector space into a finite set of values. A lower-space vector requires less storage space, so the data is compressed. Appropriate to the density matching property of vector quantization, the compressed data contains errors that are inversely comparative to density.

Basic working of Vector Quantization is as following:

- Input image.
- Find the closest match code/vector for each image block from the directory or codebook.
- Replaces code /vector by transmit key of code for additional processing.
- Above property is used to reduce the storage space of image.

Image compression techniques are becoming very vital role in the area of image analysis, analysis statistical, analysis text, data mining, web mining etc. The image compression in terms of compression ratio and compression size. The proposed algorithm achieves lower bits rate at the same image quality. After performing different images it is concluded that proposed AVQ algorithm is the best one in all compression ratio parameters. The proposed work is compared with other existing method which produced better results for image compression.

An Enhanced Vector Quantization Method for Image Compression with Modified Fuzzy Possibilistic C-Means using Repulsion

Yanfeng Zhang et al., proposed an Agglomerative Fuzzy Kmeans clustering method with automatic selection of cluster number (NSS-AKmeans) approach for learning optimal number of clusters and for providing significant clustering results. High density areas can be detected by the NSS-AKmeans and from these centers the initial cluster centers with a neighbor sharing selection approach can also be determined. Agglomeration Energy (AE) factor is proposed in order to choose an initial cluster for representing global density relationship of objects. Moreover, in order to calculate local neighbor sharing relationship of objects, Neighbors Sharing Factor (NSF) is used. Agglomerative Fuzzy k-means clustering algorithm is then utilized to combine these initial centers to get the desired number of clusters and create improved clustering results. Experimental observations on several data sets have proved that the proposed clustering approach was very significant in automatically identifying the true cluster number and also providing correct clustering results.

Xiao-Hong et al., presented a novel approach on Possibilistic Fuzzy C-Means Clustering Model Using Kernel Methods. The author insisted that fuzzy clustering method is based on kernel methods. This technique is said to be kernel possibilistic fuzzy c-means model (KPFCM). KPFCM is an improvement in Possibilistic Fuzzy CMeans (PFCM) model which is superior to fuzzy c-means (FCM) model. The KPFCM model is different from PFCM and FCM which are dependent on Euclidean distance. The KPFCM model is dependent on non-Euclidean distance by implementing kernel methods. In addition, with kernel methods the input data can be mapped into a high-dimensional attribute space where the nonlinear pattern now looks linear. KPFCM can deal with noises or outliers superior than PFCM. The KPFCM model is interesting and provides better solution. The experimental observation shows that KPFCM provides significant performance.

Zhang Zhe et al., proposed an improved K-Means clustering algorithm. K-means algorithm is extensively utilized in spatial clustering. The mean value of all the cluster centroid in this technique is taken as the Heuristic information; hence it has some demerits such as sensitive to the initial centroid and instability. The improved clustering algorithm referred to the best clustering centroid which is searched during the optimization of clustering centroid. This increases the searching probability around the best centroid and enhanced the strength of the approach. The experiment is performed on two groups of representative dataset and from the experimental observation, it is clearly noted that the improved K-means algorithm.

IV. SYSTEM ARCHITECTURE

Benefits of Image Compression

Below are few benefits of the Image compression technique:

- It enables a reliable cost of savings that is included withthe sending of less data on the network of switched telephone in which the cost of call is normally dependent on its duration.
- It is not only to decrease the requirements of storage but also decrease the entire time of execution.
- It decreases the chances of the errors transmission as
- some bits have got transferred.
- It enables a level of the security against monitoring the unlawful Activities.

3.2. EXISTING SYSTEM

JPEG and JPEG2000 are transform-based standards, which use Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT), respectively. The latter presents a superior compression performance and supports a set of desirable features, such as region-of-interest (ROI) coding, lossy-to-lossless coding, resolution scalability and resilience against transmission errors. The superiority of JPEG2000 is also justified by its volumetric extensions, which provide extra functionality for the compression of volumetric data, such as the Multi Component Transform (MCT) feature (Part 2 of JPEG2000) and a 3D extension, JP3D (Part 10 of JPEG2000) . While the first solution interprets the multiple image slices as separate image components, the JP3D part extends the core techniques of JPEG2000 to encode the input data as a volume, namely by expanding DWT to three dimensions (3D-DWT).

The JPEG-LS standard uses the LOw COmplexity LOssless COmpression for Images (LOCO-I) algorithm, which is based on predictive coding. Its main limitation is the inability to exploit inter-slice redundancies, which makes it inefficient for the compression of volumetric data. An efficient alternative to encode this data is to use video coding standards, namely the MPEG2 and H.264/AVC, as well as the most recent state-of-the-art High Efficiency Video Coding (HEVC) standard.

3.3. Proposed system

Huffman Algorithm with 3D predictors and masking

The general idea in the Huffman encoding algorithm is to allocate the very short code-words to those blocks of input along with the high possibilities and the long code-words are allocated to those which are having the low probabilities. The Huffman code process is dependent on the two observations mentioned below:

a. Very frequently found symbols will have the shorter codewords as compare to the symbol which found less frequently.

b. Two symbols which found least frequently may have the equal length.

To mask values here a constant matrix is used with the values of the image files. In my project am using the following matrix as mask

mask =

[1 1 1 1 0 0 0 0
1 1 1 0 0 0 0 0
1 1 0 0 0 0 0 0
1 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0];

This is used as a coder during compression and decoder using decompression. If sparse matrices can be used more lossless quality can be obtained.

The Huffman code is prepared by combining together two least possible characters and that are repeating in this process as far as there is only the one character is remaining. A code-tree is hence prepared and then a Huffman code is generated from the labeling of code tree. It is the best prefix code that is generated from the set of the probabilities and which has been used inthe different applications of the compression. These generated codes are of different length of code which is using integral number of the bits. This concept results in a decrease in average length of the code and hence the whole size of the compressed data is become smaller as compare to the original one. The Huffman's algorithm is the first that provides the solution to the issue of constructing the codes with less redundancy.

In predictive coding, information already sent or available is used to predict future values, and the difference is coded. Since this is done in the image or spatial domain, it is relatively simple to implement and is readily adapted to local image characteristics. Differential Pulse Code Modulation (DPCM) is one particular example of predictive coding. Transform coding, on the other hand, first transforms the image from its spatial domain representation to a different type of representation using some well-known transform and then codes the transformed values (coefficients). This method provides greater data compression compared to predictive methods, although at the expense of greater computation.

Algorithm (Image Compression)

Step 1. Start

Step 2. Read input Image to compress

Step 3. Compress the input image and generate Compressed image known as CompressImage.

Step 4. Read CompressImage or Compressed image to decompress and convert into original image.

Step 5. Stop.

V. CONCLUSION AND FUTURE WORK

The objective of this work was to compress an image. As in many of the devices where the full size images cannot be viewed or are not supported so the compressed images are used. The image compression also helps to save memory, as the size of the compressed image is less than the actual size of the image. A new technique for the image compression that uses the Mask matrix -Coding in the combination with the Huffman encoder has been explained here. This method uses the zero tree architecture of at the decomposition level of 8 with the Huffman encoder is a very efficiently used that proved in the higher ratio of compression and a better PSNR. My project also helps to the software developer to develop new compressed software for the compressing of any image as the lossless actual image by the use of Huffman encoding algorithm. Ones it can generate their own frame-work for the compressing any of the image with the less difficulties as a space and time.

Image compression is not suitable for images with sharp edges and lines. JPEG image format is not capable of handling animated graphic images. JPEG images do not support layered images. Graphic designer need to work on layered images in order to manipulate and edit graphic images which is not possible with JPEG Images. Modern high resolution digital cameras support 10, 12, 14 or 16 bit images. If these images are stored in JPEG format, extra information is discarded, resulting in decreased image quality. These issues have to be maintained in future.

REFERENCES

[1]. Christophe, E., C. Mailhes, P. Duhamel (2016), Adaptation of zerotrees using signed binary digit representations for 3 dimensional image coding," EURASIP Journal on Image and Video Processing, 2017, to appear in the special issue on Wavelets in Source Coding, Communications, and Networks, Paper ID 54679.

[2] Misiti, M., Y. Misiti, G. Oppenheim, J.-M. Poggi (2017) Wavelets and their applications, ISTE DSP Series.

[3] Two Dimensional True Compression /Wavelet Toolbox / http : // in.mathworks.com / help

[4] Digital Image Processing using matlab (Gonzalez)/Chapter 6/Color Image Representation in MATLAB

[5] IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 17, NO. 12, DECEMBER 2016 Hyper spectral Image Compression: Adapting SPIHT and EZW to Anisotropic 3-D Wavelet Coding Emmanuel Christophe, Member, IEEE, Corinne Mailhes, Member, IEEE, and Pierre Duhamel, Fellow, IEEE [6] Emmanuel Christophe, Pierre Duhamel, and Corinne Mailhes, Adaptation of Zerotrees Using Signed Binary Digit Representations for 3 Dimensional Image Coding, INTERNATIONAL JOURNAL OF IMAGE AND VIDEO PROCESSING 1

[7] Q. Du and J. E. Fowler, Hyperspectral image compression using JPEG2000 and principal component analysis, IEEE Geosci. Remote Sens. Lett., vol. 4, no. 4, pp. 201–205, Apr. 2017.

[8] Q. Du and J. E. Fowler, Low-complexity principal component analysis for hyperspectral image compression, Int. J. High Performance Comput. Appl.,to be published.

[9] D. Van Buren, A high-rate JPEG compression system for space, in Proc. IEEE Aerospace Conf., Mar. 2016, pp. 1–7.

[10] P.-S. Yeh, P. Armbruster, A. Kiely, B. Masschelein, G. Moury, C. Schaefer, and C. Thiebaut, The new CCSDS image compression recommendation, presented at the IEEE Aerospace Conf., Mar. 2015.

[11] X. Tang, W. A. Pearlman, and J. W. Modestino, Hyperspectral image compression using three-dimensional wavelet coding, in Proc. SPIE Image and Video Communications and Processing, 2016, vol. 5022, pp. 1037– 1047. [12] H. Kim, C. Choe, and J. Lee, Fast implementation of 3-D SPIHT using tree information matrix, in Proc. IEEE Int. Geoscience and Remote Sensing Symp., Jul. 2013, vol. 6, pp. 3586–3588.

[13] X. Tang and W. A. Pearlman, Scalable hyperspectral image coding, in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, 2015, vol. 2, pp. 401–404.