Volume 4, Issue 11, November 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

An Enhanced Approach in Huffman Coding Scheme

A. Nagarajan*, V. Devendran
Department of Computer Applications,
PSNA College of Engineering and Technology
India

Abstract — Data compression means the process of compressing data or files in such a manner that they require very less space than they had been in their original form. It is the process of reducing the amount of data required for transmission or storage of a piece of information in the format of text, graphics, video, sound, etc.. The efficient technique for data and image compression, ever increasing, because the raw images need large amounts of disk space, seems to be a big disadvantage during transmission and storage. It is the process of encoding information using fewer bits than an encoded representation, is also making use of specific encoding schemes. Image compression is the application of Data compression on digital images having plenty of techniques available for itself. This paper analyses Huffman Encoding Algorithm and compares the efficiency of proposed enhanced Huffman algorithm with other common compression techniques like GIF, TIFF and traditional Huffman Scheme. Result proves that Huffman is comparatively better than other mentioned algorithms.

Keywords — Image Enhancement , Huffman, TIFF, GIF, Compression, Encoding.

I. INTRODUCTION

Image compression encoding is to store the image into bit-stream as compact as possible and the decoding is to display the encoded bit-stream back in the monitor as exact as possible as image. Now, consider an encoder and a decoder. When the encoder receives the original image file, the image file converted into a series of binary data, which is called the bit-stream. The decoder then receives the encoded bit-stream and decodes it to a form called the decoded image. If the total data quantity of the bit-stream is less than the total data quantity of the original image, then this is called image compression. Compression algorithms determine compressed bit rate – i.e., the bandwidth required to transmit video images – and the hard disk space consumed within a digital video recorder (DVR) or network video recorder (NVR) – i.e., storage space. The main objective of this paper is to compress images by reducing number of bits per pixel required to represent it and to decrease the transmission time for transmission of images and then reconstructing back by decoding using the Huffman codes. The organization of the paper is as follows: section I gives the introduction, section II has need for compression, section III discusses the types of redundancy, section IV deals with types of compression, section V explains the implementation and section VI concludes with conclusion and future work.

II. NEED FOR COMPRESSION

An effective seismic data compression method would be useful in storing industrial reflection seismic data [1][8]. These data are stored on hundreds of thousands of digital tapes costing millions of dollars. One particular seismic data acquisition company uses more digital tapes than any concern in the world, except for the U.S. Federal government. Since little original seismic data is ever discarded and since it is best kept in climate-controlled long-term storage, large rooms are cut out of salt domes to store tapes containing old reflection seismic data. Any technique [2] that would reduce seismic data storage space significantly would save the industry millions of dollars. The change from the cine film to digital methods of image exchange and archival is primarily motivated by the ease and flexibility of handling digital image information instead of the film media [3][4][5]. While preparing this step and developing standards for digital image communication, one has to make absolutely sure that also the image quality of coronary angiograms and ventriculograms is maintained or improved. Similar requirements exist also in echocardiography. At a closer look one observes that ad hoc approaches to image data compression have been applied in most digital imaging systems for the catheterization laboratory all the time [6]. An example [7][9] is recording the x-ray images with a smaller matrix of just 512 by 512 pixels (instead of the 1024 by 1024 pixel matrix often applied for real-time displays). In order to objectively assess these and other techniques of image data compression, some systematic knowledge of the trade-offs implied in different modes of image data reduction is mandatory. The simplified diagram for compression and decompression is given in Fig. 1 below.

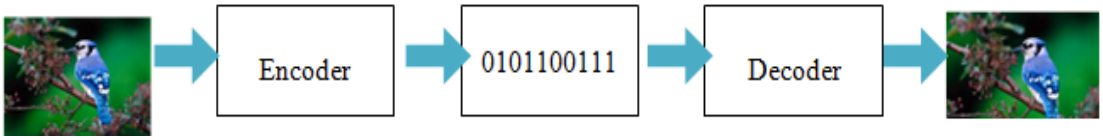


Fig 1 Flow of Compression Algorithms

© 2014, IJARCSSE All Rights Reserved

III. DIFFERENT TYPES OF REDUNDANCY

Redundancy, in theory, is the number of bits used to transmit a message minus the number of bits of actual information in the message [10]. Informally, it is the amount of wasted "space" used to transmit certain data. Data compression is a way to reduce or eliminate unwanted redundancy, while checksums are a way of adding desired redundancy for purposes of error detection when communicating over a noisy channel of limited capacity [11]. For image compression there are three types of redundancies i.e., Coding Redundancy, Inter pixel Redundancy and Psycho visual Redundancy discussed in detail below.

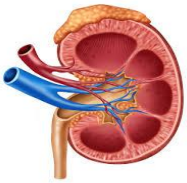
A. Coding Redundancy

It is present when less than optimal code words are used. The code-words are ordered in the same way as the intensities that they represent; thus the bit pattern 00000000, corresponding to the value 0, represents the darkest points in an image and the bit pattern 11111111, corresponding to the value 255, represents the brightest points. If the size of a code-word is larger than is necessary to represent all quantization levels, then we have coding redundancy. An 8-bit coding scheme has the capacity to represent 256 distinct levels of intensity in an image. But, if there are only 16 different grey levels in a image, the image exhibits coding redundancy because it could be represented using a 4-bit coding scheme. Coding redundancy can also arise due to the use of fixed-length code words.

B. Inter pixel Redundancy

Inter pixel redundancy results from correlations between the pixels of an image [12]. The intensity at a pixel may correlate strongly with the intensity value of its neighbours because the value of any given pixel can be reasonably predicted from the value of its neighbours. Much of the visual contributions of a single pixel to an image are redundant; it could have been guessed on the bases of its neighbour values. We can remove redundancy by representing changes in intensity rather than absolute intensity values . For example, the differences between adjacent pixels can be used to represent an image. Transformation of this type is referred to as mappings. They are called reversible if the original image elements can be reconstructed from the transformed data set.

Table 1 Sample Decoded Images with Original Images

Name of the Image	Original Image	Decode image form Compressed
EYE		
LIPS		
LEGS		
KIDNEY		

C. Psycho Visual Redundancy

Psycho visual redundancy is due to data that is ignored by the human visual system (i.e. visually non essential information) [13][14], the image with 256 possible gray-levels, for example. Uniform quantization can be applied to four bits or sixteen possible levels. The resulting compression ratio is 2:1. The false contouring is also present in the previously smoothened regions of the original image. The significant improvements which are possible with quantization, that takes advantage of the peculiarities of the human visual system. The method used to produce this result is known as improved gray-scale (IGS) quantization. It recognizes the eye's inherent sensitivity to edges and breaks them up by adding to each pixel a pseudo-random number, which is generated from the order bits of neighbouring pixels, before quantizing the result.

© 2014, IJARCSSE All Rights Reserved

IV. TYPES OF COMPRESSION

Compression can be divided into two categories, as Lossless and Lossy. In lossless compression, the reconstructed image after compression is numerically identical to the original image [3]. In lossy compression scheme, the reconstructed image contains degradation relative to the original. Lossy technique causes image quality degradation in each compression or decompression step. In general, lossy techniques provide for greater compression ratio than lossless techniques. The following are the some of the lossless and lossy data compression techniques: Run Length Encoding, Huffman Encoding, Arithmetic Encoding, Set Portioning In Hierarchical Tree, Portable Network Graphics, Tagged Image File Format, Graphics Interchange Format, LZW coding are techniques are in the category of lossless compression. . JPEG, JPEG 2000, Wavelet, Fractal Compression, High Definition Photo are few in the category of Lossy compression. Here this papers compares the proposed scheme of Huffman with Giff,

V. IMPLEMENTATION OF PROPOSED WORK

The Huffman algorithm is very simple and is most easily described in terms of how it gets generated in the prefix-code tree. It starts with a forest of trees, one for each message. Each tree contains a single vertex with weight wi = pi. It gets repeated until only a single tree remains. Then select two trees with the lowest weight roots (w1 and w2). Huffman coding suffers from the fact that the uncompression need to have some knowledge of the probabilities of the symbols in the compressed files. This can need more bits to encode the files if this information is unavailable. Compressing the file requires two passes. First Pass: find the frequency of each symbol and construct the Huffman tree. Second Pass: compress the file. This image compression method is well suited for gray scale (black and white) bit map images. This method can be improved using adaptive Huffman coding technique that is an extension to Huffman coding. Each iterations of Huffman's algorithm reduces the size of the problem by 1. The pseudo code is given as follows.

A. PSEUDOCODE

```
Huffman(W, n)
Input:  A list W of n (positive) weights.
Output: An extended binary tree T with weights
     taken from W that gives the minimum weighted path length.
Procedure:
        Create list F from singleton trees formed from elements of W
 WHILE (F has more than one element) DO
        Find T1, T2 in F that have minimum values associated with their roots
        Construct new tree T by creating a new node and setting T1 and T2 as its children
        Let the sum of the values associated with the roots of T1 and T2 be associated with the root of T
        Add T to F
 OD
 Huffman := tree stored in F
```

B. ALORITHM AND WORKING OF HUFFMAN ENCODING

1. Initial Data Sorted By Frequency
2. Combine the Two Lowest Frequencies, F and E, to Form A Sub-Tree of Weight 14.
3. Move it into its Correct Place.
4. Again Combine the two Lowest Frequencies C and B, to Form a Sub-Tree of Weight 25.
5. Move it into its Correct Place.
6. Now The Sub-Tree With Weight, 14, and D are combined to make a tree of Weight, 30.
7. Move it to its Correct Place.
8. Now the two Lowest Weights are held by the "25" And "30" Sub-Trees, So Combine them to make one of Weight, 55.
9. Move it after the A.
10. Finally, Combine the A and the "55" Sub-Tree to produce the Final Tree.

The encoding table is:

A 0 C 100 B 101 F 1100E 1101 D 111

Table 2 Comparative Results of Huffman Algorithm

File Name	Size	Huffman	TIFF	GIF	Enhanced Huffman
Eyes	1132 kb	55%	76%	79%	83%
Lips	1022kb	48%	72%	77%	86%

© 2014, IJARCSSE All Rights Reserved

Legs	987kb	51%	68%	80%	81%
Nose	555 kb	59%	77%	82%	84%
Heart	1600 kb	66%	72%	79%	82%
Chest	1900 kb	59%	65%	72%	85%
Skull	2200 kb	63%	69%	76%	89%
Heart 1	2565 kb	61%	67%	75%	84%
Heart 2	4258 kb	51%	58%	69%	79%
Kidney	5693 kb	62%	69%	75%	90%

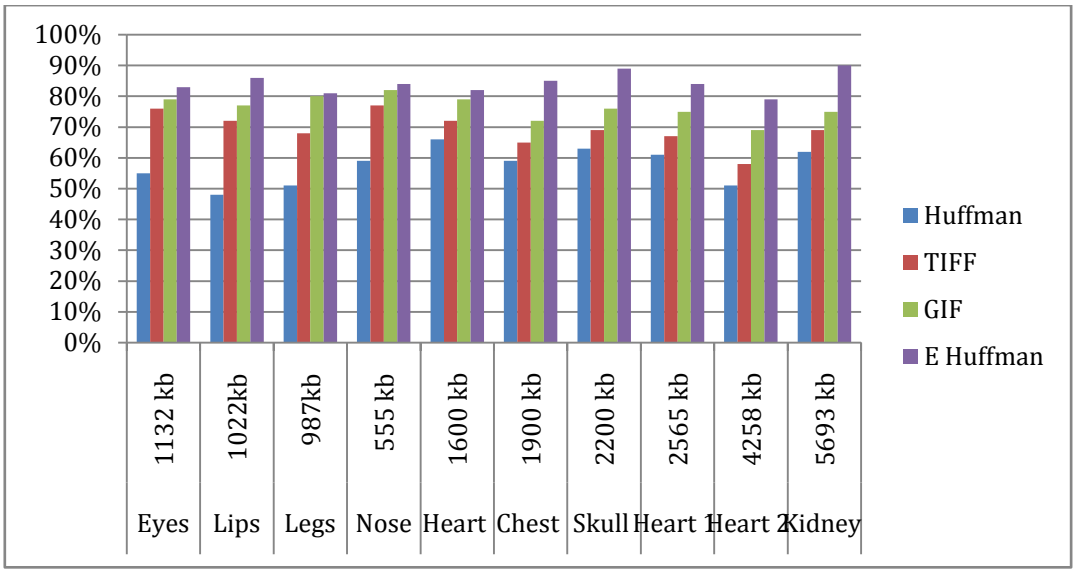


Fig. 2 Diagramatic Representation of Comparative Results of Huffman

VI. CONCLUSION AND FUTURE WORK

Data compression means the process of reducing the amount of data required for transmission or storage of a piece of information in the format of text, graphics, video, sound, etc., because the raw images need large amounts of disk space, seems to be a big disadvantage during transmission and storage. It is the process of encoding information using fewer bits than an encoded representation, is also making use of specific encoding schemes. Image compression is the application of Data compression on digital images having plenty of techniques. This paper analysed Huffman Encoding Algorithm and compares the efficiency of proposed enhanced Huffman algorithm with other common compression techniques like GIF, TIFF and traditional Huffman Scheme. Result proved that Enhanced Huffman Encoding Scheme is comparatively better than other commonly used traditional algorithms.

REFERENCES

[1] Mamta Sharma, "Compression Using Huffman Coding", International Journal of Computer Science and Network Security, Vol.10, No.5, pp. 133-141, May 2010.
[2] Darrel Hankersson, Greg A. Harris, and Peter D. Johnson Jr., Introduction to Information Theory and Data Compression, CRC Press, 1997.
[3] D.A. Huffman, "A Method for the Construction of Minimum-Redundancy Codes", Proceedings of the I.R.E., pp. 1098–1102. Huffman's original article, September 1952.
[4] Terry Welch, "A Technique for High-Performance Data Compression", Computer, June 1984.
[5] J. Ziv and A. Lempel, "A Universal Algorithm for Sequential Data Compression", IEEE Transactions on Information Theory, May 1977.
[6] Dzung Tien Hoang and Jeffery Scott Vitter, "Fast and Efficient Algorithms for video Compression and Rate", Control, June 20,1998.
[7] D.E. Knuth, "Dynamic Huffman Coding", Journal of Algorithms, Vol. 6, pp. 163-180, 1983.
[8] V. Bhaskaran and K. Konstantinides. Image and video compression standards, Kluwer Academic Publishers, Boston, MA, 1995.
[9] J.S. Vitter, "Design and analysis of Dynamic Huffman Codes", Journal of the ACM, Vol. 34 (3), pp. 823-843, 1987.

© 2014, IJARCSSE All Rights Reserved

[10] Vo Dinh Minh Nhat, Sung Young Lee, "Two-Dimensional Weighted PCA algorithm for Face Recognition," Proceedings of IEEE International Symposium on Computational Intelligence in Robotics and Automation, pp. 219-223, June 27-30, 2005.

[11] J. Ziv and A. Lempel, ``Compression of Individual Sequences Via Variable-Rate Coding", IEEE Transactions on Information Theory, Vol. 24, pp. 530--536, 1978.

[12] Chiu-Yi Chen, Yu-Ting Pai, Shanq-Jang Ruan, "Low Power Huffman Coding for High Performance Data Transmission", International Conference on Hybrid Information Technology, Vol .1 (9-11), pp. 71 – 77, 2006.

[13] Lakhani, G, "Modified JPEG Huffman coding", IEEE Transactions Image Processing, Vol. 12(2), pp. 159 – 169, 2003.

[14] Rudy Rucker, "Mind Tools", Houghton Mifflin Company, 1987.

© 2014, IJARCSSE All Rights Reserved