H.264 Based Video Compression

Pranob K Charles[1] Ch.Srinivasu[2] V.Harish[3] M.Swathi[3] Ch.Deepthi[3]

[1](Associate Professor, Dept. of Electronics and Communication Engineering, KLUniversity.)
[2](Professor, Dept. of Electronics and Communication Engineering, KLUniversity.)
[3](Project Students, Dept. of Electronics and Communication Engineering, KLUniversity.)

Abstract

A raw video takes up a lot of space for storage and bandwidth for transmission. Uncompressed video from a video recording device may take up about 20 MB per second of video. Thus, there is a clear necessity for an efficient mechanism which enables the video to be stored and transmitted over limited bandwidths. Compression can be broadly classified into two types, lossy compression and lossless compression. In lossy compression, the video is compressed to very low bit rates which considerably reduce the quality of the video whereas in lossless compression no loss of data is present and it is more visually appealing than lossy compression. This paper discusses about H.264 video compression standard which is a motion-block oriented codec standard developed by ITU-T. The aim of this algorithm is to provide visually better video quality with fewer amount of information transfer. Various motion estimation algorithms are also studied.

Keywords – H.264, lossy compression, motion estimation

I. Introduction

Video compression technology has enabled many services like video calling, video conferencing, internet streaming. The two major parameters of consideration for designing a standard are Bandwidth and storage space. New video compression standard should be compatible and more advanced than the previously developed ones. It is obvious that a new standard should use less bandwidth for transmission and minimum storage space .

Two major groups that are involved in creating the video standards are Video Coding Experts Group (VCEG) from ITU-T and Moving Picture Experts Group (MPEG) from International Organization for Standardization (ISO)/International Electro-technical

Commission (IEC). A previously developed video codec standards, and their released year are mentioned here. The

first video codec standard is H.120 developed in 1984 by ITU-T which is a lossy compression. The newer standard defined in 1990 by ITU-T, is called H.261. It is popular for

video conferencing and telephony. ISO/IEC also released MPEG-1 part 2 in 1993, which is implemented in video CD. In 1995 ITU-T and ISO/IEC jointly released their new video codec known as H.262, popular in Video CD/DVD/Blu-ray, Video broadcasting. In 1995, H.263 and 1999 MPEG-4 part 2 are developed by ITU-T and ISO/IEC respectively. Finally, H.264 video codec was announced by ITU-T and ISO/IEC, which is popular for all the modern day applications [8].

H.264 is widely accepted video codec standard and is also known as MPEG-4 part-10 AVC (Advanced Video Coding). It took approximately four years to have a Trade off in terms of coding efficiency, complexity on hardware, and cost which are balanced in new proposed standard [1]. Some of its features are high definition resolution, interlaced/progressive scan mode, variable frame rate, high bit rate, supporting I, B, & P-frames, 9 different prediction modes, variable size block matching motion estimation, and motion compensation, quarter pixel accuracy, adaptive length coder, and image enhancement filter. H.264 is compatible with almost all kinds of recent video coding tools which enable it to standout from other video coding standards.

Ii. Basics of video coding

The human visual system works as RGB model. Cones sense the red, green and blue color components and merges them to create a image. Also the human visual system is less sensitive to color component than luminance/brightness of image. Another fact of human eye is that if we pass at-least 15 frame per second in front of human eye, human eye will not differentiate frame boundary and it seem as movie. These limitations of human eye are advantageous to our video compression system.

The raw video is compressed mainly by removing the redundant data present in the video and only transmitting or storing the required amount of data.There are four types of redundancies present in a raw video.

A. Temporal Redundancy

A video is nothing but a group of consecutive pictures called frames. Temporal redundancy is the redundancy present in between two consecutive frames in a video.

B. Spatial Redundancy

The correlation between the neighboring pixels present in the same frame leads to spatial redundancy.

Fig 1. Spatial Redundancy data

It can be clearly seen from the figure 1 that the pixels present inside the box almost have same pixel values.

C. Color Spatial Redundancy

A human eye is insensitive to sudden color variations but is effective in identifying the changes in brightness of a picture. We can save bandwidth if we only send selective samples of chrominance components.

D. Phyco-visual Redundancy

It refers to a high-level limitation of HVS. When HVS visualizes an image, the partial information of image is important. The rest of the information is less important, so we can represent less important data with less number of bits.

Iii. H.264 Video Coding

H.264 video coding consists of Inter frame prediction block, motion compensation block, Discrete Cosine Transfom block, Quantizer and Enctropy Encoder . Figure 2 shows the block diagram of H.264 video coding.

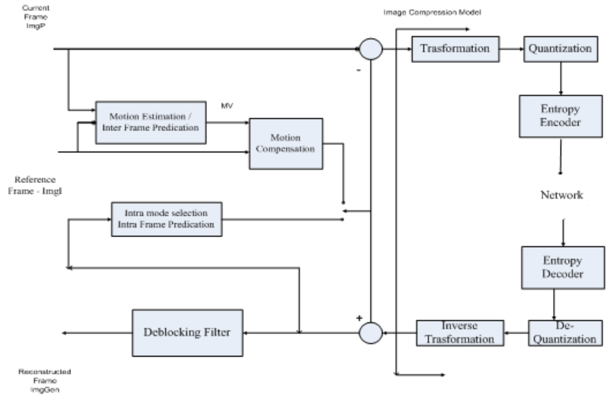

Fig.2. H.264 Video Coding Block Diagram

Motion estimation block is also known as inter frame prediction. The main purpose of the motion compensation block is to remove the temporal redundancy. The prediction of the current frame can be done by using only the previous frame or a group of previous frames.

The residual image which is the output of the entire prediction stage is then sent to DCT block. The DCT of the residual image will convert the image into frequency domain thus differentiating between the high frequency and low frequency components present in the current frame. The main limitation of the human visual system is that it cannot recognize sudden changes in the frames,in other words the human eye is insensitive to high frequency. The quantizer exploits this limitation of the human eye and removes the high frequency components by having a threshold value. Everything that falls below this threshold value will be made zero and the remaining are left intact.

Iv. Prediction and motion estimation

Out of all the blocks present in the coding process, Prediction stage is more computationally intensive. In this paper, we have studied various block matching algorithms and arrived at the best block matching algorithm by considering both the computational intensity as well as the PSNR values. The algorithms that are studied are Exhaustive Search, New Three Step Search, Simple and Efficient TSS, Four Step Search, Diamond Search and Adaptive Rood Pattern Search.[17]

The algorithms are tested for the computational intensity and the Peak Signal to Noise Ratio and the best method is chosen by having a tradeoff between the computations and the PSNR values. The best algorithm should have a good PSNR value and at the same time with minimum possible computations.

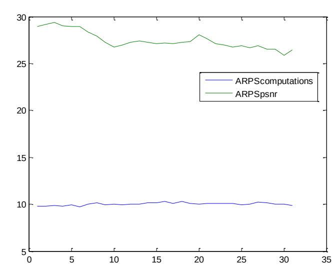

Fig 3. Computations and PSNR of Adaptive Rood Pattern Search

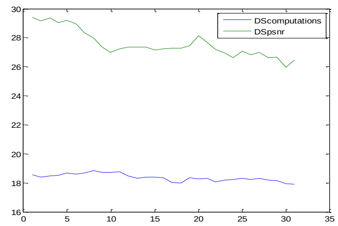

Fig 4. Computations and PSNR of Diamond search

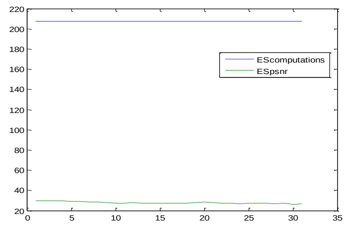

Fig 5. Computations and PSNR of Exhaustive search

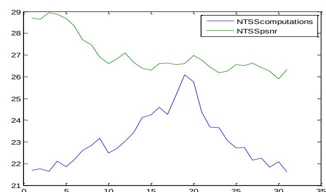

Fig 6. Computations and PSNR of New Three Step Search

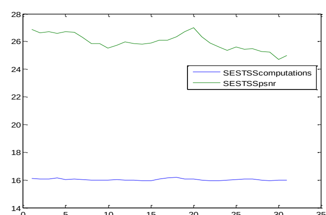

Fig 7. Computations and PSNR of Simple and Efficient TSS

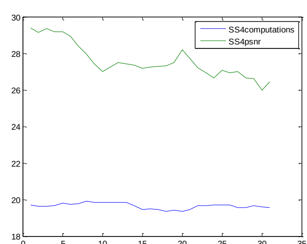

Fig 8. Computations and PSNR of Four Step Search

Four Step Search, Diamond Search and Adaptive Rood Pattern Search prove to be effective with low computations and high PSNR value. When, Computational intensity is critical, Adaptive Rood Pattern Search proves to be effective.

V. Encoding and decoding

The output from the prediction stage is now divided into blocks then a discrete cosine transform is performed on each of the block followed by quantization, zig zag scanning and entropy coding. The block diagram for encoding is given in figure 9.

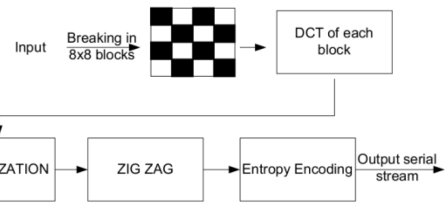

Fig 9. Blocks present in Encoding stage

A. DCT

DCT is performing a math function on the original matrix. It converts a signal into the frequency domain matrix. It represents components in terms of the sum of the cosine function [5]. A typical 8 x 8 DCT matrix element can be represented as shown in figure 10. we can observe that the frequency component is increased from left to right, and top to bottom.

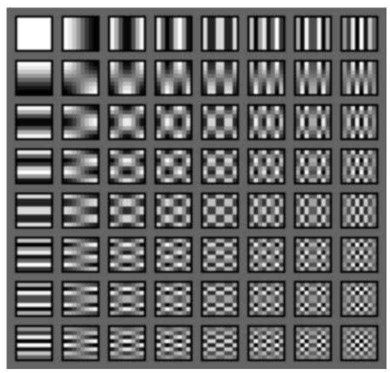

Fig 10. An 8 x 8 DCT matrix in frequency domain

B. Quantization

An 8-by-8 DCT block is ready for the next level of compression, which is known as quantization. It is a process of transformation of continuous set of values to the finite set of small values. When the quantization is performed on the original signal, the part of information is lost from the original signal. It means higher the compression but lower the quality of image, and vice versa.

C. Entropy Coding

The H.264 uses content based adaptive variable length coding, and it is processed in major five steps.

1. Encode the number of coefficient and the trailing ones.
2. Encode the sign of each trailing ones.
3. Encode the level of non-zero coefficient.
4. Encode the total number of zeros before the last coefficient.
5. Encode the location of encoded zeros.

Vi. Results

The PSNR (peak signal to noise ratio) based on MSE (mean square error) is used as a measure of "quality." MSE and PSNR are given by the following relations:

$$MSE = \frac{1}{m * n} \sum_{i=1}^{n} \sum_{j=1}^{m} (x_{i,j} - y_{i,j})^2$$

$$PSNR = 10 \log[(255)^2 / MSE]$$

Figures 11 and 12 show one of the original and decoded frame . We can observe a little degradation in quality of the decoded frame ,the PSNR values for various resolutions are given in table 2.

Table 1 . Computations and PSNR of various Prediction Methods

Method	Computations	PSNR
Exhaustive search	207.4	27.83
Simple and Efficient Three Step Search	16.02	25.97
New Three Step Search	23.09	27.02
Four Step Search	19.65	27.59
Diamond Search	18.36	27.56
Adaptive Rood Pattern Search	**10.01**	**27.48**

Fig 11. Left : Original frame Right: decoded frame

Table 2. PSNR values for various video resolutions

Video size	PSNR
176 x 144	31.5865
352 x 288	65.9162

Vii. Conclusions

H.264 is the most advanced and widely used video coding standard available today. H.264 uses more accurate prediction algorithms and motion compensation techniques to achieve better compression and same quality video at the low bit rate without compromising the image quality.

References

[1] G. J. Sullivan, P. Topiwala, and A. Luthra, "The H.264/AVC Advanced Video Coding Standard: Overview and Introduction to the Fidelity Range Extensions", Presented at the SPIE Conference on Applications of Digital Image Processing XXVII Special Session on Advances in the New Emerging Standard: H.264/AVC, August 2004.

[2] http://www.drtonygeorge.com/video_codec.htm

[3] Thomas Wiegand, Gary J. Sullivan, "Overview of the H.264/AVC Video Coding Standard", IEEE Trans. on Circuits and Systems for Video Technology, vol. 13, no. 7, pp. 560–576, July 2003.

[4] H. Jozawa, K. Kamikura, H. Kotera, H.Watanabe, "Two-Stage Motion Compensation Using Adaptive Global MC and Local Affine MC", IEEE Trans. on Circuits and Systems for Video Technology, vol. 7, no. 1, pp. 75–85, February 1997.

[5] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete Cosine Transform", IEEE Trans. Computers, 90-93, Jan 1974

[6] http://www.cs.uga.edu/~sadiq/pdf/vasim.pdf

[7] Soon-kak Kwon, A. Tamhankar, K.R. Rao, "Special Issue on Emerging H.264/AVC video coding standrad" J. Visual Communication and Image Representation, vol 17.

[8] http://en.wikipedia.org/wiki/Video_compression

[9] K. Peter, "Algorithms,complexity analysis and VLSI architectures for MPEG-4 motion estimation", Kluwer Academic Publishers, Boston, 1999.

[10] Jianhua Lu and Ming L. Liou, "A Simple and Efficient Search Algorithm for Block-Matching Motion Estimation", IEEE Trans. on Circuits and Systems for

Video Technology, vol. 7, no.21, pp. 429-433, April 1997.

[11] R. Li, B. Zeng, and M. L. Liou, "A new three-step search algorithm for block motion estimation" IEEE Transactions on Circuits and Systems Video Technology,vol. 4, pp. 438–443, Aug. 1994.

[12] P. Lai-Man and M. Wing-chung, "A novel four-step search algorithm for fast block motion estimation", IEEE Transactions On Circuits and Systems for Video Technology, vol. 6, pp. 313-317, Jun. 1996.

[13] M.H. Chan,.B. Yu, A.G. Constantinides, "Variable size block matching motion compensation with applications to video coding", IEE Proceeding, Vol. 137, Pt. I , No. 4, pp 205-212, August 1990

[14] Shou-Yi Tseng, Che-How Mak, "A Motion Vector Coding Scheme Based on Bottom-up Merging Procedure", IEEE Transaction on advances in multimedia, pp 125-129, July 2009

[15] L. Kizewski, Student Thesis on "Image Deblocking Using Local Segmentation", School of computer science and software engineering Monash University, November 2004

[16] http://www.vcodex.com/files/H264_intrapred_wp.pdf

[17] http://www.mathworks.com/matlabcentral/fileexchang e/8761

Mr. Pranob K Charles, Associate Professor, currently associated with School of Electrical sciences, Department of ECE, K L University, Andhra Pradesh, India. He received B.Tech in Electronics and Communication Engineering from JNTUH and M.E in Applied Electronics from Anna University, Coimbatore. He is a Life Member in IETE. He authored nearly 10 research papers in various International journals and conferences. His research interests include Image Processing, Video Processing, and Artificial Neural Networks.