ANALYSIS AND APPLICATIONS OF H.264 VIDEO CODING STANDARD

[1] Imran Ullah Khan, Research Scholar,Dept. Electronics &Comm. Engg.,Mewar University, Mewar
[2] M.A.Ansari Member IEEE, Professor & Head, Dept. Electrical, Ins. & Control Engg. GCET, Gr. Noida

Abstract

This paper investigates the effects of many parameters, which are crucial for performance of H.264 video coder .H.264 is the latest international video coding standard. It was jointly developed by the video coding expert group (VCEG) of the ITU-T and moving picture expert group (MPEG) of ISO/IEC. The main goals of the H.264 standardization effort have been enhanced compression performance. H.264 has achieved a significant improvement in rate distortion efficiency relative to existing standards. It uses state of the art coding tools and provide enhanced coding efficiency for a wide range of applications including video telephony , video conferencing , streaming video , digital video authority , digital cinema and many others.

Introduction

We ask authors to follow these guidelines and make the paper look exactly like this document. The easiest way to do this is simply to download this template and replace the content with the text of your manuscript.

Emerging H.264 video coding standard developed and standardized collaboratively by both the VCEG and ISO/IEC MPEG organizations. H.264 represents a number of advances in standard video coding in terms of both coding efficiency enhancement and flexibility for effective use over a broad variety of networks types and application domains. It is widely use for transmission of standard definition (SD) and high definition (HD) TV signals over satellite, cable and storage of high quality SD video signals on DVDs.[2,7]

However an increasing number of services and growing popularity of high definition TV are creating greater needs for higher coding efficiency. Moreover, other transmission media such as Cable Modem, XDSL or UMTS. offer much lower data rates than broadcast channels or high quality video representations within existing digital transmission capacities. Video coding for telecommunication applications has evolved through the development of the ITU-T H.261, H.262 (MPEG-2), and H.263 video coding standards (and later enhancements of the H.263 known as H.263+ and H.263++) and has diversified from ISDN service to embrace PSTN ,mobile wireless networks , and LAN/ Internet network delivery. Throughout this evolution, continued efforts have been made to maximize coding efficiency while dealing with the diversification of network types and their characteristic formatting and less/error robust requirements. In early 1998 the video coding expert group (VCEG-ITU-T SG16) issued a call for proposals on a project called H.26L, with the target to double the coding efficiency (which means halving the bit rate necessary for a given level of fidelity) in comparison to any other existing video coding standards for a broad verity of applications. The first draft design for that new standard was adopted in the October 1999.In December 2001, VCEG and the moving picture expert group (MPEG-ISO/IEC JTC I/SC 29/WG 11) formed a joint video team (JVT), with a charter to finalize the draft new video coding standard for formal approval submission as H.264 in march-2003.

Video formats and quality

Video coding is a process of compressing and decompressing a digital video signal. This paper examines the structure and characteristic of digital images and video signals and introduces concepts such as sampling formats and quality matrices that are helpful to an understanding of video coding. Digital video is a representation of a natural (real world) Visual scene, sampled at point in time to produce a frame (a representation of a complete video scene at that point in time) or a field (consisting of odd-even numbered lines of spatial samples).

Natural video scenes

A typical 'real world' or 'natural' video scene is composed of multiple objects each with their own characteristic shape, depth, texture and illumination. The colour and brightness of a natural video scene changes with varying degrees of smoothness throughout the scene (continues tone).

Video formats

In practice, it is common to capture or convert to one of a set of 'intermediate formats' prior to compression and transmission. The choice of frame resolution depends on the application and available storage or transmission capacity. For examples, 4CIF is appropriate for standard definition TV and DVD video; CIF and QCIF are popular for video conferencing applications; QCIF or SQCIF are appropriate for mobile multimedia applications.

PSNR

Peak signal noise ration (Equation1) is measured on logarithmic scale and depends on the mean squared error (MSE) of between an original and impaired image or video frame, relative to $(2n-1)2$ (the square of the highest possible signal value in the image , where n is the number of bits per image sample.

$$PSNRdB = 10 \log10(2n -1)2/MSE$$

PSNR can be calculated easily and quickly and is therefore a very popular quality measure, widely used to compare the quality of compressed and decompressed video images.

For a given image or image sequence, high PSNR usually indicates high quality and low PSNR indicates low quality. A distorted version of the original image from PSNR rating don't necessary correlate with true subjective quality. Sometime a human observer gives a higher importance to the face region and so is particularly sensitive to distortion in area.

This section provides details for typesetting your manuscript according to the formatting guidelines set for the IJATER Journal. Please use a 10-point Times New Roman font for the entire document. The goal is to have a 10-point text, as you see here. The text should be "full justified."

Technical overview of H.264

The H.264 design supports the coding of video (in 4:2:0 choma format) that contains either progressive or interlaced frames, which may be mixed together in the same sequence. Generally a frame of video contains two interleaved field the top and the bottom field. The two fields of an interlaced frame which are separated in time by a field period (half the time of the frame period), may be coded separately as two field pictures or together as a frame picture. A progressive frame should be coded as a single frame picture; however, it is still considered to consist of two fields at the same instant in time.

Network abstraction layer(NAL)

The VCL, which is described in the following section, is specified to represent, efficiently, the content of the video data. The NAL is specified to format that data and provide header information in a manner appropriate for conveyance by the transport layers or a storage media. All data are contained in NAL units , each of which contains an integer number of bytes. A NAL unit specifies a generic format for use in both packet oriented and bit stream systems. The format of NAL units for both packet oriented transport and bit stream delivery is identical , except that each NAL unit can be preceded by a start code prefix in a bit stream oriented transport layer.

video coding layers

The video coding layer of H.264 is similar in script to other standards such as MPEG-2 video. It consist of a hybrid tem-poral and spatial prediction, in conjunction With transform coding. Figure-1 shows the H.264 encoder. In common with earlier coding standards, H.264 does not explicitly define a Codec but rather defines the syntax of an encoded video bit stream together with the method of decoding this bitstream.

Basic of video coding

A digitized video signal consists of a periodical sequence of images called frames. Each frame consist of a two dimensional array of pixels. Each pixel consist of three colour components R, G and B. Usually, pixel data is converted from RGB to another colour space called YUV in which U and V components can be sub-sampled. A block-based coding approach divides a frame into macro blocks (MBs) each consisting of say 16x16=256 Y components. Each of three components of a MB, a hybrid of three techniques are used: prediction, transformation & quantization and entropy coding. This procedure works on a frame of video.

Prediction

Prediction tries to fine a reference MB that is similar to the current MB under processing so that, instead of the whole current MB, only their (hopefully small) difference needs to be coded. Depending on their reference MB comes from, prediction is classified into inter-frame prediction and intra-frame prediction. In an inter-predict (P &B) mode, the reference MB is somewhere in a frame before or after the current frame, where the current MBs resides. It could also be some weighted function of MBs from multiple frames. In an intra-predict (I) mode, the reference MB is usually calculated with the mathematical functions of neighbouring pixels of current MB. The difference between the current MB and its prediction is called residual error data (residual). It is transformed from spatial domain to frequency domain by means of discrete cosine transform. Because human visual system is more sensitive to low frequency image and less sensitive to high frequency image, quantization is applied such that more low information is retained while more frequency information is discarded. The third and final type of compression coding is entropy coding. A variable length coding gives shorter codes to more probable symbols and long codes to less probable ones such that the total bit count is minimized. After this phase, the output bit stream is ready for transmission or storage. There is also a decoding path in the encoder. One has to use the reconstructed frame as the reference for prediction since in the decoder side only the reconstructed frame is available. The restored residual data is obtained by performing inverse transformation. Adding the restored residual to the predicted MB, the reconstructed MB is obtained that is then inserted to the reconstructed frame. Now the reconstructed frame can be referred to either by the current I type compression or P-type or B-type prediction. Prediction exploits the spatial or temporal redundancy of a video se-

quence so that only the difference between the actual and predict instead of the whole image data need to be encoded. There are two types of prediction intra prediction of I-type frame and inter prediction for P-type (predictive) and B-type (bidirectional predictive) frame.

Intra Prediction

Their exist high similarity among neighbouring blocks in video frame. Consequently, a block can be predicted from its neighbouring blocks. The prediction is carried out by means of a set of mathematical functions. In H.264, an I-type 16x16, 4:2:0 MB has its luminance component (one 16x16) and chrominance (two 8x8) separately predicted. There are many ways to predict a macro block. The luminance component may be intra predicted as one single Intra 16x16 blocks or 16 intra 4x4 blocks. When using the Intra 4x4 case, each 4x4 utilizes one of nine prediction modes (one DC mode and eight directional prediction modes). When using the Intra 16x16 case, which is well suited for smooth image area, a uniform prediction is performed for the whole luminance component of a macro block. Four prediction modes are defined. Each chrominance component is predicted as a single 8x8 block using one of the four modes.

Inter Prediction

High quality video sequences usually have high frame rate at 30 or 60 frames per second. Therefore, two successive frame in a video sequences are very likely to be similar. The goal of the inter prediction is to utilize this temporal redundancy to reduce data needed to be encoded. When encoding frame t, we only need to encode the difference between frame (t-1) and frame t, instead of whole frame t. this is called motion estimated inter frame prediction. In most video coding standards, the block based motion estimation (BME) is used to estimate for movement of a rectangular block from the current frame. For each MxN pixel current block in the current frame, BME compares it with some or all possible MxN candidate blocks in the search area in the reference for the best match. The reference frame may be a previous frame or the next frame in P-type coding or both in B type coding. A proper matching criteria is to measure the residual calculate by subtracting the current block from the candidate block , so that the candidate block that minimizes the residual is chosen as best match. The cost function is called sum of absolute difference (SAD), which is sum of pixel by pixel absolute difference between predicted and actual image. There are three new features of motion estimation in H.264: variable block size, multiple reference frames and quarter pixel accuracy. [8,9]

Variable block size – block size determine tradeoff between the residual error and the number of motion estimation (FBSME) spends the same efforts when estimating the mo-

tion of moving objects and background (no motion). This method causes low coding efficiency. Variable block size motion estimation (VBSME) uses smaller block size for moving objects and larger block size for background, to increase the video quality and the coding efficiency. H.264 specifies multiple block sizes starting from 16x16 going down to 4x4. Multiple reference frames- In previous video coding standards, there is only one reference for motion estimation. In H.264, the number of reference frames increases to 5 for P frame and to 10 (5 previous frames and 5 next frames) for B frame. More reference frame result in smaller residual data and therefore lower data rate.
Nevertheless, it requires more computation and more memory traffic. Quarter pixel accuracy – In previous coding standards, motion vector accuracy is half pixel at most. In H.264, motion vector accuracy is down to quarter pixel and results in smaller residual data.

Compensation

Corresponding to prediction, there is also two kind of compensation, intra compensation for I-type frame and inter compensation for P-type and B-type frame.
3.2.1.3 Transformation and Quantization
Discrete cosine transform (DCT) is a popular block based transform for image and video compression. It transforms the residual data from time domain representation to frequency domain presentation. Since most image and video are low frequency data, DCT can centralize the coding information. The main transformed coefficients and to reduce the coding information. The H.264 standard employs a 4x4 integer DCT.

In-loop filter

One of the advantages of block-based video coding is that discontinuity is likely to appear at the block edge. In order to reduce this effect, the H.264 standard employs the deblocking filter to eliminate blocking artifact and thus generates a smooth picture of the deblocking filter for highly compressed [10].

Entropy coding

There are two popular entropy coding methods; variable length and arithmetic coding. The H.264 standard defines two coding methods: context adaptive variable length coding (CAVLC) and context based adaptive arithmetic coding (CABAC).For baseline profile, only CAVLC is employed.[2,7]

H.264 and its Parameters

As in H.263 the video coding layer (VCL) of H.264 uses translation block based motion compensation and transform based residual coding. It also features scalar quantization

with adjustable quantization step size for output bit rate control, zigzag scanning and run length VLC coding of quantized transform coefficients. However there are significant differences in the details .H.264 uses a more flexible and efficient model for motion compensation. Use of multiple reference pictures and seven different block sizes are supported for performing motion compensation. Motion vectors can be specified with 1/8 pixel accuracy. The standard also specifies use of an improved deblocking filter within the motion compensation loop in order to reduce visual artifacts and improve prediction.
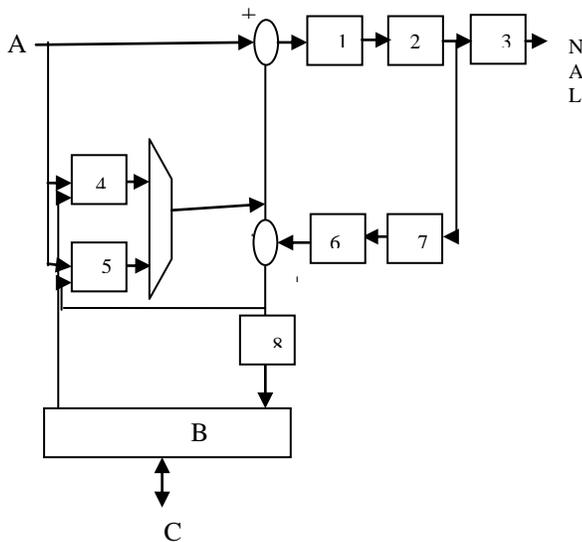
Figure1: Block diagram of H.264 coder

A: Video Input
B: Memory Interface
C: External Memory
1: DCT-Discrete Cosine Transform
2: Q-Quantizer
3: Lossless Compression
4: Motion Estimation/Compensation
5: Intra Prediction
6: IDCT-Inverse Discrete Cosine transform
7: Q-1
8: Deblocking Filter
9. NAL: Network Adaptation Layer

It features an integer transform instead of DCT.This transform reduces blocking and ringing artifacts, and also eliminates encoder-decoder mismatch in the reverse transform which was present in the earlier floating-point DCT.

The standard also features a more complex and efficient context based arithmetic coding (CABAC) for entropy coding of the quantized transform coefficients.

Applications

The increasing no of digital video services and the growing popularity of high definition TV are creating the need for higher video coding efficiency. MPEG-2, developed almost a decade ago, is currently used for such applications. However, the coding efficiency of MPEG-2 is not adequate for lower bit rate transmission channels such as 3GPP UMTS mobile units, cable modems, and DSL. Even DVD storage, for digital video, can only contain high quality standard definition video signals. H.264 has been developed to address a wide range of applications with varying bit rates, resolutions, quantities and services. Given the profile and many levels, the standard attempts to be as widely applicable as possible. Some of the application areas for H.264 include the following:
Video conferencing over DSL, Ethernet, LAN, Wireless and mobile networks. High definition streaming video over the internet, DSL, LAN even wireless and mobile networks
High definition broadcasts video to wireless and mobile networks ,High definition video storage on DVD,Cable TV over copper wired networks.

Direct broadcast satellite and terrestrial digital video services Multimedia mailing –a state-of art video codec H.264, delivers stunning quality at remarkably low data rates. Ratified as part of the MPEG-4 standard (MPEG-4 part10), this ultra efficient technology gives you excellent results across a broad range of bandwidths, from 3G for mobile devices to iChat AV for video conferencing to HD for broadcast and DVD. Master quality, minimal files –H.264 uses the latest innovations in video compression technology to provide incredible video quality from the smallest amount of video data. This means you see the crisp, clear video in much smaller files, saving you bandwidth and storage cost over previous generations of video codecs.H.264 delivers the same quality as MPEG-2 at a one third to half the data rate and up to four times the frame size of MPEG-4 part 2 at the same data rate.H.264 is truly a sight to behold.

Conclusion

H.264 represents a major step forward in the development of video coding standards. It typically outperforms all existing standards by a factor of two and especially in comparison to MPEG-2, which is the basis for digital TV system world wide. H.264 is now mandatory for HD-DVD and bluary specifications (the format for high definition DVDs) and ratified in the latest version of the DVB (digital video broadcasters) and 3GPP (3rd Generation partnership project) standards. Numerous broadcast, cable, video conferencing and consumer electronics companies consider H.264 the video codec of choice of their new products and services.

References

[1] "Image and Video coding –Emerging standards and beyond "IEEE transaction on circuits and systems for video technology,vol.8, No.7,1998 by Barry Ghas kell,Atulpari,Lawrence Rabiner, Lean Bottim Pattrik Haffner.

[2] Joint video Team (JVT) ITU-T VCEG, Joint Model Reference Software version 4.2.

[3] http://spmg.ece.ubc.co/h.263pleas/h.263 plus.html

[4] D.Marpe, T.Wiegand and G.J.Sullivan,"The H.264/MPEG Advance video coding standard and its applications"IEEEcommunications Magazine, vol.44, no.8, pp.134-144, Aug.2006.

[5] T.Wiegand: Joint final committee Draft Doc.JVT-E146d37ncm, joint video team (JVT) of ISO/IEC MPEG and ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16), November 2002.

[6] Video coding for low bit rate communicationITU-T, Recommendation H.264/H.263, ver.1, 1995.

[7] "Compressed video communications", by Abdul H. Sadka,published by John Wiley and sons,2002.

[8] Ralf Schafer,Thomas Wiegand,and Heiko Schwarz, "The emerging H.264/AVC standard",EBU TECHNIC AL REVIEW-January 2002 Heinrich Hertz Insti tute,BERLIN,germany

[9] Iain E.G. Richardson"H.264 and MPEG-4 Video Com pression",Video Coding for next generation multime dia.John Wiley & Sons Ltd.,PO19 8SQ,England 2003.

[10] ITU-T Recommendation H.264,ISO/IEC 13818-2(MPEG-4):Generic coding of moving pictures andas sociated audio information,Part-2:Video ITU-t and ISO/IEC JTC1,November 1994.

Acknowledgments

Biographies

IMRAN ULLAH KHAN obtained B.Tech Degree (with Honors in Electronics and Communication Engineering) and M.Tech (Communication and information systems) from Aligarh Muslim University, Aligarh.He is persuing Ph.D. from Mewar University,Chittorgarh. He is currently working as a Assistant professor, Department of Electronics and Communication Engineering, Maharana Institute Professional Studies , Kanpur. Presently Working as Coorinator(student Affairs). Previously he has worked as HOD (Department of Electronics and Communication Engineering) in Pranveer singh institute of Technology ,kanpurand Faculty of Engineering and Technology, Agra College, Agra and has been involved with teaching and guiding the students in the areas of Signals and Systems, Integrated circuits, Communication Systems etc. at undergraduate level. He is also a member of IEEE (Institute of Electrical and Electronics Engineers).Several papers were presented and published in National Conference. Rated as an Excellent Faculty for the session2007-08 and 2008-09 in Kanpur institute of Technology, Kanpur. He has ten years of teaching experience. His two books is published with Acm learning and I.K.Intrnational publications.

imranuk79@gmail.com

DR.M.A.ANSARI obtained B.Tech Degree (with Honors in Electronics and Communication Engineering) from Aligarh Muslim University, Aligarh and M.Tech (Electronics) from I.I.T. Roorkee.He is Ph.D. from I.I.T. Roorkee.

He has more than eleven years experience in teaching of Engineering subjects. Presently working as Professor & Head in the Department of Electrical, Instrumentation & Control Engineering at Galgotias College of Engineering & Technology (UP Technical University), Gr. Noida, India.I have also worked as Head of the Electrical and Electronics Department, College of Engg Roorkee, India, Coordinator of Alumni Section and Branch Counselor of IEEE. I have also worked as Coordinator of Student Exchange Programme (MOU) with the Wilkes University, USA and College of Engineering Roorkee. Currently working as head of the EE/ICE departments and the professor in charge of student's affairs & branch counselor of IEEE at Galgotias College of Engineering & Technology, Gr. Noida, India. ma.ansari@ieee.org