Study on the Core Technology of New Video Coding Standards H.264/AVC

Dr.T.Charan singh
Assoc Professor,EEE SBIT
Khammam,TS,India.
Charanphd734@gmail.com

Mrs.T.Samatha
Asst.Professor,EEE,SBIT
Khammam,TS,India
samsai1791@gmail.com

Mrs.M.Sirisha
Asst. Professor,EEE,SBIT
Khammam/TS/India
sirishalaxman79@gmail.com

Abstract

Here, we provide a quick primer on the fundamentals of H.264/AVC, a popular codec and decoder. Motion estimation and compensation, in-frame and inter-frame prediction, integer transform and quantization analysis, entropy coding methods, deblock filter, new photographic image type, aspect-oriented IP and wireless environment, etc. are all aspects of the H.264/AVC standard that we have studied in depth. The H.264/AVC standard has reconciled coding efficiency with image quality, and its effect is clear; however, many of the advantages gained have come at the expense of computational simplicity. As a result, researchers will focus on simultaneously improving coding efficiency and reducing computational complexity in future work.

Keywords:

H.264/AVC; Video coding; Intra-frame and inter-frame prediction; Discrete Cosine Transform (DCT) integer transform; entropy coding; deblock filter.

Introduction

Both the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) have contributed to the present international standards for the video encode, which are administered by the International Telecommunication Union (ITU-T). ITU-T's proposed standard for video encoding is known as the H.26X series, and includes filename extensions like "H.261," "H.263," and "H.264." MPEG-1, MPEG-2, and MPEG-4 are all examples of ISO/IEC's MPEG-X video encoding standards. Real-time video communication (such as video conferencing or a videophone) makes extensive use of the H.26X family of standards; video storage (DVD), transmission, and streaming media make extensive use of the MPEG family of standards. With the exception of H.262/MPEG-2, the two groups typically create their own standards in isolation from one another. Resuming their partnership from 1997, ITU-T VCEG and ISO/IEC MPEG formed the Joint Video Team (abbreviated "JVT") with the goal of creating the next generation of video encoding standards, H.264/AVC, which was eventually released in May of 2003. The system was designated H.264 by the ITU-T and ISO/IEC 14496-10/MPEG-4 AVC. To accommodate various video applications, including video calling, video conferencing, video storage, broadcasting, and monitoring, the standard's primary objective is to devise a simple and effective coding technology with a high compression ratio and the ability to be easily transmitted over the internet.

H.264/AVC encoding/decoding fundamentals

The Coder and Decoder How They Operate

Figures 1 and 2 depict the coding and decoding processes involved in H.264/AVC, respectively.

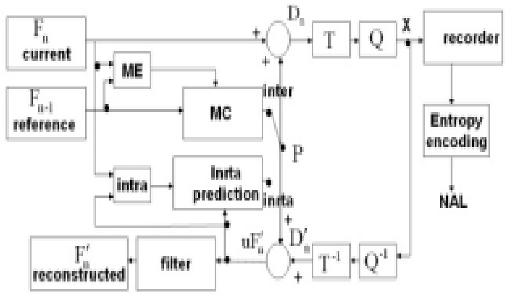

Fig 1 The block diagram of the encoder of H.264/AVC

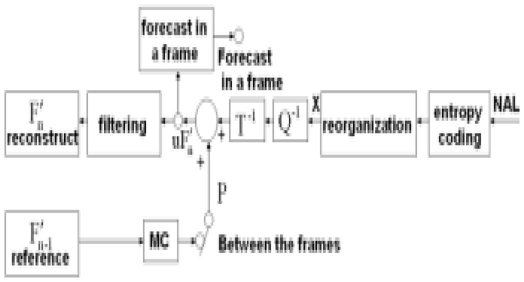

Fig 2 The block diagram of the decoder of H.264/AVC

In Fig 1, the input frame/field Fn is handled by the encoder with the macro block unit of original image 16×16 pixel. Firstly, the input frame/field Fn is handled with the method of forecasting-coding in a frame or between the frames. If the method of forecasting-coding in a frame was adopted, its forecasting value PRED (represented with P in Fig 1) was derived from the encoded reference area for Motion Compensation (MC) in the current image, the reference image is represented with F′ n-1. In order to increase the forecasting accuracy and compression ratio, the actual image can be chosen in the frames of encoded/decoded reconstruction and filtering in the past/future (displayed order). The difference of forecasting value PRED and current block produced a residual block Dn, we got a series of transformation parameters X after Block Transform and Quantization, the transformation parameters X encoded by the entropy combine with some required information relating to the decoding, such as Prediction Mode Quantization parameters and Motion Vectors, into a compressed bit stream for transportation and storage via NAL (Network automatically adapting layer). In order to provide further forecasting reference image, the encoder must have the function of image reconstruction. Therefore, we must handle the residual image by the inverse Quantization and inverse Transform treatment and obtain D′ n, D′ n added to the forecasting value P is equal to uF′ n (un-filtering frame). We designed a loop filter to remove the noise generated by the encoding/decoding loop and increase the image quality of reference frame, which increase the performance of image compression. The filtered output F′ n is the reconstructing image which can be used as the reference image. Known from Fig 1, NAL of the encoder exports a compressed bit stream of H.264/AVC in Fig 2, a series of quantitative transformation parameters X were obtained by the entropy coding, we must handle the residual image by the inverse Quantization and inverse Transform treatment and obtain D′ n. The decoder which takes advantage of the head message came from the bit stream generates a forecasting block PRED, it is the same as the original PRED of encoder. When PRED of the decoder added to the residual difference D′ n is equal to uF′ n, after filtering, we get the restored image F′ n finally.

Syntactic analysis of bit stream

In the output bit stream of the encoder, the basic unit of data is the sentence element. Each sentence element consists of several bits, it represents a particular physical meaning, for example, Macroblock type, Quantitative Parameters, etc. The syntax represents the structure of syntax element, the semantic illustrates the specific meaning of semantic element. All the video encoder standards formulate the work flow of encoder/decoder by defining the syntax and semantic. In the output code stream, each bit belongs to one of the sentence elements. In other words, a code stream consists of a series of syntax elements in sequence. The code stream has no specific content to control or synchronous except syntax element. In the code stream defined by H.264/AVC, syntax elements are organized into hierarchical structure, respectively describe all levels of information. The hierarchical structure of syntax elements can save code stream effectively. For example, in a picture, many areas have the same data. If each area carried these data respectively, the code stream must be wasted. An effective approach is to extract the public information from these images and form the image level of syntax element. The syntax elements in H.264/AVC describe five levels of information including sequence, picture, area, macro block and subblock respectively. The biggest difference of hierarchical structure in H.264/AVC is to have cancelled sequence layer and image layer, most of syntax elements originally belonging to series and image head are drifted out to form two layers of reference set including sequence and picture, the rest is put into the area layer. The new adding syntax elements on the area layer identify the numbers of parameter sets, each area carries its own basic information including the number of pictures, size and so on. In encoding, H.264/AVC formulates that these independence data units such as parameter sets and area should be put into a group as complete as possible to transmit.

Core technology of H.264/AVC standards Although H.264/AVC

standards adopted a mixed coding method based on block, it also uses a large number of different techniques, its video encoding performance is far superior to any other standards. Its main techniques include the following: (1) Each video image is divided into pixel macro blocks of 16×16. The dividing method has the video image handled in pixel macro block. (2) Takes advantage of the relevance of time domain. The relevance of time domain lies in those continuous image blocks, which makes it need to encode those differentials in coding time. Generally we take advantage of the relevance of time domain through motion estimation and motion compensation. A pixel block derives motion vectors from a preceding frame encoded or a few of preceding frame in search of the relevant pixels, but the motion vectors in encoding/decoding end are used to forecast the current pixel block. (3) Makes use of the airspace redundancy of residual error. After motion estimation, the encoding end only needs the encoding residual error, namely encodes the difference between the current block and corresponding forecasting block. The encoding process includes the following steps: transformation, quantization, scanning output and entropy encoding. (4) Other techniques. Besides those we have just described, H.264/AVC also includes: oversampling relationship between traditional chromaticity data and luminous intensity data of 4ĩ2ĩ0; block motion vector; motion vector surpassing the image edge; size partition of transformation block; graded quantization; I, P and B image type, etc. The differences between H.264/AVC and other encoding methods mainly embodied in the following:

Motion estimation and motion compensation

Firstly, H.264/AVC has adopted the macro block partitions and sub-partitions methods of different sizes and shapes. The luminous intensity value of a macro block of 16×16 can be divided in accordance with unit 16×16, 16×8, 8×16 or 8×8, but if the unit of 8×8 was selected, it can be sub-divided in accordance with unit 8×8, 8×4, 4×8 or 4×4.

Intra prediction and inter prediction

The preceding video encoder standards all adopted the inter prediction method, however only the intra encoding image was called I image. The I image transforms the numerical value of pixel block directly; the treatment results will have a large number of redundant information contained in I image to low compressing efficiency. H.264/AVC adopted a new intra prediction model employing the correlation of adjacent pixel based on the same nature possibly owned by the adjacent pixel. We can forecast by the left of the current pixel block and the top pixels (encoded the reconstructed pixels), only encode the differentials between the actual value and forecasting value, so we can use as few bit number as possible to express the pixel block information of intra encoding. The luminous intensity value in the standard of H.264/AVC has 9 kinds of 4×4 block and 4 kinds of 16×16 block intra prediction models, however 4 kinds of chromaticity models of 8×8 is the same as 4 kinds of luminous intensity models of 16×16.

The inter predictions are used to reduce the time domain correlation of image, accurately forecast the next flame to reduce the transmitting data quantity by employing many flames method of reference and smaller motion prediction method. Each luminous intensity macro block is divided into different shapes' region describing motion, the dividing methods have 4 kinds of 16×16, 16×8, 8×16, 8×8. When the model of 8×8 was selected, it can be further divided into 4 sub-regions of 8×8, 8×4, 4×8 and 4×4.Each contains its own motion vectors, each motion vector and selected region information must be transmitted by encoding. Therefore, when a large region was selected, the data quantity describing motion vectors and selected regions decreased, but the residual error after motion compensation will increase; when a little region was selected, the residual error will decrease, the prediction became more accurate, but the data quantity describing motion vectors and selected regions increased. A large region is fit for reflecting the homogeneous part between the frames, a little region is fit for describing the detail part between the frames.

Integer DCT transform and quantization

H.264/AVC uses the integer transform similar to the Discrete Cosine Transform (DCT) of 4×4 to transform the residual result of motion estimation and intra prediction from the time domain to the frequency domain, all operators use integer algorithm, the transform core is mainly addition and shift. In the whole process of transform and quantization, H.264/AVC only carries out the integer algorithm of 16bit and a multiplication operation, not the floating-point transform similar to MPEG-2 and MPEG-4. Therefore, H.264/AVC has a series of virtues such as good effect and fast computation (only the addition and shift operators), its inverse transform process without mismatch problems. Meanwhile, the transformation of block size from 8×8 to 4×4 can lessen the block effect and ringing effect. Although the preceding standards of video encoding/decoding took advantage of the quantization principle to compress the code, the quantization of H.264/AVC has its uniqueness, here, the quantization is a very important step for data compression. The transformation coefficient of H.264/AVC is quantified through un-extended classified quantization. Its basic formula is:

$$Z = round\left(\frac{Y}{Q_{step}}\right)$$

Value of the input coefficient (Y) and the quantizing step (Qstep) are denoted as (Z, Y, Qstep). Quantitative parameter (QP) determines quantification type for each macroblock out of 52 possible types. QP goes up by 1, hence the quantification step goes up by 12.5%; in the previous norms, Qstep went up by a constant amount. Although the quantization of the luminous intensity coefficient may be somewhat harsh at times, the chromatic aberration signals have taken on a more nuanced quantizing step, making their fidelity superior to that of the luminous intensity coefficient. Combining the transform and quantification into a single step greatly decreased the number of operations required to compress the code.

H.264/AVC, an entropy coding technique

using a mixed bag of entropy coding techniques (CABAC, or Context-based Adaptive Binary Arithmetic Coding) and CAVLC (Context-based Adaptive Variable Length Coding) and UVLC (Universal Variable Length Coding). Prior norms embraced UVLC, and all UVLC symbols made use of a code table based on a statistical model of probability. Code words need to have integral units of bit, the correlation between symbols is ignored, conditional probabilities are not applied, and the system may not be particularly realistic despite its seeming simplicity. Because of these flaws, UVLC's compressive effect at moderate and high compression rates suffers.

Deblocking filter

The block structure is an important part of the encoding features. The block effect is a perceived quality loss that arises from the block reconstruction's reliance on quantization errors of pixel values at the block border. The block effect is reduced, both in terms of perceived and objective quality of the decoded picture, and a better reference image is provided thanks to the implementation of a content-based deblock effect filter. The filter is used to even out the transition between adjacent blocks when there is little variation in picture quality across the border; otherwise, it is bypassed. As a result, the block effect is mitigated, strain on the image's contents is prevented, and the bit rate is decreased by a factor of five to ten without sacrificing perceived quality.

According to H.264/AVC specifications, the deblock effect filter is based on the block border of 44, meaning that for each macro block, we need filter 4 horizontal boundaries and 4 vertical boundaries of luminance component of 1616, but for an 88 pixel block, we only need filter 2 horizontal boundaries and 2 vertical boundaries of luminance component.

Brand-new slice types for images H.264/AVC

supports both the standard image slice type and the newer switching image slice types, SP (Switching P) slice and Si (Switching I) slice, that are used for switching between different code streams. After an SP slice and SI slice have been added to a code stream, the decoder can switch fast between streams with the same information but a different coding rate, and it can also play back segments at random and in real time. SP slice use the Inter Prediction technique, which adjusts the size of quantization values to effect a transition across picture streams with varying coding rates. When transmission faults prevent SI from using the Inter Prediction technique, the SI slice is the best available alternative to the SP slice.

Algorithm with a hierarchical structure

H.264/AVC has two levels, the Video Encoding Layer and the Network Abstraction Layer, each of which performs a specific role. The Network Abstraction Layer (NAL) successfully compresses video data for transmission across various networks, while the Video Coding Layer (VCL) effectively describes the video content. Therefore, excellent encoding efficiency and network friendliness are achieved by VCL and NAL independently. After an encoder compresses a code sequence, the resulting data is known as VCL data. Prior to transmission or storage, encoded VCL data is mapped or encapsulated into a NAL unit. A collection of NAL headers matching to the video encoding data is included inside each NAL unit in the form of a Remote Batch Station Processor (RBSP). Fig. 3 displays the NAL unit's sequence structure.

Fig 3 NAL unit sequence

NAL is responsible for using the segmentation format of low-level network to encapsulate the VCL data, including the framing, signalling of logical channels, timing information utilization and sequence ending signal, etc. For example, NAL supports the video transmission format in the circuit switching channel, the video transmission format on the Internet utilizing RTP/UDP/IP. NAL contains its own head information, segment structure information and real load information, namely the top VCL data (If the data partitioning technology was adapted, the data might consist of a few parts). The hierarchical structures are shown in Fig 4

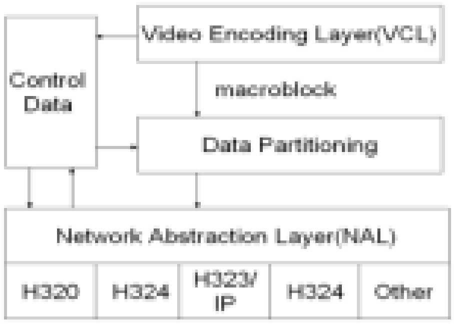

Fig 4 Hierarchy Structure of H.264/AVC

After implementing the aforementioned changes, the H.264/AVC encoding technique became much more efficient, allowing for a 50 percent reduction in bit rate during picture reconstruction while maintaining quality. (8) In an IP and wireless-based setting for the compressed video flow to transmit in an error code and dispersed packet loss environment (such the moving channel and IP channel), H.264/AVC proposals include the tools used to remove errs and boost the transmitting resilience. To prevent transmission problems in

H.264/AVC video streams, the inter picture refresh is employed to finish the time synchronization, while slice structured coding facilitates the space synchronization. After an error code has been detected, the video data in a still image may be used as a resynchronization point. Furthermore, the encoding efficiency may be taken into account, and the encoder can be tailored to the specifics of various transmission channels using the inter macro block refresh and multi-reference macro block permission encoding techniques. H.264/AVC not only adjusts to the channel coding rate by altering the rate at which quantization is performed, but it also handles the transition between rates by splitting data. The idea behind data partitioning is that the encoder should produce separate priority video data to back up QoS in the network. In wireless communications applications, adjusting the quantization accuracy or space/time resolution of each frame allows for support of the maximum bit rate modification of wireless channel. However, in a multi-broadcasting scenario, it is impractical to require the encoder to adapt to a wide range of variable bit rates. As an alternative to MPEG-4's (inefficient) Fine Granular Scalability technique, H.264/AVC uses stream switching inside the SP frame un lieu of hierarchical encoding.

Conclusion

The H.264/AVC standards represent an improvement over their predecessors in many respects, including system structure and efficiency, thanks to the many innovative technologies that went into their development. The H.264/AVC technique reduces the coding rate by about 50 percent compared to earlier standards like H.263 and MPEG-4 while maintaining the same level of visual quality. Additionally, it features a robust error resilience characteristic that allows it to adjust to the challenging conditions of wireless video transmission, including a high packet loss rate and significant disruption. Therefore, H.264/AVC is adaptable to video transmission across networks and allows hierarchical encoding transmission in various network resources, resulting in stable picture quality. Constant tension exists between picture quality and encoding efficiency in the field of video applications. H.264/AVC standards' encoding efficiency was boosted by incorporating the successful encoding techniques of CAVLC and CABAC into its entropy coding technology. H.264/AVC's better performance comes at a cost, however, since its higher computer complexity is a direct result of this superior performance. Estimates place the encoding's computational complexity at at three times that of H.263, with decoding placing it at around two times that of H.263. The next focus of research will be on finding ways to both reduce complexity and increase encoding efficiency.

Reference

[1] Huanchang Qin. Mode Decision Strategy for I-Frames in H.264/AVC Based on Low -Pass Filter[J].Key Engineering Materials Vols. 460-461 (2011) pp 810-815

[2]]Huanchang Qin. Optimization on H.264 Motion Estimation Algorithm in Real-time Video[J]. Advanced Research on Computer Science and Information Engineering CSIE 2011, Part I CCIS 152, pp : 475-481, May.2010.

[3] Huanchang Qin. Improvement on control algorithm of video code rate based on the H.264 standards [J], The Journal of Changchun University of Science and Engineering, 2010ƨd'5Đġ180l181.

[4] Zhaoming Yu, etc, Image encoding standards H.264 technology [M], Beijing: People's telephone and postal press, 2006.

[5] Houjie Bi, New generation of video compression encoding standards——H.264/AVC[M], Beijing: People's telephone and postal press, 2005.

[6] Feng Liu, Encoding technology of video image and international standards [M], Beijing: People's telephone and postal press, 2005.