2012 AASRI Conference on Computational Intelligence and Bioinformatics

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

Huanchang Qin[*]

Department of Physics and Electronics Information Science, Baise University,
No 21 , The Second Road of Zhongshan，Baise City, Guangxi, 533000, China

Abstract

We introduce briefly the basic principle of the standardized coding/decoding H.264/AVC . We have analyzed and studied in detail the core technology of H.264/AVC standard, including motion estimation and motion compensation, forecasts in a frame and between the frames, integer transform and quantization analysis, entropy coding methods, deblock filter, new photographic image type, aspect-oriented IP and wireless environment, etc. The H.264/AVC standard has solved the contradiction between the image quality and the coding efficiency, its effect is obvious, but many advantages acquired exchange for the sacrifice of computing complexity, therefore, to achieve greater coding efficiency will be the next studying emphasis at the same time of reducing the computing complexity.

[*] Corresponding author. Tel.: 0776-2848172、13097765363; fax: 0776-2825195.
E-mail: qin8638@163.com.

1. Introduction

International Telecommunication Union - Telecommunication Standardization Sector (ITU-T) and International Organization for Standardization (ISO)/ International Electrotechnical Commission (IEC) are two different organizations of the current international standards for the video encode. The video encoder standard formulated by ITU-T is called as the video encoder proposal, named as H.26X series, for example, H.261, H.263. The video encoder standard formulated by ISO/IEC is called as MPEG-X, for example, MPEG-1, MPEG-2, MPEG-4. The H.26X series is mostly used for real-time video communication, for example, video conferencing, videophone; The MPEG series is mostly used for video storage (DVD), video broadcast and video streaming media. Except for the joint development of H.262/MPEG-2, in most cases, the two organizations formulate independently the relevant standards. In 1997, ITU-T VCEG cooperated with ISO/IEC MPEG again, they set up the Joint Video Team （JVT）, committed to develop a new generation of the video encoder standard H.264/AVC and published the standard in May 2003. ITU-T named the system as H.264, ISO/ IEC called it 14496-10/MPEG-4 AVC. The main goal of the standard is to try to design simple and effective coding technology owning the ability that is high compression ratio and easy to be transmitted by internet, to meet the occasion of video application, such as video phone, video conference, video storage, broadcast and video monitoring.

2 Basic principle of encoding/decoding of H.264/AVC

2.1. Working processes of the coder and decoder

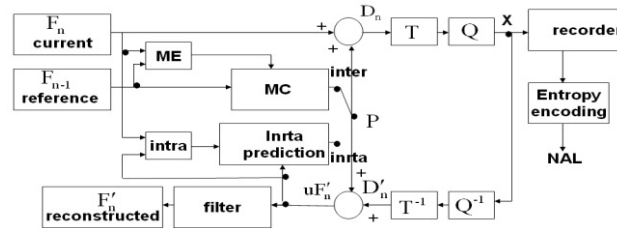The functions of the coder and decoder of H.264/AVC are shown as in figure 1 and 2.

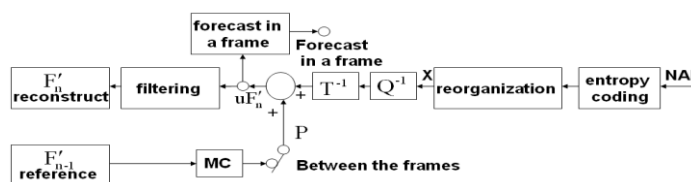Fig 1 The block diagram of the encoder of H.264/AVC

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

In Fig 1, the input frame/field F_n is handled by the encoder with the macro block unit of original image 16×16 pixel. Firstly, the input frame/field F_n 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 D_n, 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 entropycoding, 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.

2.2 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 picture, 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.

The parameter sets are an independent data unit, not dependent on other syntax elements out of the parameter sets. The parameter sets are only cited when the syntax elements of area layer need them, and a parameter set do not correspond to a specific image or sequence, the same sequence of parameter set can be cited by several sequences of image parameter sets. Similarly, the same image parameter set can be cited by several images.

Only when the encoder thinks you need to up the contents of the parameter sets, it will send out new parameter sets. Due to the independence of parameter sets, it can be retransmitted many times or protected employing special technology.

3. Core technology of H.264/AVC standards

Although H.264/AVC standards adopted a mixed coding method based on block, it also use 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 4：2：0; 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:

(1) 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.

(2) 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 are 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.

(3) 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:

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

Where Z is quantizing value, Y is input coefficient value, Qstep is quantifying step. The quantifying step has 52 kinds, the quantitative parameter (QP) decides the quantifying step of each macro block. QP increases 1, accordingly the quantifying step increases 12.5%, in the preceding standards Qstep increases a constant every time. Sometimes the quantization of luminous intensity coefficient is very rough, however the chromatic aberration signals in the quantization process adopted a more delicate quantifying step, the fidelity of chromatic aberration signals became better than the luminous intensity coefficient. Due to putting the transformation and quantifying together, the operations of compressing the code were reduced effectively.

(4) Method of entropy coding

H.264/AVC used 2 kinds of entropy coding method, namely the combination of CAVLC（Context-based Adaptive Variable Length Coding）and UVLC（Universal Variable Length Coding）, CABAC（Context-based Adaptive Binary Arithmetic Coding）. The preceding standards adopted UVLC, all symbols of UVLC all used a code table derived from a statistical probability distribution model. Although it is simple, may have the following fault: probability distributions may not be very fit for reality; probability distributions are still; the correlation of symbol is neglected, conditional probabilities are not made used of; code words must have integral units of bit. These shortcoming affect the compressing effect of UVLC on middle and high compression rate

(5) Deblock effect filter

One of the encoding characteristics based on block lies in its block structure. The quantization errors of pixel value of block boundary form the block effect which affects the subjective quality of image due to the block reconstruction. In order to eliminate the block effect and improve the subjective and objective quality of decoding image and to provide better reference image, the deblock effect filter based on content is introduced. When the difference between the images at the block boundary is little, the filter is used to smooth this difference, if the image characteristics at the boundary are obvious, the filter is not used. Therefore, it can weaken the influence of block effect and avoid straining off the image contents, and at the same time the bit rate in the same subjective quality is reduced 5％～10％.

In the standards of H.264/AVC, the deblock effect filter is based on the block border of 4×4, namely for each macro block, we need filter 4 horizontal boundaries and 4 vertical boundaries of luminance component of 16×16, however we need only filter 2 horizontal boundaries and 2 vertical boundaries of luminance component of 8×8, on the block border, the filtering intensity is related to the encoding patterns, motion vectors and residual values, for a single pixel, the threshold values based on quantizing coefficient may cancel the filter for any single pixel

(6) Type of new image slice

H.264/AVC supports not only the type of traditional image slice but also the switching image slices between the new code streams which are called as SP (Switching P) slice and Si (Switching I) slice.

After the code stream contains SP slice and SI slice, the decoder can transfer quickly between the code streams which have similar contents but different code rate, at the same time supports the random access and quick playback model. SP slice uses the Inter Prediction method to realize the transformation between the image streams of different code rate by changing the size of quantizing values. SI slice is a best slice similar to SP when SI slice can not uses the Inter Prediction method due to transmission errors.

(7) Hierarchy structure of algorithm

The function of H.264/AVC is divided into two layers: the Video Encoding Layer and the Network Abstraction Layer. The Video Coding Layer (VCL) accomplish the effective description for the video content, the Network Abstraction Layer (NAL) accomplish the packed transmission of video data on different networks. Therefore, VCL and NAL separately complete the tasks of high encoding efficiency and network friendship.VCL data is the output video data sequence after the encoder compresses the code.

These encoding VCL data is mapped or encapsulated into NAL unit before VCL data is transmitted or stored.Each NAL unit includes a Remot Batch Station Processor (RBSP), a group of NAL headers corresponding to the video encoding data. The sequence structure of NAL unit shown in Fig 3:

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, signaling 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.
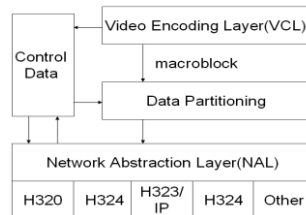

Fig 4 Hierarchy Structure of H.264/AVC

The encoding efficiency of H.264/AVC algorithm was greatly improved after the above measures was adopted, in the reconstruction of the same image quality, H.264/AVC can save 50％ or so bit rate than H.263.

(8) Facing IP and wireless environment

The proposals of H.264/AVC contain the tools, which are used to eliminate errs and increase the

transmitting robustness so that the compressed video flow transmits in the error code and disseminated packet loss environment (such as the moving channel and IP channel).

The inter image refresh is used to complete the time synchronization of H.264/AVC video flow in order to resist transmitting errors, the slice structured coding supports the space synchronization. The video data in a picture provides some resynchronization point after error code.

In addition, the inter macro block refresh and multi-reference macro block permission encoder can not only consider the encoding efficiency but also be adapted to the characteristics of different transmitting channels. Besides it is adapted to the channel code rate utilizing the change of quantization pace, H.264/AVC deals with the change of channel code rate utilizing the data partitioning method.

Generally speaking, the concept of data partitioning is that the encoder generates different priority video data to support the Quality of Service (QoS) in network.

In wireless communications applications, we can support the maximum bit rate change of wireless channel by change the quantization precision or space/time resolution of each frame. However, in the situation of multi-broadcasting, it is impossible to demand the encoder in response to all kinds of changing bit rates. Therefore, H.264/AVC adopted the SP frame of stream switching to take the place of hierarchical encoding, other than the Fine Granular Scalability approach employed by MPEG-4 (low efficiency).

4. Conclusion

Due to the results of using a lot of new technologies, the H.264/AVC standards in all aspects such as system structure and efficiency are superior to the preceding video encoding standards. In the same image quality, the H.264/AVC algorithm saves 50% or so code rate in terms of preceding standards such as H.263 or MPEG-4. At the same time, it has a strong error resilience property to adapt itself to the video transmission of wireless channel with a high packet loss rate and serious disturbance. So H.264/AVC supports the hierarchical encoding transmission in different network resources, achieves a steady image quality and can be adapted to the video transmission between different networks.

In the video application domain, the image quality and encoding efficiency always are a dilemma. The encoding efficiency of H.264/AVC standards is improved because its entropy coding technology adopted the encoding of CAVLC and CABAC, successes to solve the problem. But the price of the superior performance of H.264/AVC is that the computing complexity increased greatly. It is estimated that the computing complexity of encoding is about three times as large as H.263, the computing complexity of encoding is about two times as large as H.263. At the same time of reducing the complexity, to achieve greater efficiency of encoding will be the next priority of study.

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, 2010，（5）：180～181.

[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.