Chapter 2 Introduction to H.264/AVC

H.264/AVC [1] is the newest video coding standard of the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG). The main improvements of H.264/AVC are the increased compression performance and the offering of a network-friendly video representation. The overall coding structure of this standard is basically similar to that of all prior major digital video standards, like H.261, MPEG-1, MPEG-2 / H.262, H.263, and MPEG-4 part 2. In this chapter, we will briefly describe this H.264/AVC standard.

2.1 The H.264/AVC Codec

As shown in Fig 2-1, only the decoder of H.264/AVC is standardized, by imposing restrictions on the bitstream and syntax, and by defining the decoding process of the syntax elements. This scope restriction permits maximal freedom to the encoder. The following description is simplified to provide an overview of encoding and decoding processes.

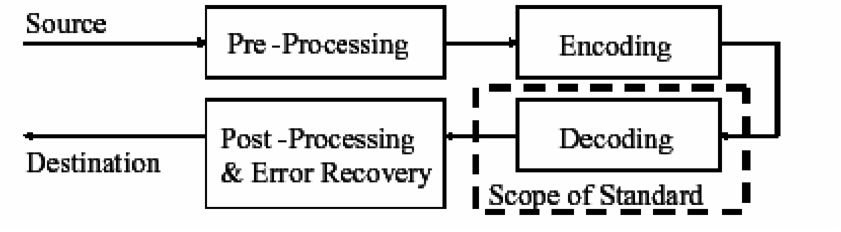

Fig 2-1 Scope of H.264/AVC standardization[2]

2.1.1 The H.264/AVC Encoder

The H.264/AVC design covers a Video Coding Layer (VCL), designed to efficiently represent video contents, and a Network Abstraction Layer (NAL), design to format the VCL representation and to provide header information in a manner appropriate for conveyance by a variety of transport layers or storage media (see Fig 2-2).

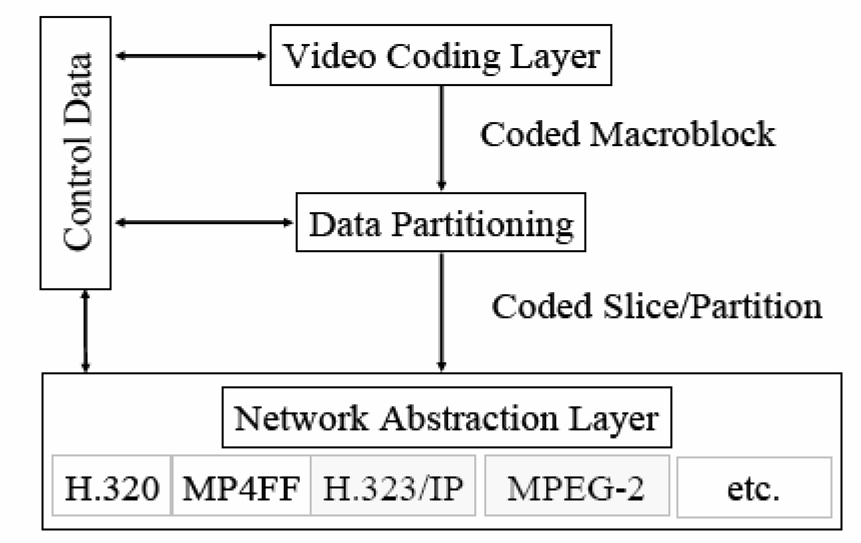

Fig 2-2 Structure of H.264/AVC video encoder. [2]

A block diagram of a typical H.264/AVC encoder is shown in Fig 2-3. With the exception of the deblocking filter, most functional elements (prediction, transform, quantization, entropy encoding) had been presented in these previous standards. However, some important changes in the details of each functional block occur in H.264.

The encoder includes two dataflow paths, a forward path and a reconstruction path. During the forward path, in order to encode the input frame (field) Fn, each macroblock is encoded one by one in either intra or inter mode. After prediction, the residual Dn is produced from the subtraction of the prediction P and the current macroblock. Then, Dn is transformed and quantized to generate the coefficient set X. The coefficient set reordered and entropy encoded with some side information which is required to decode the macroblock. Finally, this encoded data are passed to a Network Abstraction Layer (NAL) for transmission or storage.

In the reconstruction path, the encoder decodes the coefficient to provide a reference frame for prediction. There is a deblocking filter to reduce the blocking effects.

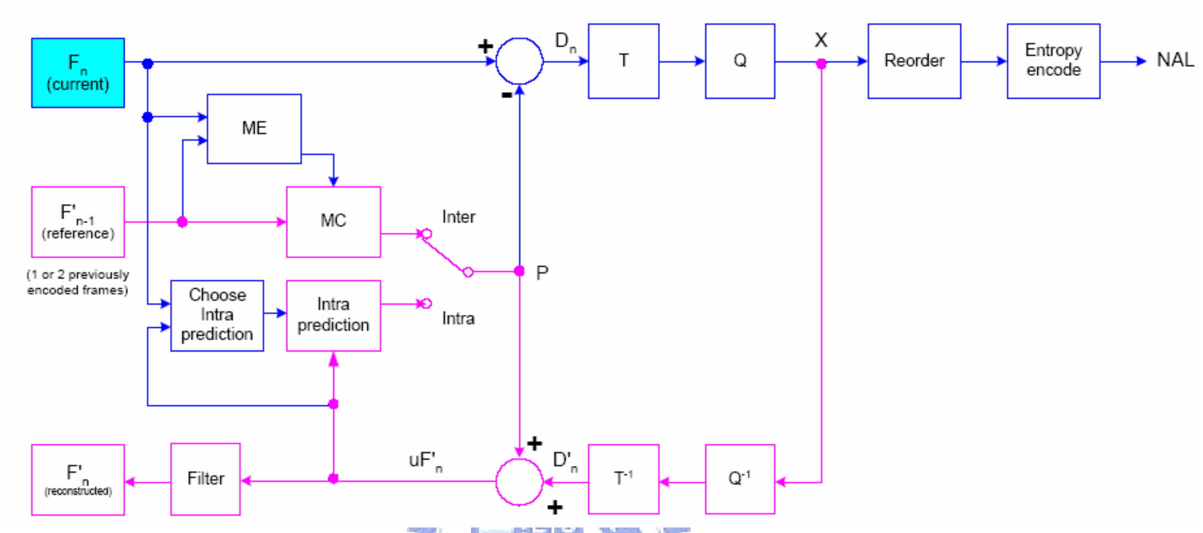

Fig 2-3 H.264 encoder [3]

2.1.2 The H.264/AVC Decoder

Fig 2-4 shows the block diagram of the H.264/AVC decoder. The decoder receives the bitstream NAL and decodes it to give the residual Dn. Using the header information, the decoder creates a prediction block. The prediction block is added to Dn to produce uFn’, which is filtered to create each Fn’.

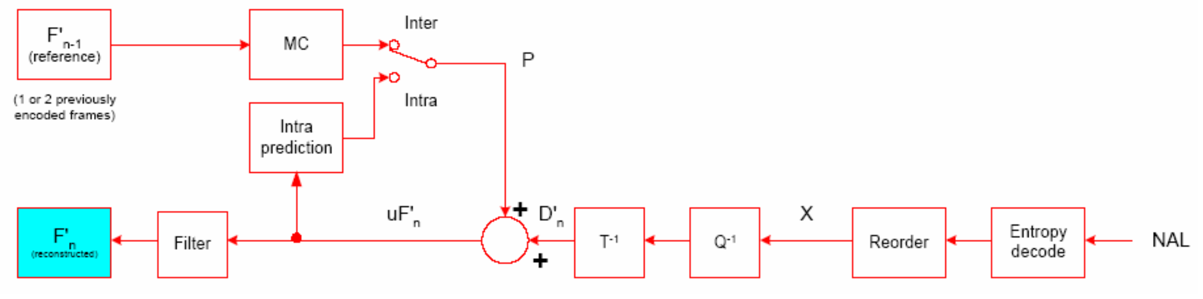

Fig 2-4 H.264 decoder [3]

2.2 H.264 Structure

There are three profiles defined in the H.264 standard, baseline profile, main profile, and extended profile. The ***Baseline profile*** supports intra coding and inter coding, together with entropy coding with CAVLC [4]. The ***Main profile*** supports interlaced video, B-picture, inter coding using weighted prediction and entropy coding using CABAC. The ***Extended profile*** does not support interlaced video and CABAC, but adds modes to enable switching between bitsreams and to improve error resilience. Fig 2-5 shows the relationship among these three profiles.

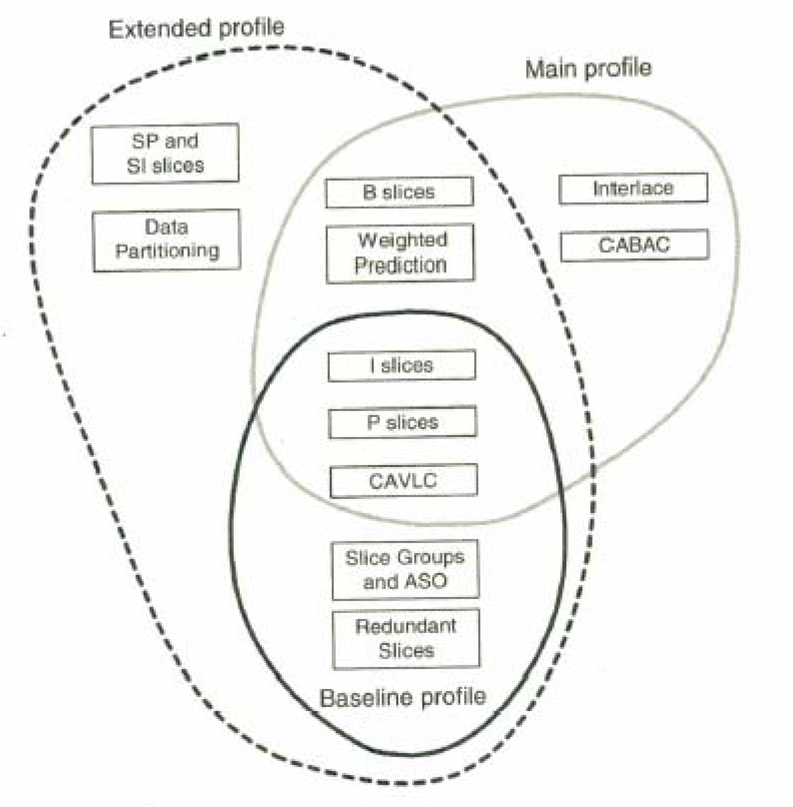

Fig 2-5 H.264 profiles [5]

2.2.1 The Baseline Profile

The Baseline Profile supports I slice and P slice. An I slice contains only intra coded macroblocks, and P slice contains intra coded, inter coded and skipped macroblocks.

2.2.1.1Slices and slice groups

Slices are a sequence of macroblocks processed in the raster scan order. A picture is split into one or several slices as shown in Fig 2-6. One slice can be correctly decoded without the use of others. However, the application of deblocking filter across slice boundaries may need some information from other slices. Besides, flexible macroblock ordering (FMO) can be used in H.264/AVC to partition a picture into several slice groups. Each slice groups containone or more slices. A slice is a sequence of macroblocks which have been processed in the order of raster scan. Using FMO, a picture can be split into many macroblock scanning patterns. Fig 2-7 shows two scanning patterns.

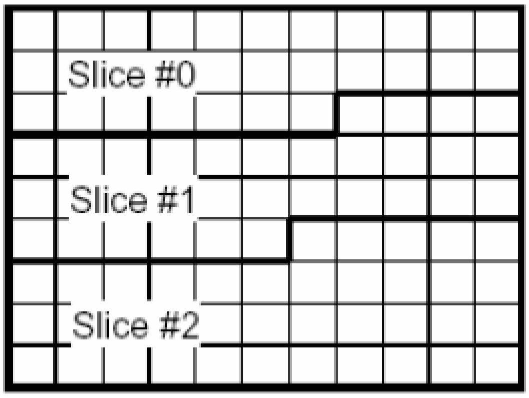

Fig 2-6 Subdivision of a picture into slices [2]

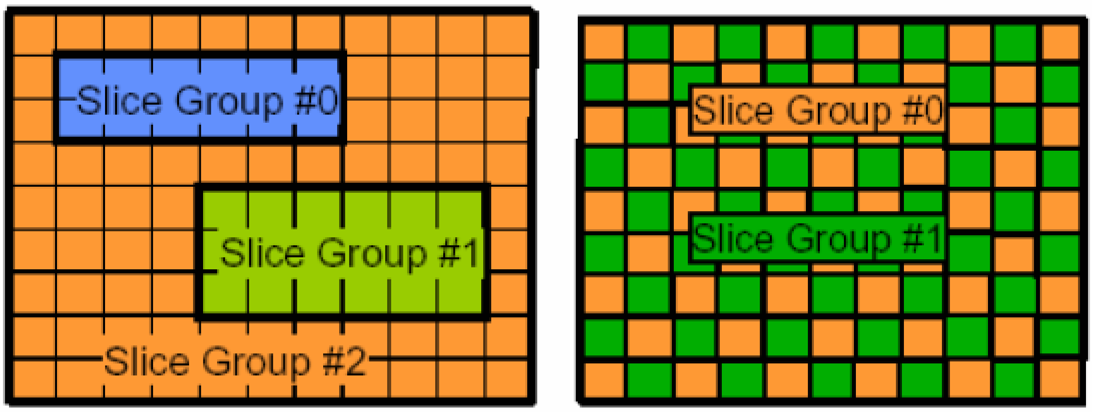

Fig 2-7 Subdivision of a picture into slices using FMO [2]

2.2.1.2 Intra prediction

Each intra block is formed based on encoded and reconstructed blocks. For luma blocks, they are is formed for each 4 × 4 block or each 16 × 16 macroblock. There are nine optional prediction modes for each 4 × 4 luma block, four modes for each 16 × 16 block and four modes for chroma components.

When using intra prediction, each block is predicted by spatially these neighboring samples that have been previously decoded. Fig 2-8 shows five of nine prediction modes. The remaining 4 modes are called vertical-right, horizontal-down, vertical-left, and horizontal-up predictions. They are suited for the prediction of texture in the specified direction. For 16 X 16 blocks, the prediction modes are similar to those of 4 × 4 modes but with only four prediction modes.

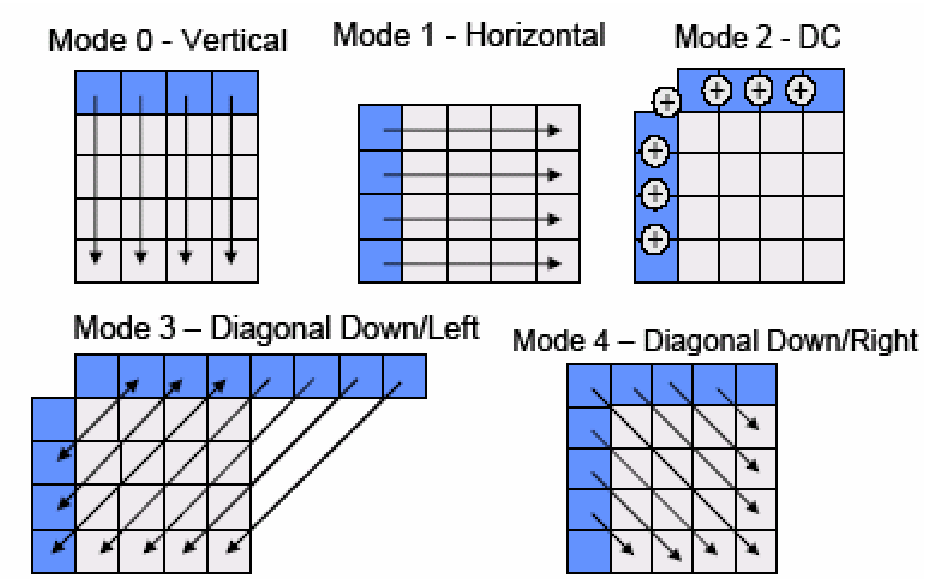

Fig 2-8 Five of nine prediction modes [2]

2.2.1.3 Inter prediction

In the inter prediction mode a prediction block is formed based on previously encoded frame, by using block-based motion compensation. There are four partitions for each macroblock: one 16 × 16 partition, two 16 × 8 partitions, two 8 × 16 partitions, or four 8 × 8 partitions. If the 8 × 8 partitions is chosen, it can be further split into four partitions: one 8 × 8 partition, two 8 × 4 partitions, two 4 × 8 partitions or four 4 × 4 partitions, as shown in Fig 2-9.

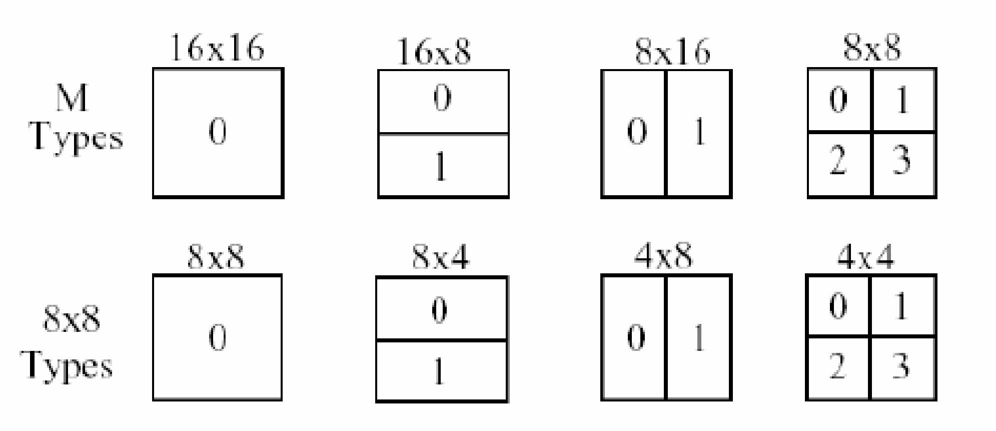

Fig 2-9 Segmentations of macroblock for motion compensation [2]

The accuracy of motion compensation is in units of one quarter of the distance between luma samples. For sub-pixel motion compensation, the corresponding samples are obtained by using interpolation to generate sub-pixel image data.

2.2.1.4 Deblocking Filter

The deblocking filter is applied after the inverse transform in both the encoder and the decoder. This filter decreases blocking effects and improves the visual quality. The filter also improves coding efficiency because a filtered image is often a more faithful reconstruction of the original frame. With this filter, subjective quality is significant improved, as shown in Fig 2-10. This filter also reduces bits rate by 5-10 % typically.

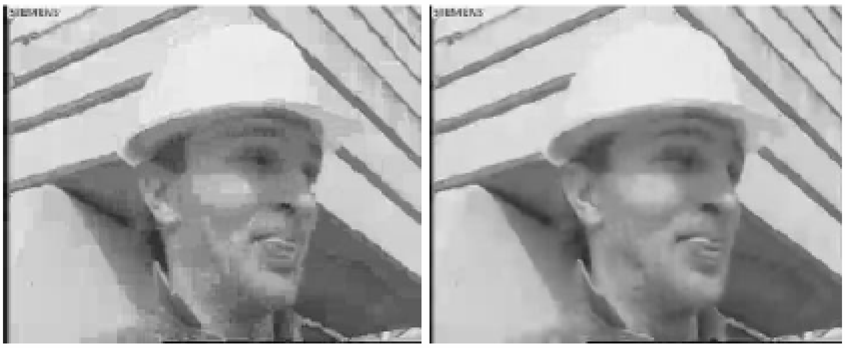

Fig 2-10 Performance of the deblocking filter for highly compressed pictures [2]
Left: without deblocking filter, right: with deblocking filter

2.2.1.5 Transform and Quantization

H.264 uses three transforms for the coding of residual data : the Hadamard transform for 4×4 arrays of luma DC coefficients (for 16×16 intra macroblock only), a Hadamard transform for 2 × 2 arrays of chroma DC coefficients, and DCT based transform for all the other 4 × 4 blocks in the residual data.

H.264 uses a quantization parameter to determine the quantization step for the quantization of transform coefficient. This quantization parameter takes 52 values. Increasing one in the value of QP means an increase of the quantization step size by approximately 12%. An increase of step size by 12% also means a reduction of bit rate by approximately 12%.

2.2.1.6 Entropy coding

Above the slice layer, syntax elements are encoded as fixed or variable-length binary codes. Below the slice layer, elements are coded using Content Adaptive Variable Length Coding (CAVLC) or Content Adaptive Binary Arithmetic Coding (CABAC). In the baseline profile, CAVLC is adopted. In the CAVLC method, the VLC tables are designed to match the conditioned statistics and the coding performance is better than the scheme with fixed VLC tables.

2.2.2 The Main Profile

2.2.2.1 B slices

Each macroblock in a B slice may be predicted from the past or future pictures. Depending on the reference pictures stored in the decoded reference buffer, there are many choices of reference picture for a macroblock in the B slice. Fig 2-11 shows three examples.

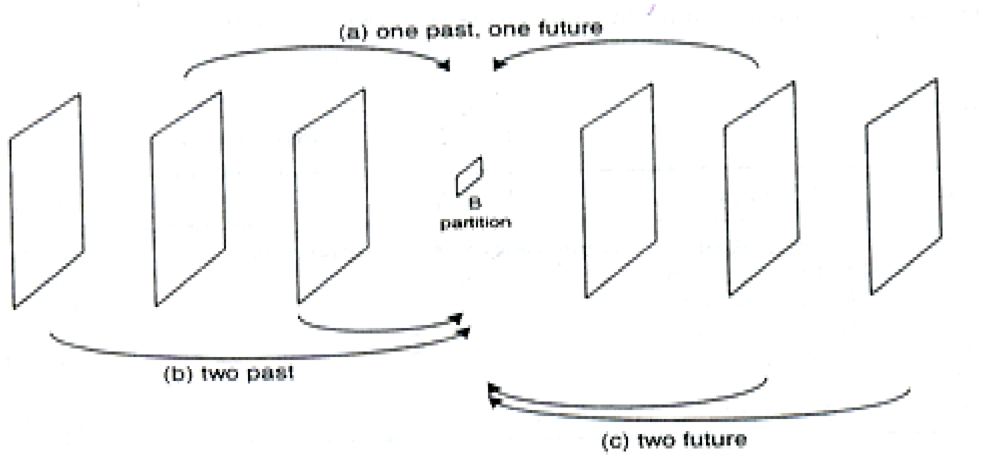

Fig 2-11 (a) past/future (b) past (c) future [3]

2.2.2.2 Interlaced Video

In interlaced frames with of moving objects or camera motion, two adjacent rows tend to show reduced degree of statistical dependency. Fig 2-12 shows the difference between progressive and interlaced frames. If field coding is adopted, the type of picture is specified in the header information. In macroblock adaptive field/frame coding mode (MBAFF), the coding type is specified at the macroblock level. In this mode the current slice is processed in units of 16 luminance wide and 32 luminance high called a macroblock pair. The macroblock pair concept is illustrates in Fig 2-13.

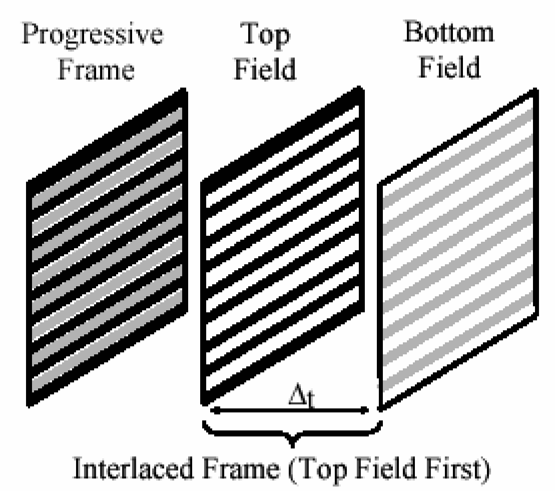

Fig 2-12 Progressive and interlaced frames [2]

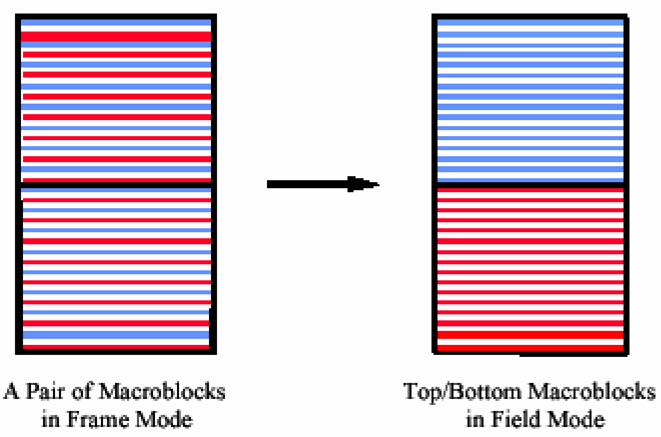

Fig 2-13 Conversion of a frame macroblock pair into a field macroblock pair. [2]

2.2.2.3 CABAC

Context-Adaptive Binary Arithmetic Coding (CABAC) can improve the coding efficiency of entropy coding. CABAC achieves good compression performance through choosing appropriate probability models for each syntax element, adapting probabilities estimates based on local statistics, and using arithmetic coding rather than variable length coding. CABAC provides a reduction in bite rate between 5~15% [2].

2.2.3 The Extend Profile

2.2.3.1 SP and SI slices

SP and SI slice enable switching between two video streams and random access foe decoder. SP slices support switching between similar coded sequences without increased bitrate in I slices. SI slices can switch to I slice that allows an exact match in an SP slice for random access or error control purposes.

2.2.3.2 Data Partition

There are three types of data partition in H.264:

1. Header information

2. Intra slice
3. Inter slice

If Partition 1 is lost, the decoder cannot to reconstruct the slice. On the other hand, Partitions 2 and 3 can be made to be independently decodeable. Hence a decoder may decode Partitions 1 and 2 only, or Partitions 1 and 3 only .

2.3 Error control techniques in H.264

The problems of error resilience and error concealment become more important in video transmission. The goals of error control are to decrease the spatial-temporal error propagation and to increase the subjective visual quality. The details of these techniques will describe in chapter 5.