MPEG-4 Part 10

Advanced Video Coding

Joint Video Coding

H.264

White Paper

July 12, 2004

For Further information,
Please contact:

Ramin Soheili, Ph.D.

SEDA Solutions Corp.
One Daniel Burnham Court
Suite 240C
San Francisco, CA 94109
USA
Phone: (415) – 409 2544
Fax: (415) – 409 2546

ramin@sedasolutions.com

1 Table Of Contents

2 Table of Figures

3 Introduction

H.264/AVC is the newest video-coding standard jointly developed by ITU-T Video Coding Expert and the ISO/IEC Moving Picture Expert Group. The main objective behind this joint development was to achieve enhanced compression performance and superior video quality at comparable bit rates to existing standards. The recommendation has been approved in 2003 by ITU-T as H.264 and by ISO/IEC as International Standard 14496-10 (MPEG-4 part 10) Advanced Coding Standard (AVC).

3.1 Advantages of H.264/AVC

Several features are added to the new standard, which distinguish it from prior video coders. The main advantages are:

- Network friendliness: the network adaptation layer (NAL) facilitates the mapping of the video coding layer which carries the picture samples, onto standard transport layers such as RTP/IP. Effectively, the NAL organizes the video data into fixed number of bytes for efficient transport. The NAL addresses both circuit switched network and packet switched network types.

- Error resilience: the standard provides tools such as slicing and data partitioning to deal with packet loss in error –prone environments such as those that occur in wireless or mobile networks.

- Up to twice the compression for a given quality compared to MPEG-2 and MPEG-4.

- High quality video at low bit rates.

4 Applications

The increasing number 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/AVC has been developed to address a wide range of applications with varying bit rates, resolutions, qualities, and services. Given the profiles and many levels, the standard attempts to be as widely applicable as possible.

Some of the application areas for H.264/AVC 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 broadcast 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

5 Profiles and Levels

The H.264/AVC defines three profiles, each supporting a set of coding tools that a decoder must comply with if using that particular profile. In addition, 15 levels, independent from the profiles, are defined, which place limits to parameters such as picture size, bit rate, and memory requirement. Hence any of the profiles can encode picture sizes from SQCIF (128x96) up to 4096x2304 pixels and up to 240Mbps.

The three profiles are: Baseline, Main, and Extended. The profiles and the grouping of the tools are illustrated in the figure below.

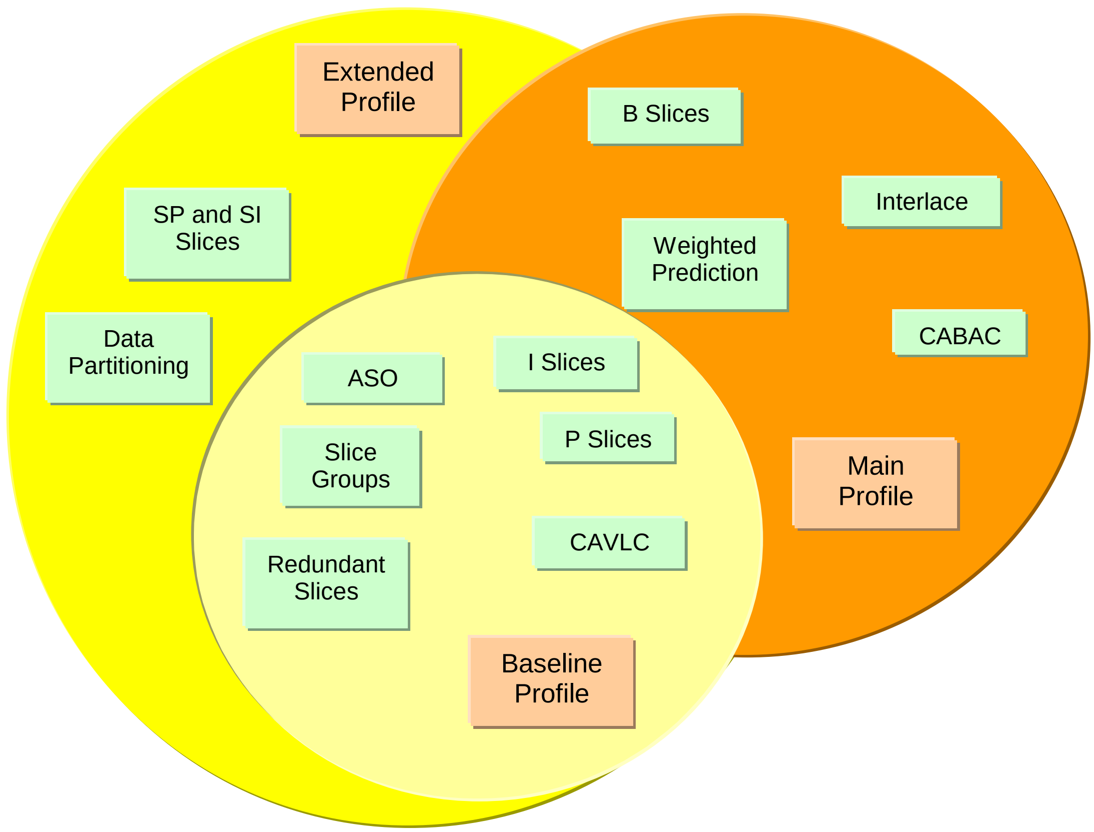

Figure 1 - Three basic profiles of H.264/AVC/JVC Codec

The baseline profile is a subset of the extended profile but not of the main profile. The baseline profile is potentially applicable to videoconferencing, conversational services and wireless mobile networks, where the arbitrary slice grouping and redundant slice tools can combat packet losses in these error-prone environments. The extended profile adds SP and SI slices that allow the decoder to smoothly switch between video streams of different bit rates. A potential application for this profile would be streaming media over IP networks. The main profile does not have error resilience tools but adds interlaced frame capability. This allows this profile to be used for television broadcasting and video storage applications, mainly the entertainment services. However, given the high flexibility of these profiles when different levels are taken into consideration, the application boundaries between the profiles become vague.

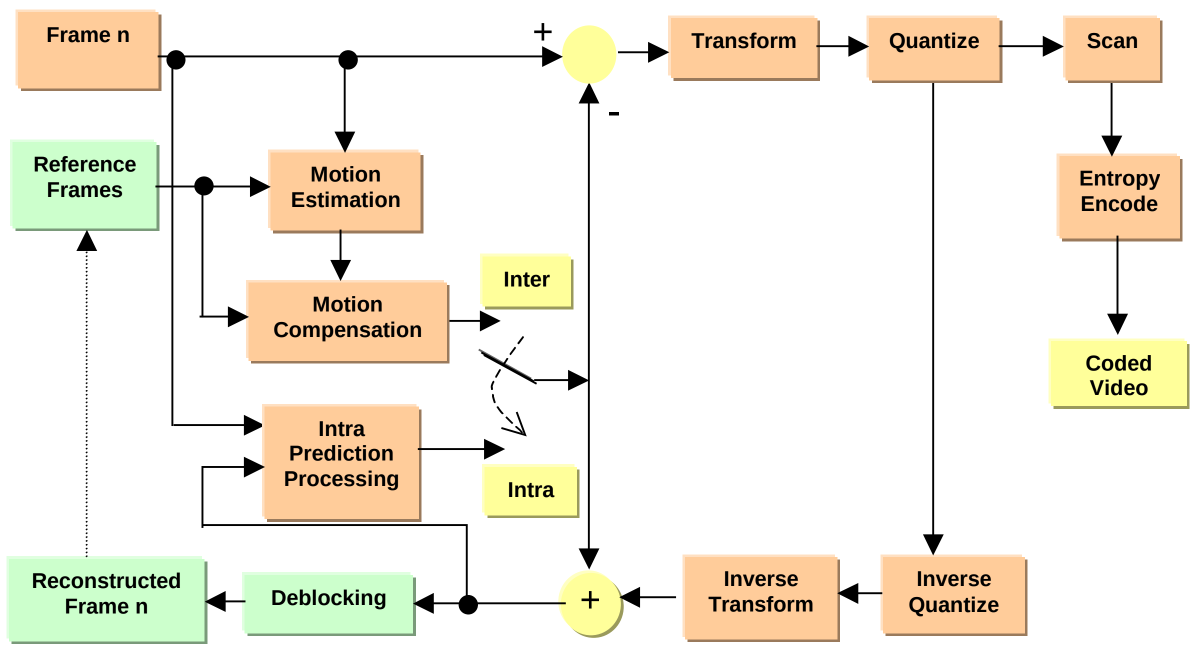

Figure 2 - Block diagram of AVC/JVC/H.264 Video Encoder

5.1 H.264/AVC Coding Features and Highlights

On the surface, the H.264/AVC standard seems to have simplified or eliminated many of the complex features and tools provided by MPEG-4. The non-bit exact DCT and IDCT 8x8 transforms in prior art are replaced with a simpler integer based 4x4 block transform. The new transform is bit exact and involves only additions and shifts. The shape coding found in MPEG-4 is eliminated in the new standard. A fixed scalar quantization is defined instead of a matrix quantization as in MPEG-4. Only one scan type is used for each of the interlaced and non-interlaced modes instead of 3 types in MPEG-4. However, the H.264/AVC standard has added many powerful spatial and temporal prediction tools to improve coding efficiency, a mandatory in-loop deblocking filter for visual and coding improvement, two new frame types for smooth bit-stream switching and synchronizing, and several error resilience tools such as arbitrary slice ordering. Depending on which tools are utilized, the complexity can be up to 3 times that of the MPEG-4 of a similar profile/level. The complexity can be much higher depending on the motion estimation algorithm used and the number of intra prediction modes included. Although the different prediction modes are optional in the encoder, the decoder must implement all tools in its declared profile.

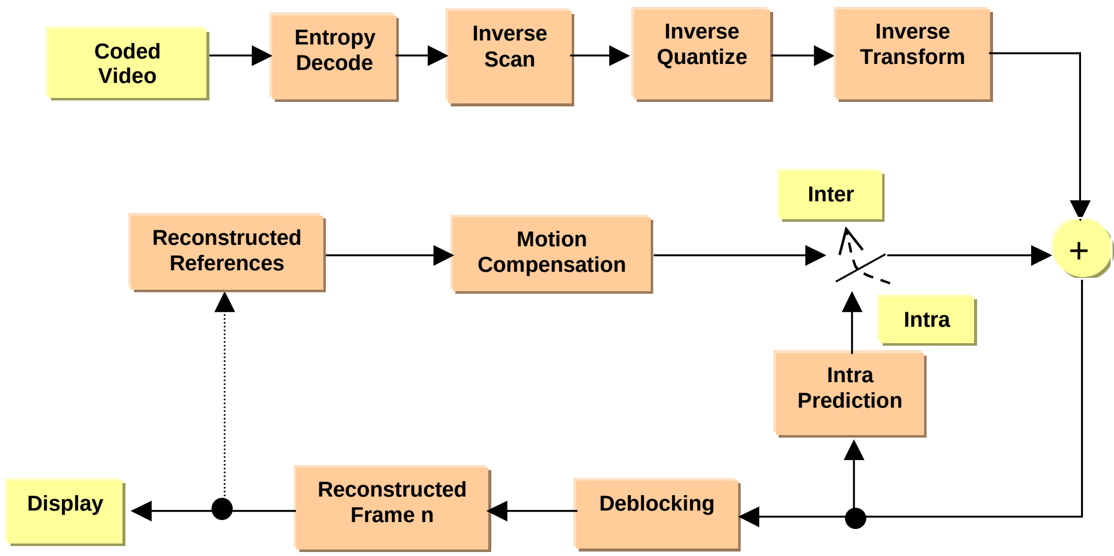

Figure 3 - Block diagram of AVC/JVC/H.264 Video Decoder

It should be noted that there is no single coding feature that provides the majority of the coding gain achieved by this new standard. Rather, it is the incremental improvements of the different tools that add up to and sometimes better than 50% of coding gain in relation to prior art.

Here are the main features of the H.264/AVC standard and the estimated coding gains where appropriate:

- **Intra prediction:** Removes spatial redundancy.

  Luminance blocks come in either 16x16 or 4x4 pixel block sizes. The 16x16 mode has 4 prediction modes, whereas the 4x4 mode has 9 prediction modes. The chrominance blocks are 8x8 pixel sizes and have 4 prediction modes. In total there are 17 prediction modes for an Intra coded picture.

- **Inter prediction:** Removes temporal redundancy between successive frames using variable block sizes.

  This type of prediction is based on motion estimation in the encoder and motion compensation in the decoder. The standard provides highly flexible block sizes and shapes ranging from 16x16 down to 4x4 pixels for luminance blocks, and from 8x8 down to 2x2 pixels for chrominance blocks. Up to 16 motion vectors may be transmitted for a single macroblock. *It is estimated that having different block sizes can achieve up to 15% of bit savings compared to using only a 16x16 block size*.

- **Quarter pixel motion estimation accuracy:** Improves motion smoothness and coding efficiency.

- A 6-tap horizontal and vertical filtering is used to derive the ½ pixel resolution samples. The quarter sample resolution value is derived by averaging the adjacent pixel values. *It is estimated that a quarter sample spatial accuracy can improve the coding efficiency by as much as 20% compared to integer-based accuracy specially at higher bit rates*.

- **Multiple reference frames:** Improves coding efficiency.

  Multiple previously decoded frames that are before or after the current frame in display order, may be used as reference for block motion estimation. *It is estimated that if 5 frames are used for reference instead of 1, up to 14% of coding efficiency can be achieved at high bit rate sequences*.

- **Integer 4x4 transform:** Eliminates decoder mismatches and improves visual quality.

  The 4x4 block transform is a rough approximation of the floating point DCT transform of the same size. The 4x4 transform involves only integer additions, subtractions, and shifting, in just 16-bit arithmetic. No rounding errors and no mismatches can occur between decoders as the ones encountered in a floating point DCT. All normalization factors are combined with the quantization step outside the transform. In some modes, the dc coefficients undergo an additional Hadamard transformation to decorrelate the blocks further. The smaller block size transform reduces artifacts such as ringing in the picture.

- **In-loop deblocking filter:** Improves visual quality and code efficiency.

  All horizontal and vertical edges in luma and chroma 4x4 blocks in a macroblock are adaptively filtered to smoothen 'blocking' distortion. The filtering is applied to the reconstructed picture, which in turn is used in temporal prediction. The filtering is not applied to the reconstructed macroblock used for spatial prediction in intra-coded macroblocks. The filtering involves deriving a strength factor and two thresholds that define the filter depth or disable it all together. This involves a complex decision making process. *It is estimated that up to 9% of bit savings can be achieved by using the deblocking filter versus not using it.*

- **Two types of entropy coding:** CAVLC and CABAC.

  The Context Adaptive Variable-Length Coding (CAVLC) is used together with Universal Variable-Length Coding (UVLC) to assign shorter codewords to symbols with higher probability of occurrence, and longer codewords to those with lower probability of occurrence. A single UVLC table lookup is used for all symbols except for transform coefficients. For transform coefficients, CAVLC is used to take advantage of the coefficient structure and properties of a transformed block. CAVLC uses multiple tables adaptively based on recently coded levels. The second entropy coding method is Context Adaptive Binary Arithmetic Coding (CABAC), which uses probability models for each syntax element based on local statistics, and then uses arithmetic coding to encode a data symbol. *It is estimated that the CABAC method can achieve up to 14% of bit savings over CAVLC.*

- **Weighted prediction:** Improves coding efficiency.

  This feature allows the motion compensated predictor signal to be weighted by a global multiplier and adjusted by an offset. This improves coding efficiently for scenes involving fades, lighting changes, and for other special effects as well.

- **Five slice types:** I, P, B, SI, and SP.

  Two new slice types are defined in the H.264/AVC standard; SI and SP. These types of slices are used for synchronization and switching. They replace the need to encode a high bit consuming I-slice by allowing a special prediction from a previously encoded picture. They are used in cases such as; switching between two similar video contents streamed at different bit rates, recovery from data losses or errors, and fast-forward or fast-reverse.

- **Redundant pictures:** Improves error resiliency.

  This feature allows the encoder to send redundant or duplicate slices of regions in the pictures, enabling recovery of data errors in networks prone to error.

- **Arbitrary slice ordering (ASO):** Reduces processing delay.

  Since slices are decoded independently of each other, this feature allows the decoder to process slices as they arrive. This reduces overall latency and improves the end-to-end delay in IP networks where packets might arrive out of order.

- **Scalar quantization:** No wide dead-zone and finer fidelity for the chrominance block.

  Significant compression occurs during the quantization process. Fifty-two levels are defined with no wide dead-zone at low levels. Level compression increases at a logarithmic rate where an increase of 6 in quantization parameter, increases the step size by a factor of 2. In addition, a finer step size is defined for the chroma blocks compared to prior art, improving the chrominance fidelity. *It is estimated that a change by 12% in step size (increment of 1 in quantization parameter) means a reduction of about 12% in bit rate.*

6 Summary

The high flexibility of the H.264/AVC standard with respect to the available coding tools as grouped by 'profiles', together with the parametric limits determined by the 'level' chosen, allows the implementation of a wide variety of video and multimedia applications and over many types of networks, at up to 50% bit-rate savings for a given quality relative to a comparable MPEG-2 or MPEG-4 implementation.