H.264 Baseline Profile Video Encoder

Johotech Solutions

INTRODUCTION

Video has always been the backbone of multimedia technology. Digital video is being adopted in an increasingly proliferating array of applications ranging from video telephony and videoconferencing to DVD and digital TV. The adoption of digital video in many applications has been fuelled by the development of video coding standards, and many video coding standards have emerged targeting different application areas. In the last two decades, the field of video coding has been revolutionized by the advent of various standards like MPEG-1 to MPEG-4 and H.261 to H.263++, each addressing different aspects of multimedia. H.264 is a new standard which adds one more step in the endeavor towards video coding excellence and provides one stop solution for wide range of applications. The standard is being developed by Joint Video Team (JVT) comprising of both ISO/IEC and ITU-T. The primary goal of H.264 is to achieve higher compression while preserving the video quality. The motivation for higher compression is to compensate for the ever present constraints of the limited channel capacity. This video coding technique follows a straight-forward “back to basics approach” with greater network friendliness, supporting sufficient error resilience for use in the unreliable networks and providing flexibility to be used in low-delay real-time applications.

Johotech Solutions (JTS) has been developing a complete video processing solution that is based on the H.264 video coding standard. The purpose of this paper is to present an overview of the H.264 standard and a discussion of JTS H.264 Baseline Profile Encoder. First, an overview of the H.264 standard and its benefits are presented. JTS H.264 Baseline Profile Encoder main features are then described , and its benefits for real-time video applications are finally discussed.

H.264: Technical Description

Overview

The main objective of the emerging H. 264 standard is to provide a means to achieve substantially higher video quality compared to what could be achieved using any of the existing video coding standards. Nonetheless, the underlying approach of H.264 is similar to that adopted in previous standards such as H.263 and MPEG-4, and consists of the following four main stages:

1. Dividing each video frame into blocks of pixels so that processing of the video frame can be conducted at the block level.

2. Exploiting the spatial redundancies that exist within the video frame by coding some of the original blocks through transform, quantization and entropy coding (or variable-length coding).

3. Exploiting the temporal dependencies that exist between blocks in successive frames, so that only changes between successive frames need to be encoded. This is accomplished by using motion estimation and compensation. For any given block, a search is performed in the previously coded one or more frames to determine the motion vectors that are then used by the encoder and the decoder to predict the subject block.

4. Exploiting any remaining spatial redundancies that exist within the video frame by coding the residual blocks, i.e., the difference between the original blocks and the corresponding predicted blocks, again through transform, quantization and entropy coding.

Schematic description of H.264 Encoder

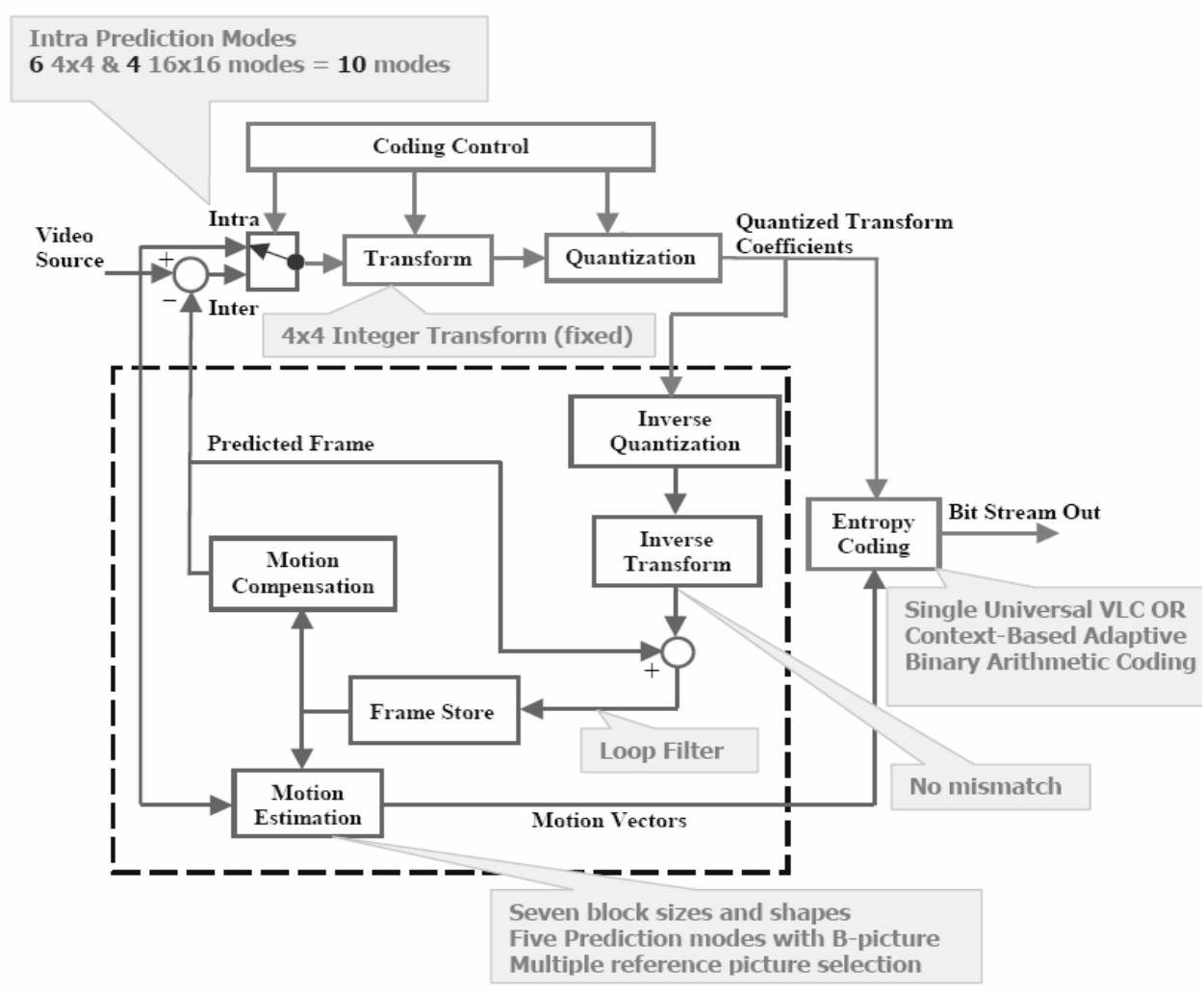


Figure 1: H.264 Video Encoder block diagram

Intra Prediction and Coding

Intra coding refers to the case where only spatial redundancies within a video picture are exploited. The resulting frame is referred to as an I-picture. I-pictures are typically encoded by directly applying the transform to the different macroblocks in the frame. As a consequence, encoded I-pictures are large in size since a large amount of information is usually present in the frame, and no temporal information is used as part of the encoding process. In order to increase the efficiency of the intra coding process in H.264, spatial correlation between adjacent macroblocks in a given frame is exploited. The idea is based on the observation that adjacent macroblocks tend to have similar properties. Therefore, as a first step in the encoding process for a given macroblock, one may predict the macroblock of interest from the surrounding macroblocks (typically the ones located on top and to the left of the macroblock of interest, since those macroblocks would have already been encoded). The difference between the actual macroblock and its prediction is then coded, which results in fewer bits to represent the macroblock of interest as compared to when applying the transform directly to the macroblock itself.

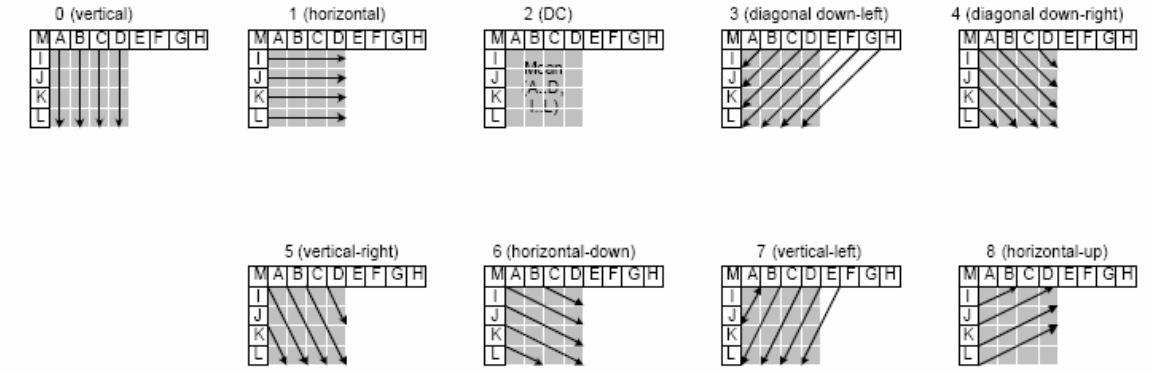


Figure 2 : Intra Prediction Modes for 4x4 luminance blocks

In order to perform the intra prediction mentioned above, H.264 offers nine modes for prediction of 4x4 luminance blocks, including DC prediction and eight directional modes. This process is illustrated in Figure 2, in which pixels A to L from neighbouring blocks have already been encoded and may be used for prediction.

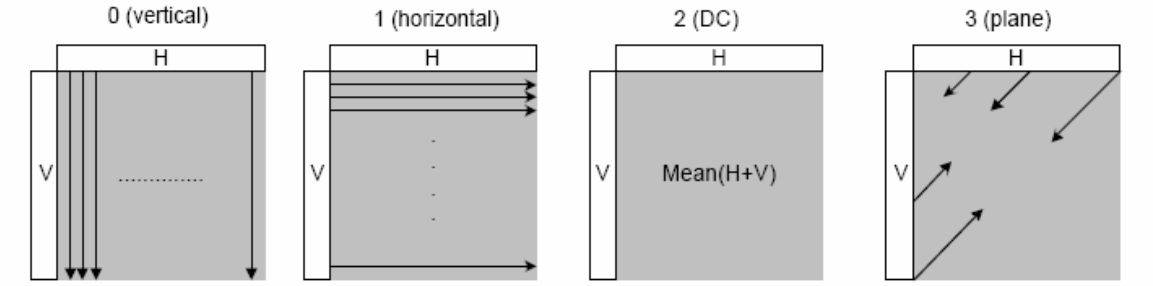


Figure 3 : Intra Prediction Modes for 16x16 luminance blocks

For regions with less spatial detail (i.e. flat regions), H.264 also supports 16x16 intra coding, in which one of four prediction modes is chosen for the prediction of the entire macroblock. Finally, the prediction mode for each block is efficiently coded by assigning shorter symbols to more likely modes, where the probability of each mode is determined based on the modes used for coding surrounding blocks.

Inter Prediction and coding

Inter prediction and coding is based on using motion estimation and compensation to take advantage of the temporal redundancies that exist between successive frames, hence, providing very efficient coding of video sequences. When a selected reference frame for motion estimation is a previously encoded frame, the frame to be encoded is referred to as a P-picture. When both a previously encoded frame and a future frame are chosen as reference frames, then the frame to be encoded is referred to as a B-picture. Motion estimation in H.264 supports most of the key features found in earlier video standards, but its efficiency is improved through added flexibility and functionality. In addition to supporting P-pictures (with single and multiple reference frames) and Bpictures, H.264 supports a new inter-stream transitional picture called an SP-picture. The inclusion of SPpictures in a bit stream enables efficient switching between bit streams with similar content encoded at different bit rates, as well as random access and fast playback modes.

The following four sections describe in more detail the four main motion estimation features used in H.264 namely,
(1) the use of various block sizes and shapes,
(2) the use of high-precision sub-pel motion vectors,

Tree structured motion compensation

H.264 supports motion compensation block sizes ranging from 16x16 to 4x4 luminance samples with many options between the two. The luminance component of each macroblock (16x16 samples) may be split up in 4 ways as shown in Figure 4-1: 16x16, 16x8, 8x16 or 8x8. Each of the sub-divided regions is a macroblock partition. If the 8x8 mode is chosen, each of the four 8x8 macroblock partitions within the macroblock may be split in a further 4 ways as shown in Figure 4-2: 8x8, 8x4,4x8 or 4x4 (known as macroblock sub-partitions). These partitions and sub-partitions give rise to a large number of possible combinations within each macroblock. This method of partitioning macroblocks into motion compensated sub-blocks of varying size is known as **tree structured motion compensation**.

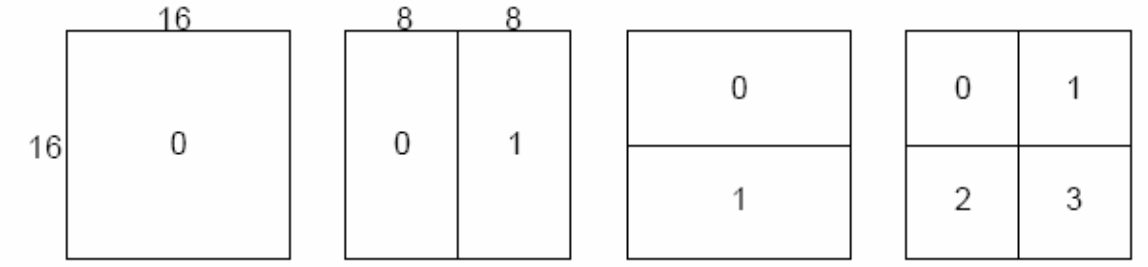


Figure 4-1 : Macroblock Partitions 16x16,8x16,16x8,8x8

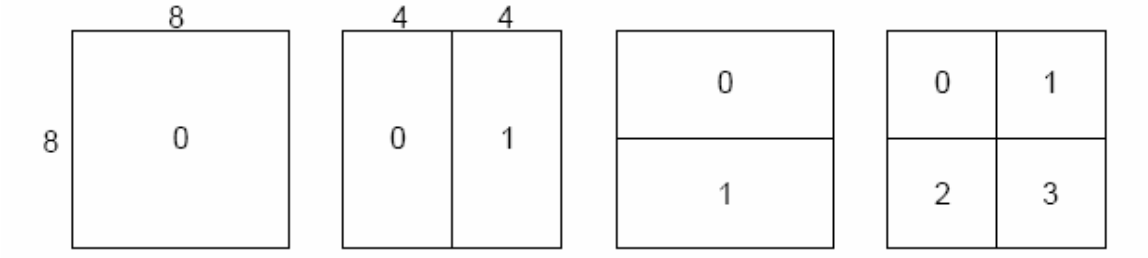


Figure 4-2 : Macroblock sub – partitions 8x8,4x8,8x4,4x4

A separate motion vector is required for each partition or sub-partition. The resolution of each chroma component in a macroblock (Cr and Cb) is half that of the luminance (luma) component. Each chroma block is partitioned in the same way as the luma component, except that the partition sizes have exactly half the horizontal and vertical resolution The horizontal and vertical components of each motion vector (one per partition) are halved when applied to the chroma blocks.

High-precision sub-pel motion vectors

The prediction capability of the motion compensation algorithm in H.264 is further improved by allowing motion vectors to be determined with higher levels of spatial accuracy than in existing standards. Quarterpixel accurate motion compensation is currently the lowest-accuracy form of motion compensation in H.26L (in contrast with prior standards based primarily on half-pel accuracy, with quarter-pel accuracy only available elsewhere in the newest versions of MPEG-4).

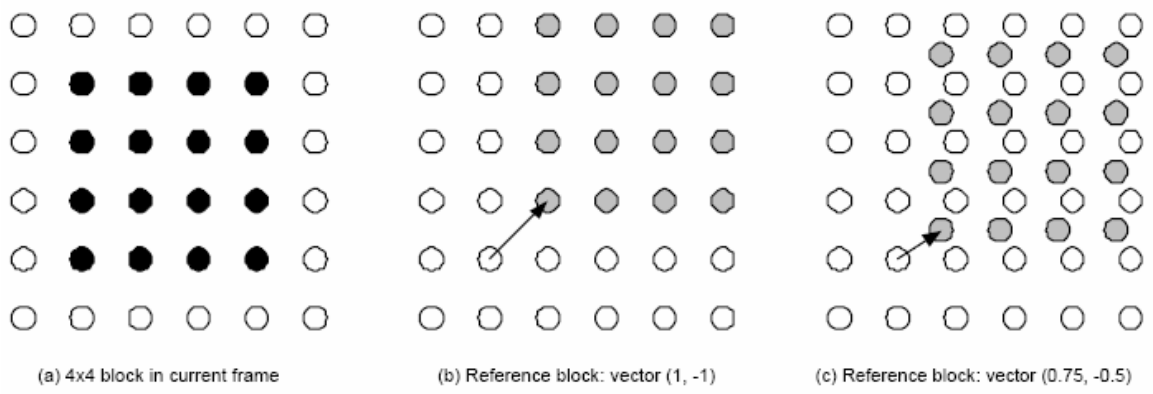


Figure 5 : Example of Integer and Sub Pixel Prediction

The luma and chroma samples at sub-pixel positions do not exist in the reference picture and so it is necessary to create them using interpolation from nearby image samples. Figure 5 gives an example. A 4x4 sub-partition in the current frame (a) is to be predicted from a neighbouring region of the reference picture. If the horizontal and vertical components of the motion vector are integers (b), the relevant samples in the reference block actually exist (grey dots). If one or both vector components are fractional values (c), the prediction samples (grey dots) are generated by interpolation between adjacent samples in the reference frame (white dots).

A predicted vector, MVp, is formed based on previously calculated motion vectors. MVD, the difference between the current vector and the predicted vector, is encoded and transmitted.

Integer Transform

Each residual macroblock is transformed, quantized and coded. Previous standards such as MPEG-1, MPEG-2, MPEG-4 and H.263 made use of the 8x8 Discrete Cosine Transform (DCT) as the basic transform. The “baseline” profile of H.264 uses three transforms depending on the type of residual data that is to be coded: a transform for the 4x4 array of luma DC coefficients in intra macroblocks (predicted in 16x16 mode), a transform for the 2x2 array of chroma DC coefficients (in any macroblock) and a transform for all other 4x4 blocks in the residual data.

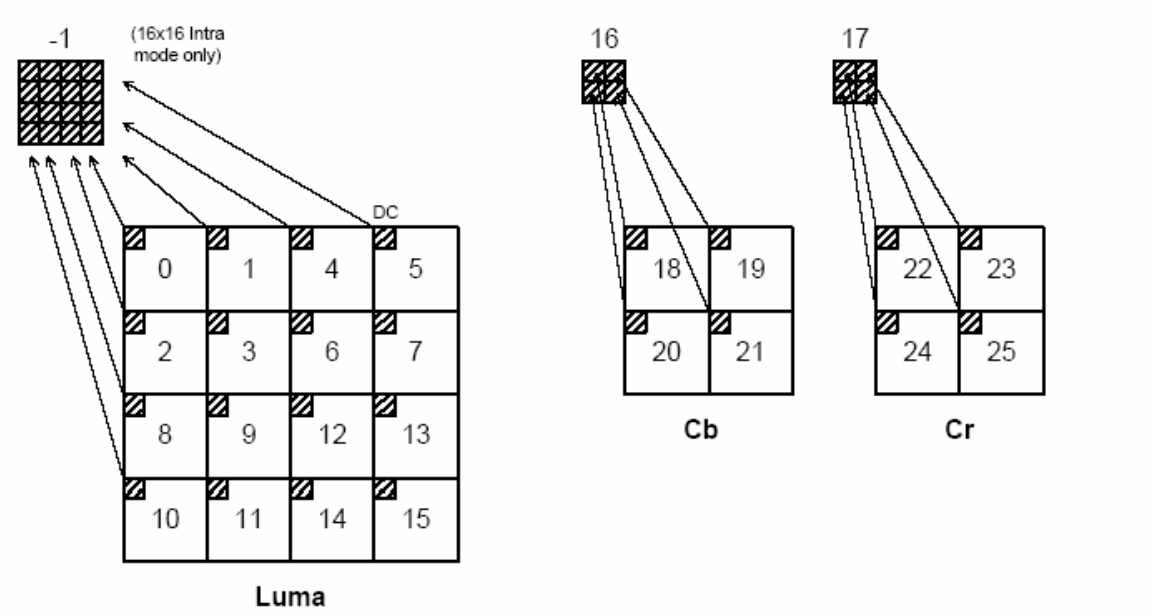


Figure 6 : Scanning order of residual blocks within a Macroblock

Data within a macroblock are transmitted in the order shown in Figure 6. If the macroblock is coded in 16x16 Intra mode, then the block labelled "-1" is transmitted first, containing the DC coefficient of each 4x4 luma block. Next, the luma residual blocks 0-15 are transmitted in the order shown (with the DC coefficient set to zero in a 16x16 Intra macroblock). Blocks 16 and 17 contain a 2x2 array of DC coefficients from the Cb and Cr chroma components respectively. Finally, chroma residual blocks 18-25 (with zero DC coefficients) are sent.

Quantization

The quantization step is where a significant portion of data compression takes place. In H.264, the transform coefficients are quantized using scalar quantization with no widened dead-zone. A total of 52 values of Qstep are supported by the standard and these are indexed by a Quantization Parameter, QP. Qstep doubles in size for every increment of 6 in QP; Qstep increases by 12.5% for each increment of 1 in QP. The wide range of quantizer step sizes makes it possible for an encoder to accurately and flexibly control the trade-off between bit rate and quality. The values of QP may be different for luma and chroma; both parameters are in the range 0-51 but QPChroma is derived from QPY so that it QPC is less that QPY for values of QPY above 30.

Entropy Coding

H.264 has adopted two approaches for entropy coding. The first approach is based on the use of Context Adaptive Variable Length Codes (UVLCs) and the second is based on Context-Based Adaptive Binary Arithmetic Coding (CABAC).

Deblock Filter

Strong motion isolation, enhanced intra prediction and arbitrary block sizes introduce visually distinguishable blocks. H.264 resolves these prominent blocking artifacts by applying a deblocking filter at the block boundaries. Use of in-loop de-blocking filter reduces the blocking artifacts, hence improving the overall picture quality.

H.264 : Applications

H.264 video codec aims at providing a single solution for a wide range of applications. The standard can be used in variety of systems viz. real time or non-real time, conversational or nonconversational, streaming or packet based and wired or wireless. Some of the target applications of H.264 are -

- Audio-visual communication over wireless networks
- Real-time conversational services, for instance Video phone.
- Internet video applications viz. streaming video and video conferencing.
- Video storage and retrieval services like Vide-on-Demand (VoD).
- Video storage and forward services, for instance video mail
- Digital video broadcast over DSL/ ADSL, DBS and CATV
- Digital terrestrial television broadcast
- Remote video surveillance

The above applications require support of varying bit rates, resolutions and quality of services. Consequently, the entire bitstream syntax is divided in terms of Profiles and Levels. Profiles are the basic subsets of the entire bitstream syntax. This division is based on algorithmic tools and features used in the codec without affecting the quality of the output video. Each profile is divided into various levels imposing constraints on maximum picture size, macroblocks per second, bitrate, and similar parameters. Hence, a very large variation in the performance of the codec can be achieved within a specific profile. Currently, the H.264 standard indicates the presence of 3 profiles namely the Baseline, Main and the Extended Profile, each of them having 14 levels.

Interaction of H.264 with various networks and applications

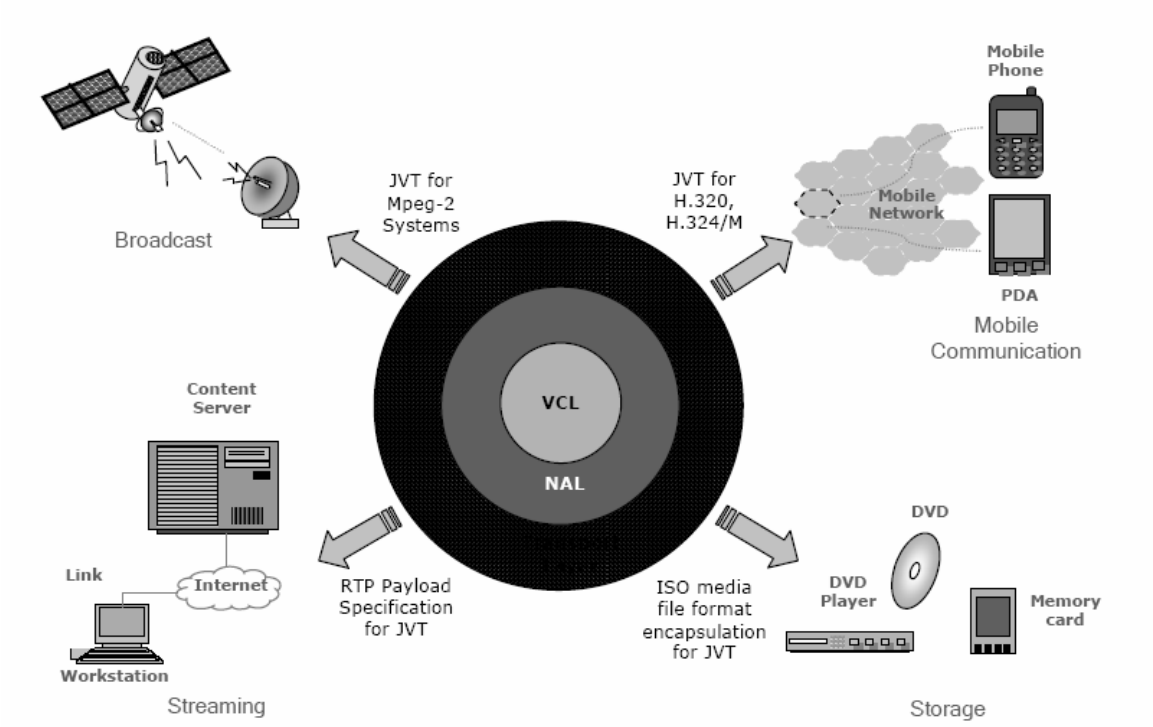


Figure 7 : Interaction of H.264 with various networks and applications

H.264 is divided into two distinct layers viz. Network Abstraction Layer (NAL) and Video Coding Layer (VCL). NAL is responsible for packaging the coded data in an appropriate manner based on the characteristics of the network upon which the data will be used. On the other hand, VCL is responsible for generating an efficient representation of the video data. Hence, NAL takes care of the constraints of the underlying network and gives VCL a network independent interface. H.264 supports both non-IP and IP based (fixed and wireless) transport mechanism viz. H.320, H.324, RTP and MPEG-2 transport stream. Figure 7 explains the interaction of H.264 with various networks and applications.

H.264 : Comparative Study

H.264 provides unmatched compression without any perceivable loss in quality. It has achieved almost 50% reduction in the bitrate as compared to its predecessors. It means that by using H.264, broadcasters can telecast more number of channels over the existing links, mobile service providers can transmit better quality video with the same bitrate and publishers can store more number of movies on the DVDs. All this has been done by conglomerating the best techniques available in the field of video compression.

H.264 follows a generally simple and straightforward design using well-known building blocks. This makes the implementation and testing of the codec both faster and easier. Though techniques like in loop de-blocking filter and quarter pixel motion accuracy have

increased the computational complexity of the codec, it can be neutralized by the tremendous growth in the computing power of processors which has happened in the recent years.

Johotech's H.264 Baseline Profile Video Encoder

Johotech Solution has implemented Baseline Profile video encoder in ANSI C for H.264. The encoder accepts input data from disk files in YUV 4:2:0 format and outputs the coded bitstream according to a set of configuration parameters . The encoder basically operates in two modes viz Constant Quality and Constant Bitrate. In Constant Quality mode, the video quality is maintained by keeping the quantization parameter constant. In Constant Bitrate encoding , the quantization parameter is varied depending on the target bitrate , bits consumed , input picture complexity and buffer fullness , thereby keeping the bitrate constant . The main building blocks of the encoder are Motion Estimation , Bitrate Control , Integer transform , Quantization , Variable Length Coding , and bit stream syntax coding .

The encoder is fully configurable and useful in portable and wireless applications.

Configuration Parameters for the encoder

Some of the configuration parameters of Johotech's H.264 Encoder include:

- Profile
- Target Bitrate
- Frame rate
- Search Range
- Type of Motion Estimation (Full ME or Fast ME)
- Type of Sub Pel Motion Estimation (Half pel or Quarter Pel)
- Quantization Parameter
- Number of reference frames
- Number of frames to be encoded
- Input YUV
- Dynamic detailed nomenclature of output .264 stream and parameter file
- Intra Period which specifies the distance between two I frames
- Max Picture Order Count

Features of Johotech's H.264 Video Encoder

Salient features of Johotech's H.264 Video encoder include:

- Completely original code . No re-use from ISO reference code. No IP issues.
- Supports Baseline profile.
- Bits per pixel = 8.
- Fully configurable using an input configuration file.
- Supports wide range of YUVs ranging from QCIF (176x144) to D1 (720x480).
- Slice types – All supported by Baseline Profile (I and P Slice)
- Half Pel and Quarter Pel Motion Estimation.
- Optimized implementation of Deblocking Filter.

- Search range of 16 pixels.
- Highly efficient and less complex algorithm implemented for Motion Estimation.
- Very powerful algorithm implemented for Intra Macroblock selection process in P Slice. This improves quality (measured by SNR) of high motion picture by 10%.
- Efficient algorithm implemented for Mode Selection in P Slice. Computing is reduced by 20% compared to trivial least SAD based selection.
- Supports Constant Bitrate control.
- Supports Constant Quality control by disabling Rate Control.
- Rigorously tested. All streams were passed by reference decoder.
- No Floating Point Calculation.
- Highly optimized C code.
- Ported the optimized C code on TI DM642 DSP and achieved 11fps VGA for processor speed of 720MHz.
- Highly flexible with capability of porting on any of available DSP chips.

Critical Analysis and Comparative Study of Johotech's H.264 Video Encoder

Evaluation of performance of any Video Encoder is with respect to Reference Encoder. Many test cases were generated to critically evaluate the performance of Johotech's Video Encoder compared to the Joint Model (JM) Reference H.264 Encoder.

Test case to evaluate H.264 standard compared to MPEG-4 Simple profile.

Test case was generated by encoding a YUV at constant bitrate and comparing the SNR's of encoded streams.

YUV Stream : f1_vga Size : VGA (640x480)						
Bitrate (bps)	SNRY (in dB)		SNRU (in dB)		SNRV (in dB)	
	MPEG4	H.264	MPEG4	H.264	MPEG4	H.264
50353112	41.27	56.29	41.95	56.68	41.81	56.68
28122024	35.05	49.46	36.22	50.33	36.01	50.42
10252648	27.12	41.6	28.19	45.62	28.21	46.69
3202536	24.69	34.59	24.7	41.33	24.83	42.53

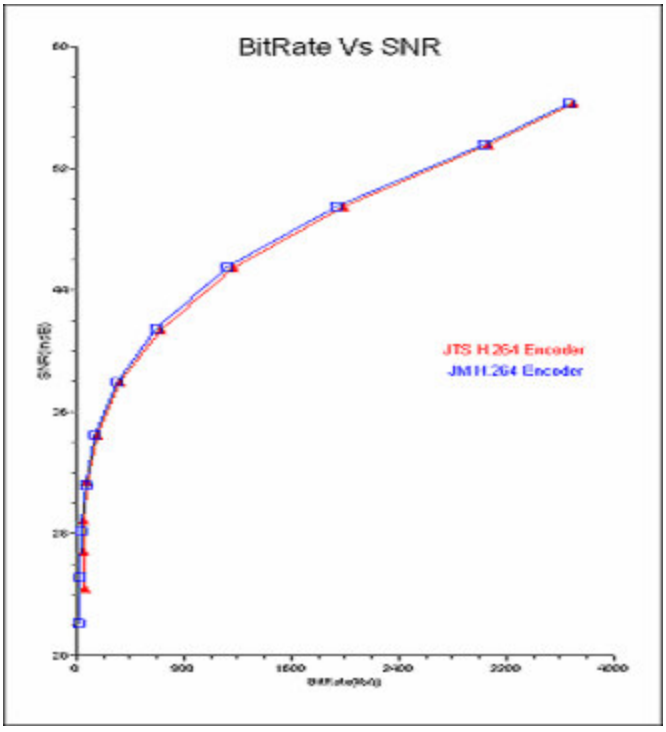
From above figures it can be concluded that H.264 outperforms MPEG 4 SP as choice of good "quality" video encoder. Though H.264 is comparable more complex than MPEG 4 SP.

Test Case to evaluate complexity of Motion Estimation module:
Complexity of Motion Estimation module can be inferred from the number of pixels (search points) traversed to conclude the best matching position (least SAD). Quality of the output stream can be inferred from SNR of Luma.

YUV / No. of frames	TYPE	QP	No. of search points		percent reduction in search points	SNR Y (in dB)	
			JM Fast ME	JTS ME algorithm		JM Fast ME	JTS ME algorithm
Forman /300	CIF	10	11896054	9750864	22	49.53	49.55
Akiyo / 300	CIF	20	10408815	8746904	19	45.1	44.64
Mobile / 300	CIF	10	10732547	9095379	18	49.37	49.36
Coastguard /300	CIF	20	11943406	9952839	20	40.81	40.79

From the above table it can be inferred that for a reduction in complexity of approximately 20 % , there is negligible reduction in quality .

Test case for Qualitative analysis of Johotech's H.264 Encoder.
264 streams were generated by both JM's as well as Johotech's H.264 Encoder for same configuration file. A plot of bitrate vs SNR is good for qualitative analysis of a video encoder.



From the above graph it can be seen that qualitatively Johotech's H.264 Encoder is at par with that of JM's (which uses plenary amount of resources compared to Johotech's Encoder) H.264 Encoder.

References

[1] Ralf Schafer, Thomas Weigand and Heiko Schwarz, "The Emerging H.264/AVC Standard",January 2003.