咨询热线：400-080-9161

您好，欢迎来到61ic！ | [登录] [注册] 忘记密码 | 设为首页 帮助

61ic®电子在线.com

. 网站首页 . 业界新闻 . 设计中心 . 移动技术 . TI专栏 . ADI专栏 . FPGA专栏 . 代码工厂 . 官方商城 .
. 活动专区 . 新品快递 . 解决方案 . 前沿科技 . TI资源 . ADI资源 . FPGA资源 . 下载中心 . 产品展示 .

» 加入收藏
» 付款方式
» 联系我们

TI Sitara ARM . TI Stellaris ARM . TI DaVinci . TI C6000 DSP和ARM . TI C6000 . TI C5000 . TI C2000 . TI MSP430 . TI模拟设计 . TI CCS .

您现在的位置： 61IC电子在线 >> TI专栏 >> TI DaVinci >> TMS320DM644x SOC >> 正文

[组图]Video codecs tutorial: Trade-offs with H.264, VC-1 and other advanced codecs 热 ★★★ 【字体：小 大】

Video codecs tutorial: Trade-offs with H.264, VC-1 and other advanced codecs

作者：德州仪器... 文章来源：德州仪器 (TI) 点击数：1080 更新时间：2011-4-24 收藏此页

In recent years, the demand for digital video products has witnessed a boom. Some examples of popular applications are video communication, security and surveillance, industrial automation, and the biggest of all, entertainment, which includes DVD, highdefintion (HD) TV, satellite TV, HD set-top boxes, Internet video streaming, digital cameras and HD video camcorders, video jukeboxes, high end displays (LCD, Plasma and DLP) and personal video recorders. A slew of new and exciting applications are currently in design or early deployment. For example, HD-DVD (Blu-ray), digital video broadcast, both to the home and the handset through terrestrial or satellite (DVB-T, DVB-H, DMB), HD videophones, digital cinema and IP Set-top boxes. End products are also increasingly becoming mobile and converged as a result of higher computational power in handsets, advances in battery technology and high-speed wireless connectivity.

Video compression is an essential enabler for all these exciting new video products. Compression-decompression (codec) algorithms make it possible to store and transmit digital video. Typically, codecs are either industry standards such as MPEG-2, MPEG-4, H.264/AVC and AVS or proprietary algorithms, such as On2, Real Video, Nancy and Windows Media Video (WMV). WMV is an exception as it was originally a Microsoft proprietary algorithm that is now also standardized by SMPTE as VC-1. Codec technology has continuously improved in the last decade. The most recent codecs, H.264/AVC and VC-1, represent the third generation of video compression technology. Both codecs are capable of squeezing very high compression ratios utilizing the available processing horsepower in low-cost ICs such as programmable DSPs and fixed-function ASICs. However choosing the right codec and optimizing its real-time implementation for a specific application remains a tough challenge. The optimal design must trade-off between compression efficiency and the use of available computational horse-power. Obtaining the optimal compression efficiency with limited computational horse-power is a tough science.

In this paper, we first provide an overview of key concepts in video coding and describe the legacy compression standards. Next, we focus on the capabilities of the latest generation of codecs including H.264/AVC, WMV9/VC-1 and AVS and provide insights into compression and complexity trade-offs that each offers. Finally, we discuss real-time implementations and key trends in end-equipment segments that may influence choices between the popular video codecs.

The Video Compression Challenge
A major challenge for digital video is that raw or uncompressed video requires lots of data to be stored or transmitted. For example, standard definition NTSC video is typically digitized at 720x480 using 4:2:2 YCrCb at 30 frames per second, which requires a data rate of over 165 Mbps. Storing one 90-minute video requires over 110 GBytes or more than 25x the storage capability of a standard DVD-R. Even lower resolution video such as CIF (352x288 4:2:0 at 30 frames/second), which is often used in video streaming applications, requires over 36.5 Mbits/s. This is many times more than what can be sustained on broadband networks such as ADSL or 3G wireless. Today, broadband networks offer between 1-10 Mbps of sustained throughput. Clearly, compression is needed to store or transmit digital video.

The main goal for video compression is to encode digital video using as few bits as possible while maintaining visual quality. Codecs are based on the mathematical principles of information theory. However, building practical codec implementations requires making delicate trade-offs that approach being an art form.

Compression Tradeoffs
There are many factors to consider when selecting the codec in a digital video system. The most important ones are the visual quality requirements for the application, the environment (speed, latency and error characteristics) of the transmission channel or storage media and the format of the source content. Also highly important are the desired resolution, target bitrate, color depth, the number of frames per second and whether the content and/or display are progressive or interlaced.

Compression often involves trade-offs between the visual quality requirements and other needs of the application. Firstly, is the application storage, uni-cast, multi-cast, two-way, or broadcast? For storage applications, how much storage capacity is available and what is the recording duration? For non-storage applications, what is the maximum bit rate? For two-way video communication, what is the latency tolerance or allowable end-to-end system delay? If not two-way, is the content that must be encoded available in advance off-line or does it require to be encoded in real-time? How error-prone is the network or storage media? The various compression standards handle these trade-offs differently depending on the primary target application.

Another trade-off is the cost of real-time implementation of the encoding and decoding. Typically newer algorithms such as H.264/AVC or WMV9/VC-1 that achieve higher compression require increased processing, which can impact the cost for encoding and decoding devices, system power dissipation and system memory.

Standards Bodies
There have been two primary standards organizations driving the definition of video codecs. The International Telecommunications Union (ITU) is focused on telecommunication applications and has created the H.26x standards for low bitrate video telephony. These include H.261, H.262, H.263 and H.264. The International Standards Organization (ISO) is more focused on consumer applications and has defined the MPEG standards for compressing moving pictures. MPEG standards include MPEG-1, MPEG-2 and MPEG-4. Figure 1 illustrates the history of video codecs standardization.

MPEG and ISO often make slightly different tradeoffs based on its primary target applications. On occasions, the groups have worked together such as in the Joint Video Team (JVT) to define the H.264 codec also known as MPEG-4 Part 10 or MPEG-4 Advanced Video Coding (AVC) in the MPEG family. In this paper, we refer to this joint standard as H.264/AVC. Similarly H.262 and MPEG-2 are identical, while H.263 Baseline Profile technology has a significant overlap in techniques with the MPEG-4 Part 2 Simple Profile codec.

Standards have been critical for the widespread adoption of codec technology. Consumers find products based on standards affordable because of economies of scale. The industry is willing to invest on standards given their assurance of interoperability between vendors. Content providers are attracted to standards given the long life and broad demand their content would see. While almost all video standards are targeted for a few specific applications, they are often used to advantage in other applications when they are well suited.

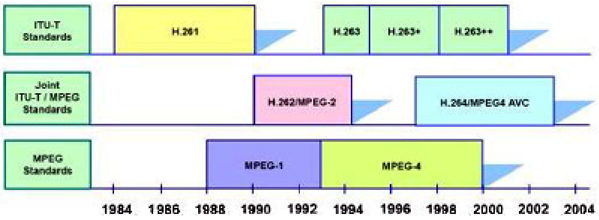


Figure 1: Chronological progression of ITU and MPEG standard [10]

ITU and MPEG continue to evolve compression techniques and define new standards for better compression and newer market opportunities. China has recently defined a national video coding standard called AVS, which we also describe later in this paper. Standards currently in the works include ITU/MPEG Joint Scalable Video Coding, an amendment to H.264/ AVC, and MPEG Multi-view Video Coding. Meanwhile, existing standards are continually evolving to satisfy newer applications. For example, H.264 has recently defined a new mode called Fidelity Range Extension to address upcoming markets such as professional digital editing, HD-DVD and lossless coding.

In addition to industry standards from the ITU and ISO, several popular proprietary solutions have emerged particularly for Internet streaming media applications. These include Real Networks Real Video (RV10), Microsoft Windows Media Video 9 (WMV9) Series, ON2 VP6, and Nancy. Due to the installed base of content in these formats, proprietary codecs can become de facto standards. In September 2003, Microsoft proposed to the Society for Motion Picture and Television Engineers (SMPTE) that the WMV9 bitstream and syntax be standardized under the aegis of that organization. This proposal was accepted and WMV9 is now standardized in SMPTE as VC-1.

Video Coding Principles
Our most preferred video standards use block-based processing. Each macro-block typically contains four 8x8 luminance blocks and two 8x8 chrominance blocks (for Chroma format 4:2:0). Video coding is based on the principles of motion compensated prediction (MC) [1], transform and quantization and entropy coding. Figure 2 shows a typical motion compensation based video codec. In motion compensation, compression is achieved by predicting each Macro-block of pixels in a frame of video from a similar region of a recently coded (“reference”) video frame. For example, background areas often stay the same from one frame to the next and do not need to be retransmitted in each frame. Motion estimation (ME) is the process of determining for each MB in the current frame, the 16x16 region of the reference frame that is most similar to it. ME is usually the most performance intensive function in video compression. Information on the relative location of the most similar region for each block in the current frame (“motion vector”) is transmitted to the decoder.

The residual after MC is divided into 8x8 blocks, each encoded using a combination of transform coding, quantization and variable length coding. Transform coding, such as discrete cosine transform or DCT, exploits spatial redundancy in the residual signal. Quantization removes perceptual redundancy and reduces the amount of data required to encode the residual. Variable length coding exploits the statistical nature of the residual coefficients. The process of redundancy removal via MC is reversed in the decoder, and the predicted data from the reference frame is combined with the encoded residual data to generate back a representation of the original video frame.

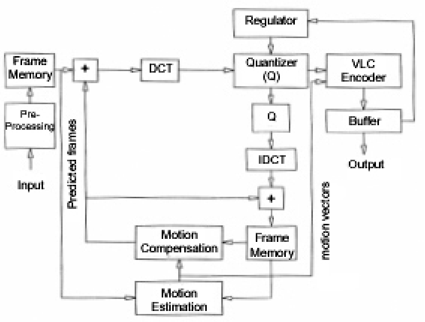


Figure 2: Standard Motion Compensated Video Coding

In a video codec, an individual frame may be encoded using one of three modes: I, P or B (see Figure 3). A few frames referred to as Intra (I) frames are encoded independently without reference to any other frame (no motion compensation). Some frames may be coded using MC with a previous frame as reference (forward prediction). These frames are referred to as Predicted (P) frames.

B frames or bi-directional predicted frames are predicted from both past frames as well as frames slated to appear after the current frame. A benefit of B frames is the ability to match a background area that was occluded in the previous frame but can be found in a subsequent frame using backward prediction. Bi-directional prediction can allow for decreased noise by averaging both forward and backward prediction. Leveraging this feature in encoders requires additional processing since ME has to be performed for both forward and backward prediction, which can effectively double the motion estimation computational requirements. Additional memory is also needed at both encoder and decoder to store two reference frames. B frame tools require a more complex data flow since frames are decoded out of order with respect to how they are captured and need to be displayed. This feature results in increased latency and thus, is not suitable for some real-time sensitive applications. Until H.264, B frames were not used for prediction allowing trade-offs to be made for some applications. For example, they can be skipped in low frame rate apps without impacting the decoding of future I and P frames.

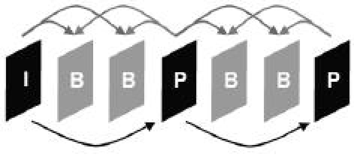


Figure 3: An illustration of inter-frame prediction in I, P and B Frames

Legacy Video Coding Standards

H.261

H.261[2], defined by the ITU, was the first major video compression standard. It was targeted for two-way video conferencing applications and was designed for ISDN networks that supported 40kbps-2Mbps. H.261 supports resolutions of 352X288 (CIF) and 176X144 (QCIF) with the chrominance resolution sub-sampling of 4:2:0. Complexity is also designed to be low since videophones require simultaneous real-time encoding and decoding. Due to its focus on two-way video, which is delay-sensitive, H.261 allows only I and P frames and no B frames.

H.261 uses a block-based DCT for transform coding of the residual. The DCT maps each 8x8 block of pixels to the frequency domain producing 64 frequency components (first coefficient referred to as DC and the rest as AC). To quantize the DCT coefficients, H.261 uses a fixed linear quantization across all the AC coefficients. The quantized coefficients are subject to run-length coding, which allows the representation of the quantized frequency coefficients as a non-zero coefficient level followed by runs of zero coefficients and a final end of block code after the last non-zero value. Finally, variable Length (Huffman) Coding converts the run-level pairs into variable length codes (VLCs) with the bit-length optimized for the typical probability distribution.

Standard block based coding results in blocky video. In H.261, this is avoided by using a loop filtering technique. A simple 2D FIR filter applied on the block edge is used to smooth out quantization effects in the reference frame. It must be applied in a bit-exact fashion on both the encoder and the decoder.

MPEG-1

MPEG-1[3] was the first video compression algorithm developed by the ISO. The driving application was storage and retrieval of moving pictures and audio on digital media such as video CDs using SIF resolution (352x240 at 29.97 fps or 352X288 at 25 fps) at about 1.15 Mbps. MPEG-1 is similar to H.261 but encoders typically require more performance to support the heavier motion found in movie content versus typical video telephony.

Compared to H.261, MPEG-1 allows B frames. It also uses adaptive perceptual quantization. For example, a separate quantization scale factor or equivalently step size is specifically applied to each frequency bin to optimize human visual perception. MPEG-1 only supports progressive video, and resultantly, an effort was started on a new standard, MPEG-2, to support both progressive and interlaced video at higher resolutions using higher bitrates.

MPEG-2/H.262

MPEG-2[4] was developed targeting digital television and soon became the most successful video compression standard thus far. MPEG-2 addressed both standard progressive video (where a video sequence consists of a succession of frames each captured at regularly spaced time instants) as well as interlaced video, which is popular in the television world. In interlaced video, two sets of alternate rows of pixels (each called a field) in the image are captured and displayed alternately. Until recently, this approach was particularly suited to the physics of most TV displays on the market. MPEG-2 supports standard television resolutions including interlaced 720x480 at 60 fields per second for NTSC used in the US and Japan and interlaced 720x576 at 50 fields per second for PAL used in Europe and other countries.

MPEG-2 builds on MPEG-1 with extensions to support interlaced video and also much wider motion compensation ranges. Since higher resolution video is an important application, MPEG-2 supports vastly wider search ranges than MPEG-1. This greatly increases the performance requirement for motion estimation versus the earlier standards. Encoders taking full advantage of the wider search range and the higher resolution, require significantly more processing than H.261 and MPEG-1. Interlaced coding tools in MPEG-2 include the ability to optimize the motion compensation supporting both field and frame based predictions and support for both field and frame based DCT/IDCT. MPEG-2 performs well at compression ratios around 30:1. The quality achieved with MPEG-2 at 4-8 Mbps was acceptable for consumer video applications, and it soon became deployed in applications including digital satellite, digital cable, DVDs and lately, high-definition TV.

In addition, MPEG-2 adds scalable video coding tools to support multiple layer video coding, namely, temporal scalability, spatial scalability, SNR scalability and data partitioning. Although related profiles were defined in MPEG-2 for scalable video applications, Main Profile that supports single layer coding is the sole MPEG-2 profile that is widely deployed in mass market today. MPEG-2 Main Profile is often referred to as simply MPEG-2. The processing requirements for MPEG-2 decoding were initially very high for general purpose processors and even DSPs. Optimized fixed function MPEG-2 decoders were developed and became inexpensive over time due to the high volumes. MPEG-2 proved that the availability of cost-effective silicon solutions is a key ingredient for the success and deployment of video codec standards.

H.263

H.263[5] was developed after H.261 with a focus on enabling better quality at even lower bitrates. One of the important targets was video over ordinary telephone modems at 28.8 Kbps. The target resolution was SQCIF (128x96) to CIF (352X288). The basic techniques are similar to H.261 with a few differences.

Motion vectors in H.263 were allowed to be multiples of ½ in either direction ("half-pel") with the reference picture digitally interpolated to higher resolution. This approach leads to better MC accuracy and higher compression ratios. Larger ranges were allowed for the MVs. A host of new options were provided for different scenarios including:

- Four motion vectors: One motion vector for each 8x8 block rather than one motion vector for the entire MB.
- 3D VLC: Huffman coding which combines an end of block (EOB) indicator together with each Run Level pair. This feature is specifically targeted at low-bit rate where many times there are only one or two coded coefficients.

However despite these techniques, adequate video quality over ordinary phone lines proved to be very difficult and videophones over standard modems are still a challenge today. Since H.263 generally offered improved efficiency over H.261, it became used as the preferred algorithm for video conferencing with H.261 support still required for compatibility with older systems. H.263 expanded over time as H.263+ and H.263++ added optional annexes supporting compression improvements and features for robustness over packet networks. H.263 and its annexes formed the core for many of the coding tools in MPEG-4.

MPEG-4

MPEG-4[6] was initiated by the ISO as a follow-on to the success of MPEG-2. Some of the early objectives were increased error robustness to support wireless networks, better support for low bitrate applications and a variety of new tools to support merging graphic objects with video. Most of the graphics features have not yet gained significant traction in products, and implementations have focused primarily on the improved low bitrate compression and error resiliency.

MPEG-4 simple profile (SP) starts from H.263 baseline and adds new tools for improved compression including:

- Unrestricted Motion Vectors: Supports prediction for objects when they partially move outside of the boundaries of the frame

- Variable Block Size Motion Compensation: Allows motion compensation at either 16x16 or 8x8 block granularity.
- Context Adaptive Intra DCT DC/AC Prediction: Allows the DC/AC DCT coefficients to be predicted from neighboring blocks either to the left or above the current block
- Extended dynamic range of quantized AC coefficients from [-127:127] in H.263 to [-2047, 2047] to support high fidelity video.

Error resiliency features added to support recovery in packet loss conditions include:

- Slice Resynchronization: Establishes slices within images that allow quicker resynchronization after an error has occurred. Unlike MPEG-2 packet sizes, MPEG-4 packet sizes are de-linked from the number of bits used to represent a MB. As a result, resynchronization is possible at equal intervals in the bitstream irrespective of the amount of information per MB.
- Data Partitioning: A mode that allows partitioning the data within a video packet into a motion part and DCT data part by separating these with a unique motion boundary marker. This allows more stringent checks on the validity of motion vector data. If an error occurs you can have better visibility into the point where the error occurs, thus avoiding the discarding of all the motion data when an error is found.
- Reversible VLC: VLC code tables designed to allow decoding backwards as well as forwards. When an error is encountered, it is possible to sync at the next slice or start code and work back to the point where the error occurred.
- New prediction (NEWPRED): Mainly designed for fast error recovery in real-time applications, where the decoder uses a reverse channel to request additional information from the encoder in the event of packet losses.

The MPEG-4 advanced simple profile (ASP) starts from the simple profile and adds B frames and interlaced tools (for Level 4 and up) similar to MPEG-2. It also adds quarter-pixel motion compensation and an option for global motion compensation. MPEG-4 advanced simple profile requires significantly more processing performance than the simple profile and has higher complexity and coding efficiency than MPEG-2.

MPEG-4 was used initially in Internet streaming and became adopted, for example, by Apple's QuickTime player. MPEG-4 simple profile is now finding widespread applications in mobile streaming. MPEG-4 ASP forms the foundation for the proprietary DivX codec that has become popular.

Tools vs. compression gains
Clearly, when we review the techniques introduced in the video codec field through H.261, MPEG-1, MPEG-2 and H.263, we observe that a few basic techniques have provided the most compression gains. Figure 4 illustrates those techniques and its relative effectiveness. Clearly motion compensation (both integer and half-pel) stands out compared to tools such as four motion vectors and quarter-pel motion compensation.

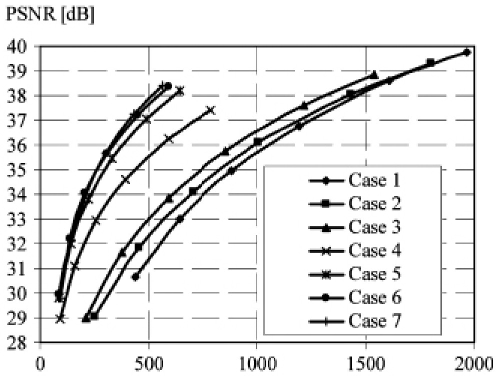


Figure 4: Effectiveness of basic techniques. 1) No MC; 2) Adding Skip mode to form a CR coder. Case 3) Allow only zero-valued MVs. 4) Allow integer-pel MC. 5) Allow half-pel MC 6) Allowing 4-MV; 7) Allowing quarter-pel MC. Refer to [7] for additional details

H.264/ MPEG-4 AVC
One of the most important developments in video coding in the last few years has been the definition of the H.264/MPEG-4 AVC[8] standard by the Joint Video Team (JVT) of the ITU and the ISO/IEC. This new standard has been referred to by many different names as it evolved. The ITU began work on H.26L (for long term) in 1997 using major new coding tools. The results were impressive, and the ISO decided to work with the ITU to adopt a common standard under a Joint Video Team. For this reason, some people refer to the standard as JVT although this is not the formal name. The ITU approved the new H.264 standard in May 2003. The ISO approved the standard in October of 2003 as MPEG-4 Part 10, Advanced Video Coding or AVC.

H.264/AVC delivers a significant break-through in compression efficiency generally achieving around 2x compression versus MPEG-2 and MPEG-4 simple profile. In formal tests conducted by the JVT[9], H.264 delivered a coding efficiency improvement of 1.5x or greater in 78% of the 85 test-cases, with 77% of those showing improvements 2x or greater and as high as 4x for some cases. The improvement offered by H.264 creates new market opportunity such as the following possibilities:

- VHS-quality video at about 600 Kbps. This can enable video delivery on demand over ADSL lines.
- An HD movie can fit on one ordinary DVD instead of requiring new laser optics.

When H.264 was standardized, it supported three profiles: baseline, main and extended. Later, an amendment called fidelity range extension (FRExt) introduced four additional profiles referred to as the high profiles. Early on, the baseline profile and main profile generated the most interest. The baseline profile requires less computation and system memory and is optimized for low latency. It does not include B frames due to its inherent latency or CABAC due to the computational complexity. The baseline profile is a good match for video telephony applications as well as other applications that require cost-effective real-time encoding.

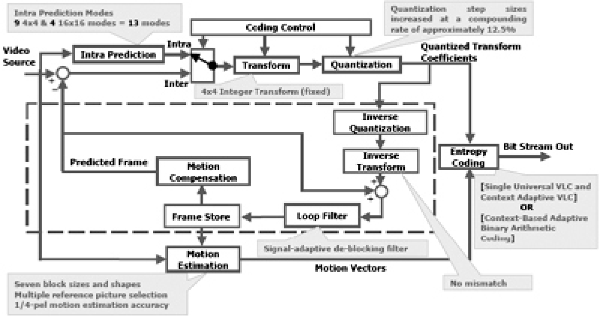
The main profile provides the highest compression but requires significantly more processing than the baseline profile, making it difficult to use in low-cost real-time encoding and low latency applications. Broadcast and content storage applications are primarily interested in the main profile to leverage the highest possible video quality at the lowest bitrate.

While H.264 uses the same general coding techniques as previous standards, it has many new features that distinguish it from previous standards, that combined improve coding efficiency. The main differences are summarized in the encoder block diagram in Figure 5 and are also described briefly below:

Intra Prediction and Coding: H.264 uses spatial domain Intra prediction to predict the pixels in an Intra-MB from the neighboring pixels in adjacent blocks. The prediction residual along with the prediction modes is coded rather than actual pixels in the block. This results in a significant improvement in intra coding efficiency.

Inter Prediction and Coding: Inter-frame coding in H.264 leverages most of the key features in earlier standards and adds both flexibility and functionality, including multiple options for block sizes for motion compensation, quarter-pel motion compensation, multiple-reference frames, generalized bi-directional prediction and adaptive loop de-blocking.

Variable Vector Block Sizes: Motion compensation can be performed using a number of different block sizes. Individual motion vectors can be transmitted for blocks as small as 4x4, so up to 32 motion vectors may be transmitted for a single MB in the case of bi-directional prediction. Block sizes of 16x8, 8x16, 8x8, 8x4 and 4x8 are also supported. Smaller block sizes improve the ability to handle fine motion detail and results in better subjective quality including the absence of large blocking artifacts.


Intra Prediction Modes
9 4x4 & 4 16x16 modes = 13 modes
Coding Control
Quantization step sizes increased at a compounding rate of approximately 12.5%
Video Source
Intra Prediction
Intra
Inter
Transform
Quantization
Quantized Transform Coefficients
4x4 Integer Transform (fixed)
Inverse Quantization
Inverse Transform
Entropy Coding
Bit Stream Out
Predicted Frame
Motion Compensation
[Single Universal VLC and Context Adaptive VLC]
OR
[Context-Based Adaptive Binary Arithmetic Coding]
Frame Store
Loop Filter
Signal-adaptive de-blocking filter
Motion Estimation
Motion Vectors
No mismatch
Seven block sizes and shapes
Multiple reference picture selection
1/4-pel motion estimation accuracy

View full size
Figure 5: H.264 Block diagram and features [10]

Quarter-Pel Motion Estimation: Motion compensation is improved by allowing half-pel and quarter-pel motion vector resolution.

Multiple Reference Frame Prediction: Up to 16 different reference frames can be used for inter-picture coding resulting in better subjective video quality and more efficient coding. Providing multiple reference frames can also help make the H.264 bitstream more error resilient. Note that this feature leads to increased memory requirement for both the encoder and the decoder since multiple reference frames must be maintained in memory.

Adaptive Loop De-blocking Filter: H.264 uses an adaptive de-blocking filter that operates on the horizontal and vertical block edges within the prediction loop to remove artifacts cause by block prediction errors. The filtering is generally based on 4x4 block boundaries, in which up to three pixels on either side of the boundary may be updated using a 4-tap filter.

Integer Transform: Previous standards that use DCT had to define rounding-error tolerances for fixed point implementations of the inverse transform. Drift caused by mismatches in the IDCT precision between the encoder and decoder were a source of quality loss. H.264 gets around the problem by using an integer 4x4 spatial transform, which is an approximation of the DCT. The small 4x4 shape also helps reduce blocking and ringing artifacts.

Quantization and Transform Coefficient Scanning: Transform coefficients are quantized using scalar quantization with no widened dead-zone. Different quantization step sizes can be chosen for each MB, similar to prior standards, but the step sizes are increased at a compounding rate of approximately 12.5%, rather than by a constant increment. Also, finer quantization step sizes are used for the chrominance component, especially when the luminance coefficients are coarsely quantized.

Entropy Coding: Unlike previous standards that offered a number of static VLC tables depending on the type of data under consideration, H.264 uses a Context-Adaptive VLC for the transform coefficients and a single Universal VLC approach for all the other symbols. The main profile also supports a new Context-Adaptive Binary Arithmetic Coder (CABAC). The CAVLC is superior to previous VLC implementations but without the full cost of CABAC.

CABAC: It uses a probability model to encode and decode the syntax elements such as transform coefficients and motion vectors. To increase the coding efficiency of arithmetic coding, the underlying probability model is adapted to the changing statistics within a video frame through a process called context modeling. Context modeling provides estimates of conditional probabilities of the coding symbols. Utilizing suitable context models, the given inter-symbol redundancy can be exploited by switching between different probability models, according to already coded symbols in the neighborhood of the current symbol. Each syntax element maintains a different model (for example, motion vectors and transform coefficients have different models). CABAC can provide up to about 10% bitrate improvement over CAVLC.

Weighted Prediction: It forms the prediction for bi-directionally interpolated macroblocks by using the weighted sum of forward and backward predictions, which leads to higher coding efficiency when scene changes fades.

Fidelity Range Extension

In July 2004, a new amendment called Fidelity Range Extension (FRExt)[11] was added to the H.264 standard. This extension introduced an additional set of tools into H.264 and also allowed the use of additional color spaces, video formats and bit-depths. Additional support for lossless inter-frame coding and stereo-view video were introduced. The FRExt amendment introduced four new profiles to H.264.

- High Profile (HP) for standard 4:2:0 chroma sampling with 8-bit color per component.
- New tools were introduced for this profile; described in more detail below.
- High 10 Profile (Hi10P) for 10-bit color with standard 4:2:0 chroma sampling for higher fidelity video displays.
- High 4:2:2 10-bit color profile (H422P) useful for source editing functions such as alpha blending.
- High 4:4:4 12-bit color profile (H444P) for the highest quality source editing and color fidelity supporting lossless coding for regions of the video, and a new integer color space transform (from RGB to YUV and back).

Among the new profiles, H.264 HP, which maintains 8-bit components and 4:2:0 chroma sampling, appears especially promising to the broadcast and DVD community. Some experiments show as much as 3x gain for H.264 HP over MPEG-2. Below are the key additional tools introduced in H.264 HP.

Adaptive Residual Block Size and Integer 8X8 Transform: The residual blocks can be switched between 8x8 and 4x4 blocks for transform coding. A new 16-bit integer transform for 8x8 blocks. The older 4x4 transform can continue to be used for smaller block sizes.

8x8 Luma Intra Prediction: Additional eight modes were added to allow luma intra macroblocks to perform intra prediction on 8x8 blocks in addition to previously 16x16 and 4x4 blocks

Quantization Weighting: New quantization weighting matrices for quantization of 8x8 transform coefficients

Monochrome: Supports coding of black and white video.

Windows Media Video 9 / VC-1

Windows Media is a leading format for music and video subscription services and streaming video on the Internet. In 2002, Microsoft introduced the Windows Media Video 9 Series codec providing a major improvement in video compression efficiency. WMV9 is also standardized in SMPTE as VC-1 [12].

Similar to H.264, it includes many advanced coding tools, although there are differences in the specifics. WMV9's ME allows ¼ pel bi-cubic (using 4-tap approximate bi-cubic filters) interpolation in addition to support for ½ pixel bilinear interpolation. It also includes an in-loop de-blocking filter similar to H.264 but with different details on the filters and decisions. Some other important features are:

Multiple VLC Tables: WMV9 main profile contains multiple sets of VLC tables that are optimized for different types of content. Tables can be switched at a frame level to adjust to the characteristics of the input video.

DCT/IDCT Transform Switch: WMV9 supports multiple DCT block sizes including 8x8, 8x4, 4x8 and 4x4 and uses a special 16-bit integer transform and inverse transform.

Quantization: Both regular step-size based quantization and dead-zone quantization is used. Use of dead-zone quantization allows substantial savings at lower bitrates.

Another interesting feature is the ability to use an explicit fading compensating for scenes involving fading. This improves the quality of motion compensation in these scenarios.

WMV9/VC-1 achieves significant performance improvements over MPEG-2 and MPEG-4 simple profile and has fared well in some perceptual quality rating comparisons with H.264[13]. Although WMV9/VC-1 delivers similar compression efficiency, it has lower complexity requirement compared to H.264 main profile. WMV9 is used heavily in the PC environment and could also become important in networked consumer appliances. WMV9/VC-1 is gaining momentum with Hollywood and independent film industry with various movie titles soon to be released encoded in WMV9/VC-1 for high-definition playback on PC DVDs. WMV9 is also standardized as a compression option for the upcoming HDDVD and Blu-ray formats.

AVS

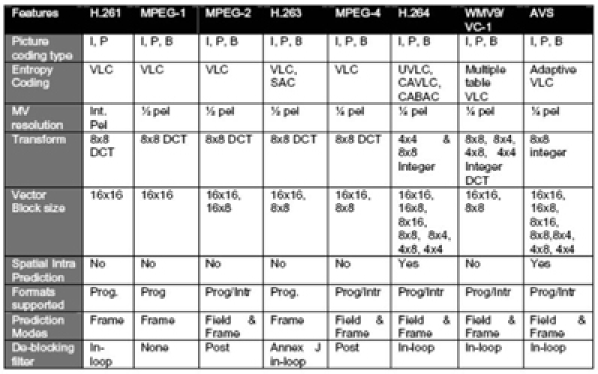
In 2002, the Audio-Video Standard (AVS) working group established by the Ministry of Information Industry of China announced an effort to create a national standard for mobile multimedia, broadcast, DVD, etc. The video standard, referred to as AVS [14], consists of two related parts, a AVS-M for mobile video applications and AVS1.0 for broadcast and DVD. The AVS standards are similar to H.264.

AVS1.0 supports both interlaced and progressive modes. AVS allows the use of 2 previous reference frames for P frames while allowing one future and one previous frame for B frames. In interlaced mode, up to four fields are allowed for reference. Frame/Field coding in interlaced mode can be performed at a frame-level only, unlike H.264 where MB level adaptation of this option is allowed. AVS has a loop filter similar to H.264, which can be disabled at a frame level. Also, no loop-filter is required in B pictures. The intra-prediction is done on 8x8 blocks. MC allows up to ¼ pel for luma blocks. The block sizes for ME can be 16x16, 16x8, 8x16 or 8x8. The transform is a 16-bit based 8x8 integer transform, similar to WMV9. VLC is based on context adaptive 2-D run/level coding. Four different Exp-Golomb codes are used. The code used for each quantized coefficient is adaptive to the previous symbols within the same 8x8 block. Since Exp-Golomb tables are parametric, table sizes are small. The visual quality of AVS 1.0 for progressive video sequences is marginally inferior to H.264 main profile at the same bitrate.

AVS-M is targeted especially at mobile video applications and overlaps with H.264 baseline profile. It only supports progressive video, I and P frames, and no B frames. The main AVSM coding tools include 4x4 block-based intra prediction, quarter-pel motion compensation, integer transform and quantization, context adaptive VLC and a highly simplified loop-filter. Similar to H.264 Baseline Profile, in AVS-M the motion vector block size can be down to 4x4, and consequently a MB can have up to 16 motion vectors. Multiple frame prediction is used but it only requires up to two reference frames. A subset of H.264 HRD/SEI messages is also defined in AVS-M. On average and with similar settings, the coding efficiency of AVS-M is about 0.3 dB worse than H.264 Baseline profile, while decoder complexity is about 20% lower.

Comparison of codecs based on features and tools

Table 1 summarizes the key compression features and tools used in the video standards that we reviewed.

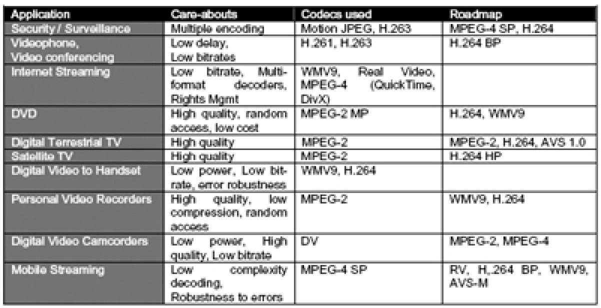


Features	H.261	MPEG-1	MPEG-2	H.263	MPEG-4	H.264	WMV9/ VC-1	AVS
Picture coding type	I, P	I, P, B	I, P, B	I, P, B	I, P, B	I, P, B	I, P, B	I, P, B
Entropy Coding	VLC	VLC	VLC	VLC, SAC	VLC	UVLC, CAVLC, CABAC	Multiple table VLC	Adaptive VLC
MV resolution	Int. Pel	½ pel	½ pel	½ pel	¼ pel	¼ pel	¼ pel	¼ pel
Transform	8x8 DCT	8x8 DCT	8x8 DCT	8x8 DCT	8x8 DCT	4x4 & 8x8 Integer	8x8, 8x4, 4x8, 4x4 Integer DCT	8x8 integer
Vector Block size	16x16	16x16	16x16, 16x8	16x16, 8x8	16x16, 8x8	16x16, 16x8, 8x16, 8x8, 8x4, 4x8, 4x4	16x16, 8x8	16x16, 16x8, 8x16, 8x8,8x4, 4x8, 4x4
Spatial Intra Prediction	No	No	No	No	No	Yes	No	Yes
Formats supported	Prog.	Prog	Prog/Intr	Prog.	Prog/Intr	Prog/Intr	Prog/Intr	Prog/Intr
Prediction Modes	Frame	Frame	Field & Frame	Frame	Field & Frame	Field & Frame	Field & Frame	Field & Frame
De-blocking filter	In-loop	None	Post	Annex J in-loop	Post	In-loop	In-loop	In-loop

View full size

Table 1: Key compression features in standard codecs

Market Trends and Applications
Video compression is enabling a growing number of digital video products in the market. End equipments using digital video compression range from battery-operated portable devices to high-performance infrastructure equipment. Table 2 shows a snapshot of some applications, key care-abouts, typical video codecs used and potential roadmap of codec usage in these applications.



Application	Care-abouts	Codecs used	Roadmap
Security / Surveillance	Multiple encoding	Motion JPEG, H.263	MPEG-4 SP, H.264
Videophone, Video conferencing	Low delay, Low bitrates	H.261, H.263	H.264 BP
Internet Streaming	Low bitrate, Multi-format decoders, Rights Mgmt	WMV9, Real Video, MPEG-4 (QuickTime, DivX)	
DVD	High quality, random access, low cost	MPEG-2 MP	H.264, WMV9
Digital Terrestrial TV	High quality	MPEG-2	MPEG-2, H.264, AVS 1.0
Satellite TV	High quality	MPEG-2	H.264 HP
Digital Video to Handset	Low power, Low bit-rate, error robustness	WMV9, H.264	
Personal Video Recorders	High quality, low compression, random access	MPEG-2	WMV9, H.264
Digital Video Camcorders	Low power, High quality, Low bitrate	DV	MPEG-2, MPEG-4
Mobile Streaming	Low complexity decoding, Robustness to errors	MPEG-4 SP	RV, H,264 BP, WMV9, AVS-M

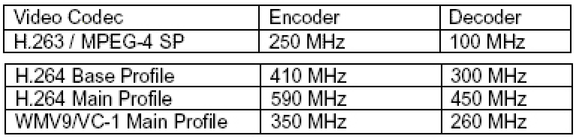
View full size

Table 2: Codecs typically used in standard applications and roadmap

Real-time implementations
The optimal processor solution for digital video depends on the specific target application. Texas Instruments has a wide variety of DSPs that support multiple standards and fit key design and system constraints. TI's solutions range from the low power TMS320C5000™ DSPs and mobile OMAP application processors to the high performance TMS320C6000™ DSPs and video optimized high-performance TMS320DM644x digital media processors. One of the processors that is currently generating a lot of interest is the recently unveiled DM6446, which we describe in this section.

TI's DM64x family of processors is specifically designed to address the requirements of highend video systems. The latest in this family of processors is the powerful DM6446[15] that leverages TI's DaVinci™ technology[16]. The dual-core architecture of the DM6446 provides benefits of both DSP and RISC technologies, incorporating a C64x+ DSP core that can be clocked at 594 MHz and an ARM926EJ-S core. The C64x+ DSPs are the highest-performing fixed-point DSP generation in the C6000 DSP platform and are based on an enhanced version of the second-generation high-performance, advanced VLIW architecture developed by TI. The C64x+ is code-compatible with older members of the C6000 DSP platform. Programmable digital media processors such as the DM644x can support all of the existing industry standards and proprietary video formats using a single programmable digital media processor. The DM6446 has on-chip memory including a two-level cache and a large set of peripherals many with video-specific features. The DM6446 also includes a Video/Imaging Co-processor (VICP) to offload many video and imaging processing tasks from the DSP core for algorithms such as JPEG, H.264, MPEG-4 and VC-1, leaving more DSP MHz available for video post-processing or other functions running in parallel.

Table 3 shows the approximate MHz required on the DM6446 to sustain D1 (720x480) resolution for various standards.



Video Codec	Encoder	Decoder
H.263 / MPEG-4 SP	250 MHz	100 MHz
H.264 Base Profile	410 MHz	300 MHz
H.264 Main Profile	590 MHz	450 MHz
WMV9/VC-1 Main Profile	350 MHz	260 MHz

Table 3: Representative MHz requirements of stand-alone video codecs on TI's DM6446 platform for D1 30fps (720X480) for YUV 4:2:0. Decoder performance numbers are for worst case bitstreams. Encoder performance can vary as a result of feature set used. High quality encoding is assumed in the examples above. The C64x+ on the DM6446 can be clocked at 594 MHz.

Note that the encoding MHz numbers shown are based on typical test data for an existing or planned implementation. Also note that the encoder loading can vary dramatically depending on the target application. Compression standards specify the required syntax and available tools, but many algorithmic decisions are left up to the implementation. Key variables include the bitrate control algorithm, single-pass versus multi-pass encoding, ratio of I/B/P frames, motion search range, motion search algorithm and the choice of available individual tools and modes that are used. This flexibility allows different trade-offs between computational loading and incremental quality improvements. Clearly, for all encoders, higher or lower MHz can be consumed to enable a different visual quality point.

Conclusion
A growing number of video compression standards offering increased compression efficiency and a wider variety of tools can be tailored for specific end applications. Also, the trend toward networked connectivity means that many products will have to support more than one standard. The proliferation of multiple standards and proprietary algorithms also makes it difficult to select one standard, especially since hardware decisions are often made far in advance of product deployment. In addition, each video coding algorithm offers a wide choice of tools and options to trade-off complexity for compression efficiency. The choice of tools and options is an iterative process that is application and use-case specific. With an increasing number of codecs requiring support and options for optimizing the codec for specific scenarios and applications become broader, there is a trend toward flexible media processors in digital video systems. Digital media processors such as the DM6446, offer the performance headroom and architecture flexibility to quickly bring to market implementations of new standards including H.264, AVS and WMV9. Algorithms can be implemented during the standard definition phase. The algorithms and tools made in software can be updated to keep up with minor and major adjustments to both the standard and changing quality tradeoffs that the application may demand.

References
[1] J. R. Jain and A. K. Jain, "Displacement measurement and its application in interframe image coding," IEEE Trans. Commun., vol. COM-29, no. 12, pp. 1799–1808, Dec. 1981.
[2] ITU-T Recommendation H.261: 1993. Video codec for audiovisual services at px64 Kbit/s.
[3] ISO/IEC 11172-2:1993. Coding of moving pictures and associated audio for digital storage media at up to 1.5 Mbit/s -- Part 2: Video.
[4] ISO/IEC 13818-2:1995. Generic coding of moving pictures and associated audio information: Video.
[5] ITU-T Recommendation H.263: 1998. Video coding for low bit rate communication. [6] ISO/IEC 14496-2:2001. Information technology – Generic coding of audio-visual objects – Part 2: Visual.
[7] G. Sullivan & T. Wiegand, "Video Compression—From Concepts to the H.264/AVC Standard", Proceedings of the IEEE, Vol 93, No.1, Jan 2005
[8] ISO/IEC 14496-10:2003. Information technology – Coding of audio-visual objects. Part 10: Advanced Video Coding
[9] Report on the Formal Verification Tests on AVC (ISO/IEC 14496-10 | ITU-T Rec. H.264), ISO/IEC JTC1/SC29/WG11, MPEG 2003/N6231, December 2003, Waikoloa.
[10] H.264 White paper UB Video Inc. www.ubvideo.com
[11] Joint Video Team of ITU-T and ISO/IEC: "Draft Text of H.264/AVC Fidelity Range Extensions Amendment", Doc. JVT-L047, Sept. 2004
[12] SMPTE 421M, Draft SMPTE Standard for TV: VC-1 Compressed Video Bitstream Format and Decoding Process
[13] Srinivasan, S.; (John) Hsu, P.; Holcomb, T.; Mukerjee, K.; Regunathan, S.L.; Lin, B.; Liang, J.; Lee, M.-C.; Ribas-Corbera, J., "Windows Media Video 9: overview and applications", Signal Processing: Image Communication, Volume 19, Issue 9, 1 October 2004, Pages 851-875
[14] L. Fan, S. Ma, F. Wu: "Overview of AVS video standard", Proceedings of the ICME 2004: 423-426
[15] "TMS320DM6446 Digital Media System-on-Chip", www.ti.com
[16] Bill Witowsky, Gene Frantz, "DaVinci™ Technology for Digital Video", www.ti.com

This paper was written for and presented at the Embedded Systems Conference Silicon Valley 2006. For more information, please visit www.embedded.com/esc/sv

About the Authors
Jeremiah Golston is a distinguished member technical staff for Texas Instruments and the chief technical officer for the DSP Video and Imaging group. He is responsible for the DSP video and imaging device architecture roadmap and leading systems and software development. Golston led the definition of the TMS32064x instruction set extensions to the C6000 DSP family with focus on added performance for video, imaging and broadband communication algorithms. He holds numerous patents in DSP architecture and optimized algorithm implementations. Golston earned a master's and bachelor's degree in electrical engineering from the University of Missouri-Rolla. He can be reached at j-golston@ti.com.

Dr. Ajit Rao is the manager for multimedia codecs in Texas Instruments. In this role he is responsible for development and technical direction for TI's multimedia codecs products. Prior to joining TI three years ago, he was a lead codec developer at Microsoft and SignalCom. With more than seven years of codecs experience, Ajit holds four U.S. patents. He has a Ph.D. and Master of Sciences in Electrical Engineering from the University of California, Santa Barbara. He also earned a Bachelor of Technology degree in Electronics & Communication from the Indian Institute of Technology, Madras. He can be reached at ajitr@ti.com

文章录入：admin　责任编辑：admin

- 上一篇文章： Analysis: Low-cost, DSP-only DaVinci chips
- 下一篇文章： IPTV ecosystems and DSP-based set top box design

【发表评论】【加入收藏】【告诉好友】【打印此文】【关闭窗口】

最新热点

DSP6748 DDR2的测试程序
TI C6478 的UART 使用例程…
evmc6748_bootloader研究
[组图]启动仿真器就是这么…
A8架构下CPSR说明
DM8168硬件调试-GPIO
DM8168裸机调试-GPMC+FPGA
解决CCS5.5 调试DM8168过程…
TMS320F28335学习之CCS5.2…
[组图]基于TMS320C6678的多…

最新推荐

[图文]TMS320VC54xx通用FL…
[图文]TMS320LF2407系统和…
[组图]烧写TMS320F2812芯片…
[组图]XDS510 DSP系列仿真…
[图文]JTAG口与仿真器连接…
[图文]TMS320VC5410硬件连…
[图文]采用ADuC834的口0与…
[图文]AD73360与TMS320VC5…
[图文]DSP5402和MAX3110硬…
[图文]主从模式的TLC320AC…

相关文章

DM642 EVM Video ADC

网友评论：（只显示最新10条。评论内容只代表网友观点，与本站立场无关！）

没有任何评论