Encoding Video for the Highest Quality and Performance

Fabio Sonnati

2 December 2008
Milan, MaxEurope 2008

Adobe

Encoding Video for the Highest Quality and Performance

- **Fabio Sonnati** media applications consultant, Flash Community Expert (FMS) and FMS Cab (Customer Advisory Board), FMS developer and beta tester since 2003 with expertise of Video Encoding optimizations. Collaborates with leading IT consultancy firms at the development of Video encoding & delivery platforms.
  Blog: http://flashvideo.progettosinergia.com
  Mail: sonnati@progettosinergia.com

Encoding Video for the Highest Quality and Performance

- To make video look the best it can is an alchemy of at least four elements:
  - **Knowledge of the inner secrets of video encoding**
  - **A good video encoder**
  - **Best Practices of HQ video encoding**
  - **Time and Passion**

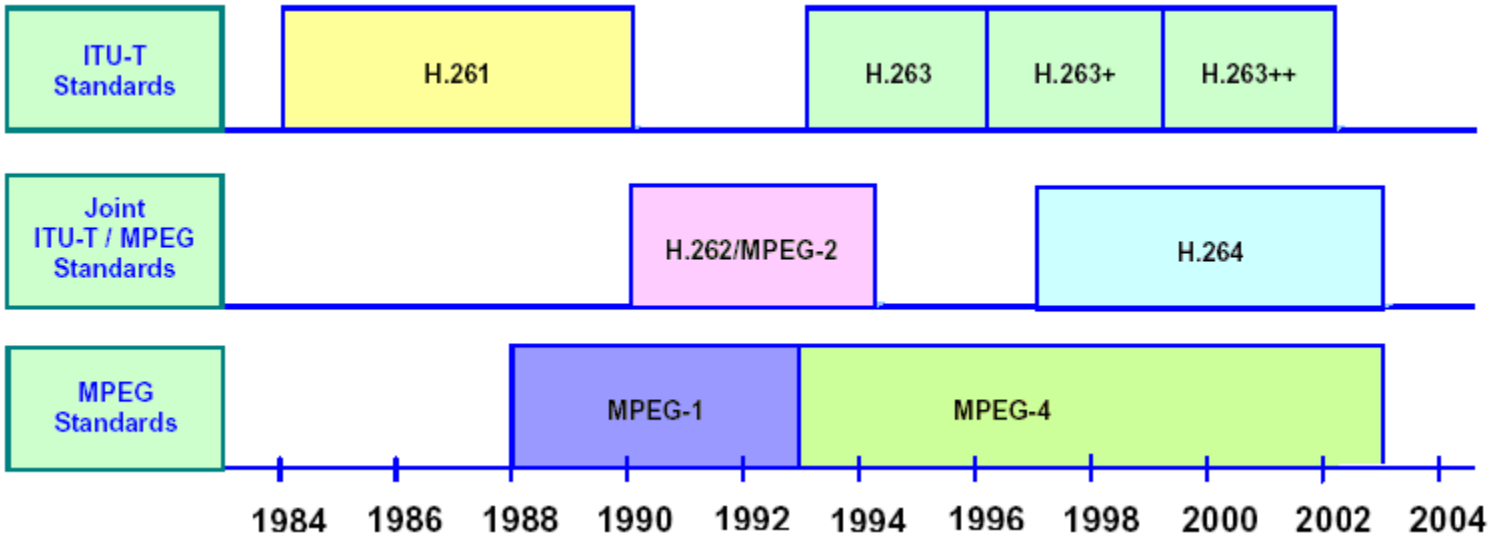
Agenda

MAX CONNECT. DISCOVER. INSPIRE.

- Video encoding standards overview
- Understanding video compression
- H.261, H.263 and H263v2
- MPEG4-AVC (H.264)
- H.264 Profiles and Levels
- Codecs supported by Flash Player
- Understanding H.264's parameters
- Best practices for Hi-Quality encoding
- Multi-bitrate and FMS 3.5
- Q&A

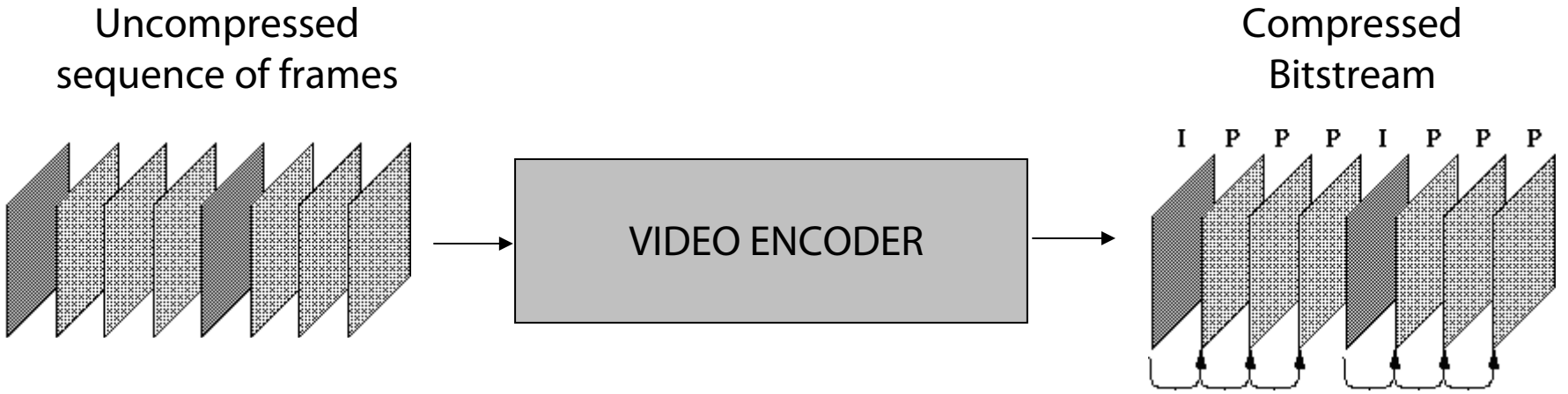
Adobe

Video encoding standards - overview



- ITU and ISO are the formal organizations for video codec standardization
- ITU develops the “core” logic of the codec (H.26x) for generic applications
- ISO defines “industry standards” for storage and broadcasting using H.26x

Understanding video compression

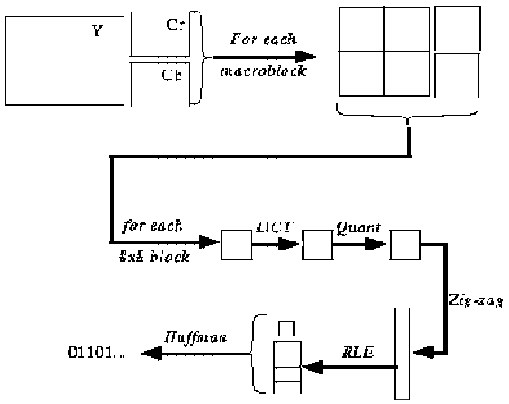


The video encoders compress data exploiting
spatial & temporal
redundancies

Compression Techniques : INTRA FRAME COMPRESSION

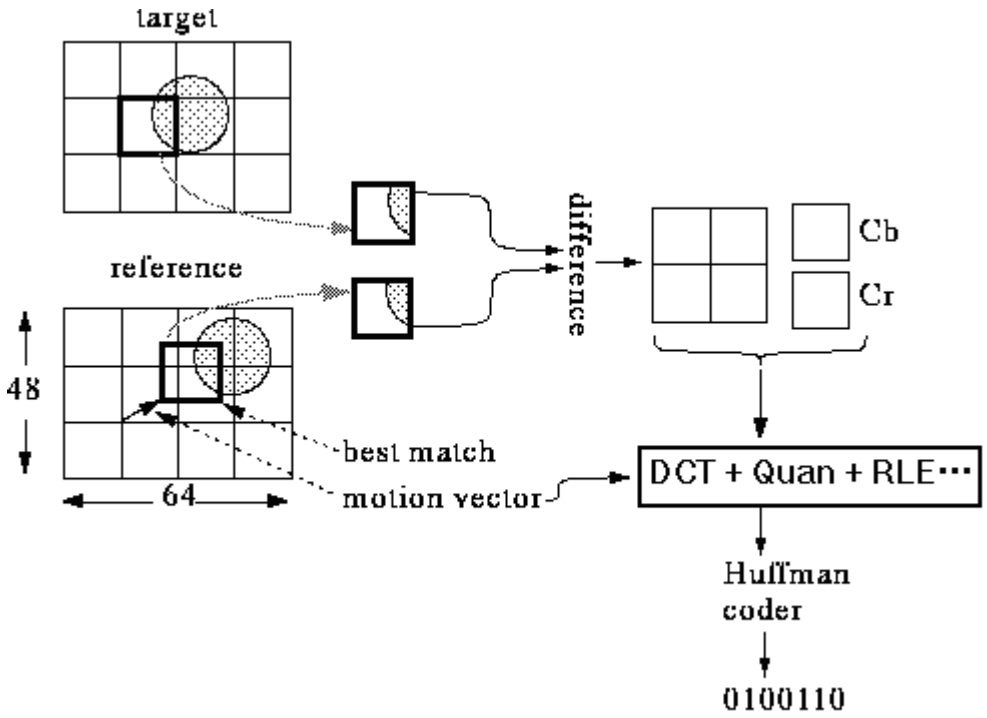
The video frame is compressed only exploiting intra-frame redundancies.
It is like JPEG compression:

- *1. The frame is converted in YCbCr color space and then the **chroma** planes (CbCr) are down-sampled (4 : 2 : 2 or 4 : 2 : 0 format) because human eye is more sensible to **luma** (Y) .*
- *2. The frame is divided in blocks of 8x8 pixels which are transformed from spatial to frequency domain using a Discrete Cosine Transform (DCT). Frequency coefficients are quantized: more bits for the lower and less for the higher frequency coefficients, because human eye is less sensible to fine details.*
- *3. Entropy Coding (Variable Length Coding) reduces residual redundancies (similar to zip compression).*



Compression Techniques : INTER FRAMES COMPRESSION

- *The video frame is compressed exploiting time-based redundancies in a frames sequence*
- *The frame is divided in macro-blocks (usually 16x16 pixels). For each macroblock in the current frame, the encoder tries to find the region in the previous frame which are more similar to it. The current macroblock is therefore "predicted" from the previous frame using a "motion vector" and the residual delta information which are transformed, quantized and entropy coded.*



Understanding video compression

MAX CONNECT. DISCOVER. INSPIRE.

Types of Video Frame :

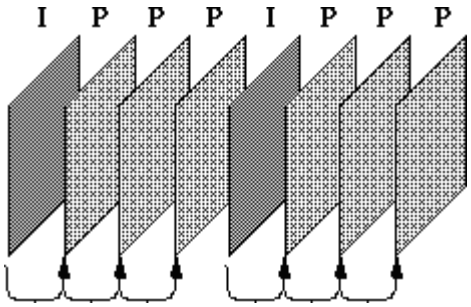
- *I-frames*
  - *Encoded only spatially (Intra). I-frames (Keyframes) are "self contained" and used for stream accessibility.*
- *P-frames*
  - *Predicted from previous reference frames using motion estimation and compensation (Inter). They are not self-contained and form a chain of references.*
- *B-frames*
  - *Bi-directionally Interpolated from previous and next reference frames (Inter). They are usually not used as reference so can be dropped without affecting the decoding. Usually used for accessibility and temporal scalability*

Adobe

H.261

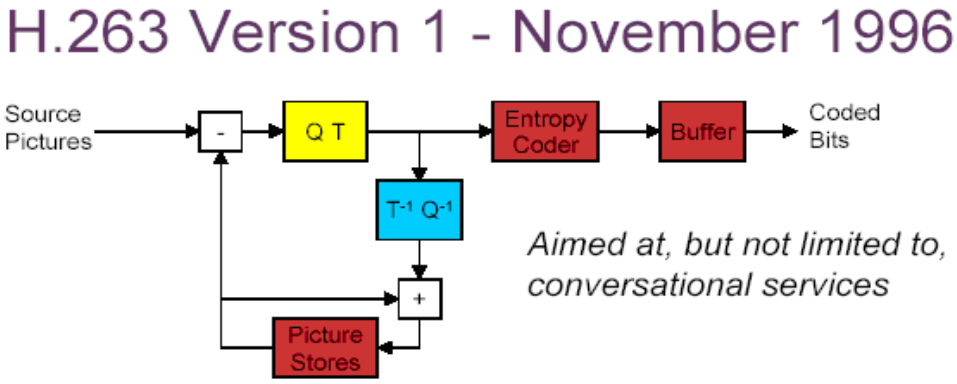
- H. 261 Compression was designed for video telecommunication applications. Developed by ITU in 1988-1990, H.261 has been widely used in video telephone applications, videoconference systems and MPEG1

  - CIF (352x288) and QCIF (176x144) resolutions in colour space YCbCr with 4:2:0 chrominance subsampling.
  - Two frame types: I and P
  - Simplified motion compensation in P-frames (single “pel” and short motion range)
  - Simple Variable Length Coding



- ITU-T developed H.263 in the years 1993-1996. H.263 has the goal to achieve better compression than H.261 with much more flexibility, especially for low bit rate IP channels.

  - H.263 may be thought as an evolution of H.261 combined with MPEG-like features and others optimizations for lower bit rates. Compared to H.261, H.263 has the following base improvements:

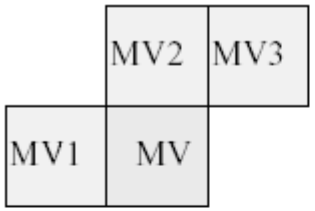


Derived from H.261, MPEG-1 and MPEG-2

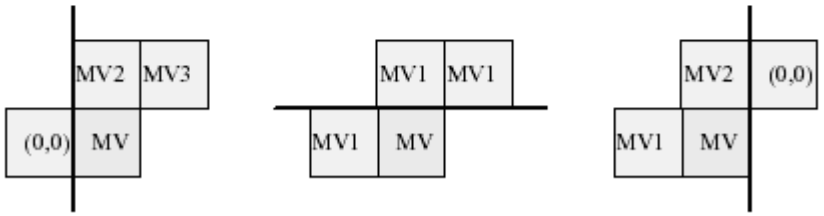
- Half Pixel Motion Compensation, 16x16 and 8x8 blocks
- Discrete Cosine Transform
- Motion vectors off picture
- Overlapped block motion compensation
- PB frames

H.263

- Resolution of Motion Vectors is now half-pel (half-pixel) in the range [-16,+15.5]
- The generic Motion Vector is predicted by the values of the surrounding MV. In fact, in a frame, the motion is a local property and adjacent blocks have similar motion vectors. The predicted value is used to reduce the amount of information transmitted. Only the difference signal (delta signal) between the real vector and its prediction is transmitted (see Figure).



MV: Current motion vector
MV1, MV2, MV3: predictors
prediction = median(MV1,MV2,MV3)



H.263 and Flash

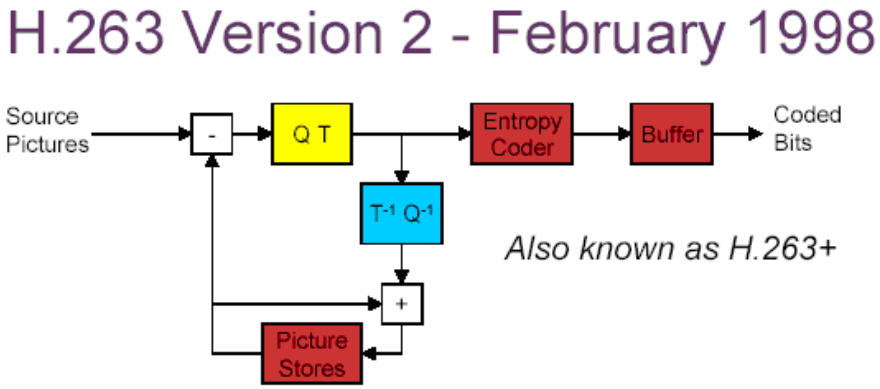
- **Flash Player 6 introduced the support for the codec «Spark»**
- **Sorenson's Spark is based on H.263v1**
- **YouTube's videos are mostly encoded with Spark**

H.263v2

H.263v2 are the natural evolution of the base standard. ITU-T developed H.263v2 in the years 1996-1998. The basic concepts and techniques of these standards may be found in the later MPEG4 standard.

Enhancements of H.263v2 over H.263 are:

- ***Extended source formats***
- ***16 negotiable optional modes***
- ***Supplemental enhancements***



Further optional modes of operation added:

- Advanced Intra Coding - using spatial prediction
- Deblocking Filter
- Reference Picture Selection
- Scalability and B pictures
- Reference Picture Resampling

H.263v2 – negotiable options

Advanced Intra Coding Mode

- The standard introduces a spatial prediction of DCT coefficients in Intra compression. This is similar to Motion Vector prediction but applied to DCT coefficients. There are 3 prediction mode: DC only, vertical DC & AC, horizontal DC & AC. 10% improvement in Intra compression.

Alternate Inter VLC Mode

- This mode uses separate VLC table for inter DCT and intra DCT. The use of different VLC tables for the various parameters allows better compression at the cost of higher coding complexity.

De-blocking Filter Mode

- Depending on quantization step size the codec applies a de-blocking filter to Blocks in order to improve visual quality perception. This is very helpful in the case of very high compression ratio.

Modified Quantization Mode

- More flexible changes of quantization step sizes and finer quantization for chrominance. The gain in SNR for chrominance levels is considerable. Less chromatic artifacts.

Improved PB-Frame Mode

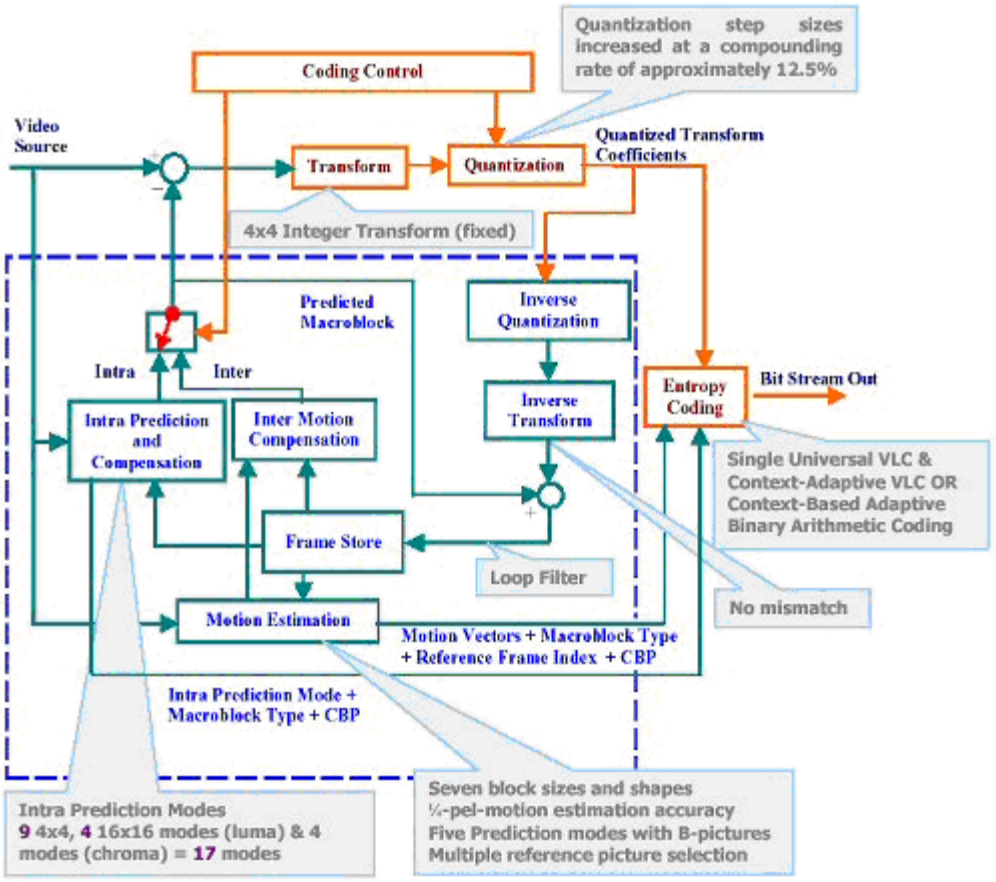
- For each Macroblock is now possible to choose for forward, backward, or bi-directional prediction. The motion vectors are predicted from the mean values of the previous and following frame. If prediction is good enough, delta signal is not transmitted at all with a consequent bandwidth gain.

On2's VP6

- Flash Player 8 introduced the codec VP6
- On2's VP6 is a proprietary technology with many points in common with H.263v2:
  - 8x8 transformation, intra prediction
  - Quarter-pel motion estimation
  - 4 motion vectors per Macroblock
  - 2 reference frames (for P frames, no B-frame support)
  - In-loop deblocking filter
  - Complex variable length coding
  - (VP6-S, disable some features like deblocking to speed up decoding)
- (Note: DivX and VC-1 use techniques similar to H.263v2)

H.264 - MPEG4-AVC

- **H.264**, **MPEG-4 Part 10 (or AVC)** was written by the ITU-T together with the ISO/IEC Movie Picture Experts Group (MPEG) as the product of a collective partnership effort known as the Joint Video Team (JVT). The ITU-T **H.264** standard and the ISO/IEC MPEG-4 **Part 10** standard (formally, ISO/IEC 14496-10) are technically identical. The final drafting work on the first version of the standard was completed in May of 2003.
- H.264 contains a number of new features that allow it to compress video much more effectively than older H.26x standards:



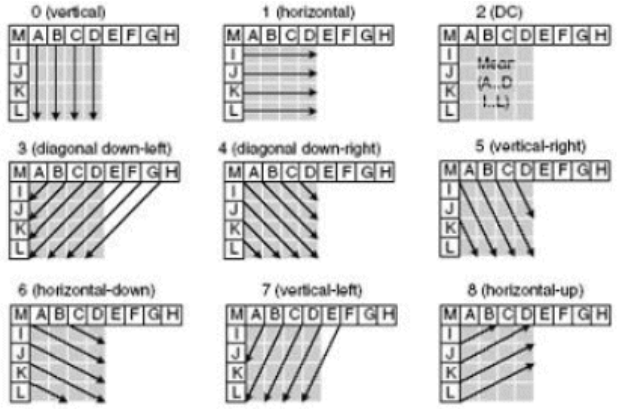
New transform design

- An exact-match integer 4×4 spatial block transform is used instead of the well known 8x8 DCT. It is conceptually similar to DCT but with less ringing artifacts. There is also a 8x8 spatial block transform for less detailed areas and chroma.

  A secondary Hadamard Transform (2x2 on chroma and 4x4 on luma) can be usually performed on "DC" coefficients to obtain even more compression in smooth regions.

Intra-frame prediction

- H.264 introduces complex spatial prediction for intra-frame compression.
- Rather than the "DC"-only prediction found in MPEG2 and the transform coefficient prediction found in H.263+, H.264 defines 6 prediction directions (modes) to predict spatial information from neighbouring blocks when encoded using 4x4 transform. The encoder try to predict the block interpolating the colour value of adjacent blocks. Only the delta signal is therefore encoded.
- There are also 4 prediction modes for smooth colour zones (16x16 blocks). Residual data are coded with 4x4 transforms and a further 4x4 Hadamard transform is used for DC coefficients.

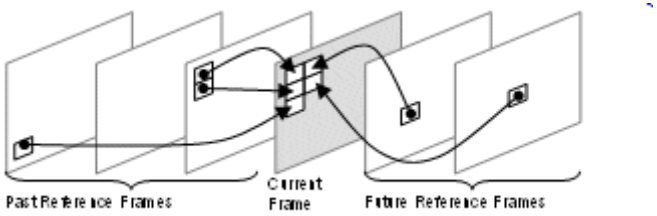


Improved quantization

- A new logarithmic quantization step is used (compound rate 12%). It's also possible to use Frequency-customized quantization scaling matrices selected by the encoder for perceptual-based quantization optimization.

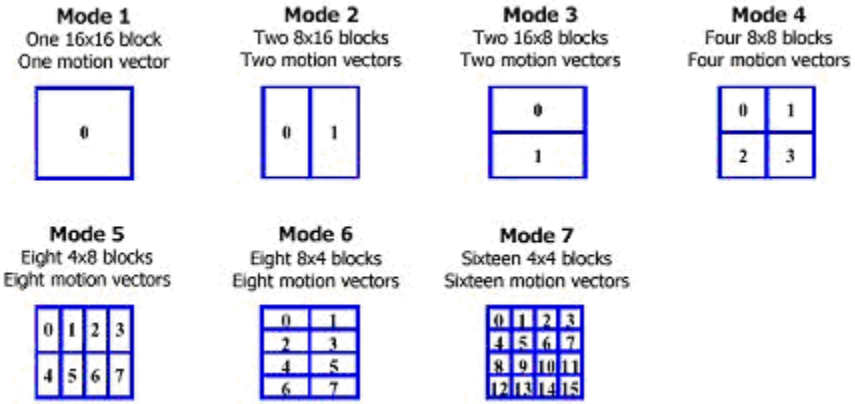
Multiple Reference Frames

- H.264 uses previously-encoded pictures as references in a much more flexible way than in past standards, allowing up to 16 reference pictures to be used (unlike in prior standards, where the limit was typically one or, in the case of conventional B frame, two). In certain scenarios, for example scenes with rapid repetitive flashing or back-and-forth scene cuts or uncovered background areas, it allows a very significant reduction in bit rate.



Enhanced Motion Compensation

- H.264 uses P frames (predicted) and B frames (interpolated) with full pixel, half-pixel and quarter pixel resolution. Motion compensation is done using 7 macroblock configurations with block size as large as 16x16 and as small as 4x4. Each macroblock can have a different reference picture. B frames are predicted from previous and/or future pictures with 5 prediction Modes (intra, forward, backward, interpolated and direct) designed to suit different scenarios. It is also possible to use B-Frames as reference for other B-Frames (B-pyramid)
- Weighted prediction allows an encoder to specify the use of a scaling and offset when performing motion compensation providing a significant benefit in performance in special cases, such as fade-to-black, fade-in, and cross-fade transitions.



In-Loop De-blocking Filter

- Loop filtering is mandatory in the encoder, it identify a blocking situation depending by two threshold factors (alpha and beta). A lot of efficiency is due to the loop filter. The strength of filter depends on intra/inter coding, differential vectors, quantization level. Up to 40% of total processing power may be required by this kind of filter. Filtering the reference frames prior to use them in prediction can significantly improve the objective and perceptual quality especially at low or medium bitrates.

Entropy Coding

- For entropy coding, H.264 may use an enhanced VLC, a more complex context-adaptive variable-length coding (CAVLC) or an ever more complex Context-adaptive binary-arithmetic coding (CABAC) which are complex techniques to losslessly compress syntax elements in the video stream knowing the probabilities of syntax elements in a given context. The use of CABAC can improve the compression of around 5-7%. CABAC may requires a 20-30% of total processing power to be accomplished.

Codec "Profiles"

Profiles define exactly what techniques and strategies can be used by the encoder and the decoder. Simple profiles require less processing power and less memory but achieve a worst quality/bitrate ratio.

- **Baseline Profile (BP):** Primarily for lower-cost applications with limited computing resources, this profile is used widely in videoconferencing and mobile applications.
- **Main Profile (MP):** Originally intended as the mainstream consumer profile for broadcast and storage applications, the importance of this profile faded when the High profile was developed for those applications.
- **High Profile (HiP):** The primary profile for broadcast and disc storage applications, particularly for high-definition television applications (this is the profile adopted into HD DVD and Blu-ray Disc).

H.264 : Profiles and Levels

MAX CONNECT. DISCOVER. INSPIRE.

Codec "Levels"

- Levels define the *max resolution of frames*, *max required memory*, *max local bitrate and buffering*. Levels are important for device compatibility. Usually the baseline profile is used with a level up to 3.1, the main profile with a level 4.1 and the high profile with levels up to 5.1
- The good new is that Flash Player supports every levels and every profiles so the best is to use High profile with level 4.1 or (5.1 for Full HD).

Adobe

MAX
CONNECT. DISCOVER. INSPIRE.

Encoding Best-Practices

Adobe

Video Codecs supported by Flash Player

MAX
CONNECT. DISCOVER. INSPIRE.

- **Sorenson's Spark (derived from H.263)**
  Flash Player 6+, 99% of Internet penetration, fast and simple.
- **On2's VP6 (mpeg4-class codec)**
  Flash Player 8+, 98-99%, fast (VP6-S) and efficient (VP6-E).
- **H.264 (implementation provided by MainConcept)**
  Flash Player 9 update 3 – FP10, 90%, "state of the art" in video encoding

Adobe

H.264 : the Flash implementation

MAX CONNECT. DISCOVER. INSPIRE.

- The H.264 decoder implemented in Flash Player is very good
- Supports **baseline**, **main**, **high** and **high10** profiles and every level
- Supports multi-core processors for decoding (up to 4 cores)
- Supports .mp4 file format as H.264 video container

Adobe

H.264 : Encoders

- Apple's QuickTime
- Nero Digital's H.264 encoder
- ***MainConcept's Reference***
- ***Adobe's CS4 Media Encoder***
- Sorenson's Squeeze
- On2's Flix

- Flash Media Encoder Server

Understanding H.264 parameters

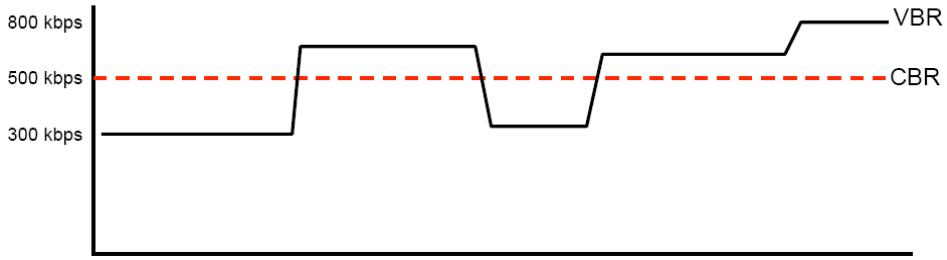
- **Profiles and Levels**

Flash Player has a robust and complete implementation of H.264, so
If the target is only the web delivery use **HIGH profile** and **4.1 level**

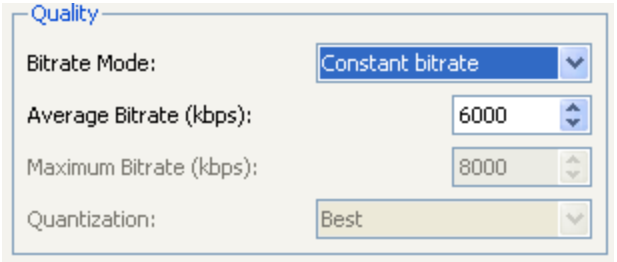
Understanding H.264 parameters

Bitrate and quality related parameters:

- Constant Bitrate
  - One bitrate for the entire video
  - Useful for streaming
  - Not optimal for quality

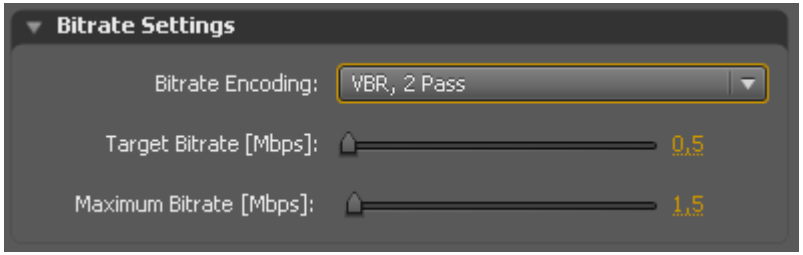


- Variable Bitrate
  - More bits in fast moving scenes, less bits in static scenes
  - Good for progressive download, bad for streaming



Multi-pass encoding

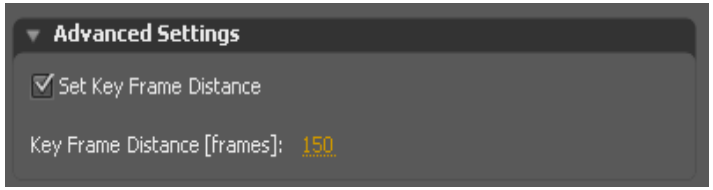
- When available use 2-pass encoding



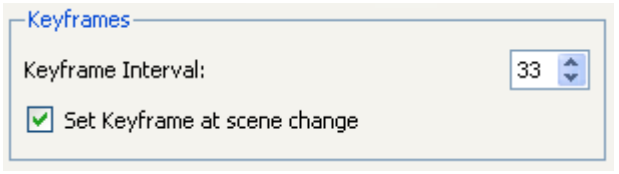
Understanding H.264 parameters

IDR frames are I-Frames which are not “crossed” by multi frames referencing. Two consecutive IDR frames isolate a Group-of-pictures (GOP).

- IDR Interval
  - It is the distance between the keyframes. The value depends by the level of accessibility you want to give to your media.
  - Recommended range 50-250 (better if “Dynamic”)

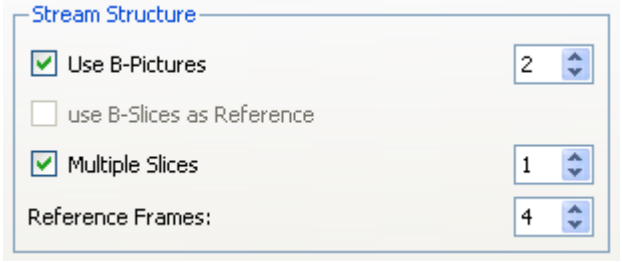


- Scene change threshold
  - Dynamic IDR positioning is guided by scene change detection. Scene change threshold is usually a value in the range 0-100. Recommended values: 40-50



Understanding H.264 parameters

- Max number of B-Frames
  - Always activate B-frames. B-frames are more useful in static scenes. The max number of consecutive B-frames can be in the range 1-16. Some encoder does not allow a value >3. Recommended values in the range 1-3.



- B-frame “Auto decision”
  - If the encoder supports it,
    enable the “automatic decision”
    for the number of B-frames really used.
- B-pyramid
  - This option allows the encoder to use B-frames as reference if needed. Enable if available.

Understanding H.264 parameters

MAX
CONNECT. DISCOVER. INSPIRE.

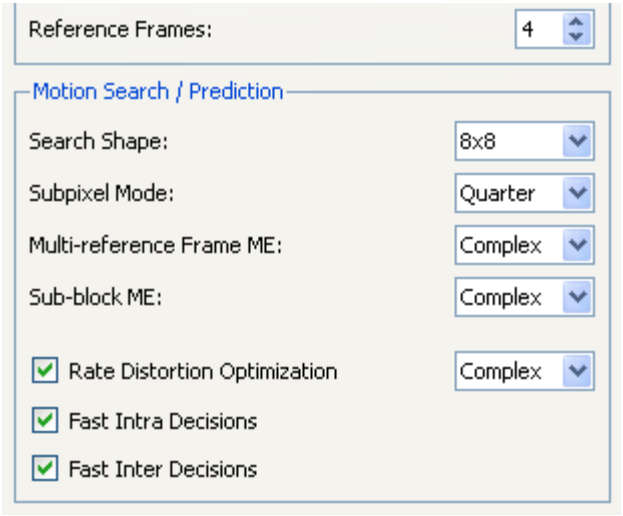
Entropy Coding

- CABAC
  - Context-Adaptive Binary Arithmetic Coding is always the best choice for quality sake.
- CAVLC
  - Requires less processing power but produce a lower quality/bitrate ratio.

Adobe

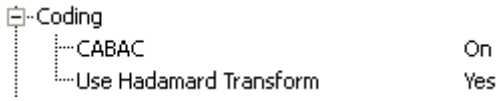
Understanding H.264 parameters

- **Motion estimation and compensation**
  these parameters can vary deeply in different encoders. Usually the more complex modes provide higher quality but slower encoding time.
- **Search modes**
  the best search mode in motion estimation is the "exhaustive" one, but it's too slow. Hexagonal search strategies are usually good and fast.
- **Search range**
  the terminology varies from encoder to encoder.
- **Search accuracy**
  set quarter-pel accuracy for final encoding.
- **Number of reference frames**
  H.264 supports up to 16 reference frames. Recommended 2-5.

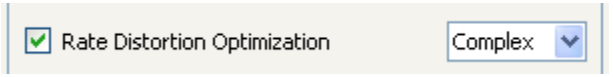


Understanding H.264 parameters

- Intra frame prediction
  Always enable Hadamard transform



- Rate Distortion Optimization
  Optimizes estimation choices for quality/bitrate
  Can be very slow and save only a few % of quality and/or bits



- De-blocking
  The complex in-loop de-blocking filter is one of the major cause of the efficiency of H.264
  Never disable de-blocking. I suggest to use the standard values for Alpha and Beta parameters which control the threshold and the strength of the de-blocking filter. But if you like a more "crisp" look try -1,-1.

Encoding speed vs Quality

- Enabling the most advanced features of H.264 consistently raises the encoding time. Depending by the specific encoder and by resolutions it may mean a x2-x4 time factor (i.e. 3 hours for encoding 1 hour of HD content on quad-core).
- Which features can be disabled to speed up encoding without affecting too much quality?
  - Reduce **reference frames** number to 2
  - Disable **b-pyramid**
  - Reduce **search range** (i.e.16, or simple) and **search mode** (bigger macroblock size)
  - Reduce **search accuracy** to half-pixel
  - Disable **rate distortion optimizations** (RDO)
  - Consider reducing resolution and pre-filtering

H.264 vs VP6

MAX CONNECT. DISCOVER. INSPIRE.

- With the knowledge and a good encoder, H.264 assures a better quality/bitrate ratio than VP6.
- However Vp6 can be useful because requires less processing power for decoding on single core processor (Athlon, Pentium4, Centrino).
- Vp6-S can help to deliver HD streaming with low processing power. On the other hand Vp6-S requires an higher bitrate to achieve the same quality.

- My suggestion:
  H.264 can be viewed by 90% of the audience today. Use Vp6 in a fall-back strategy to extend the audience to 99% and enhance the user experience on low-end computers.

Adobe

Encoding Best Practices

MAX
CONNECT. DISCOVER. INSPIRE.

How to maximize the “video-experience” :

- **1. Choose the best resolution-bandwidth mix for your video**
- **2. Pre-process source video**
- **3. Optimize video playback in the player**
- **4. Use multi-bitrate and FMS**

Adobe

Encoding Best Practices

MAX CONNECT. DISCOVER. INSPIRE.

Choose the best resolution-bandwidth mix for your video

- The level of motion in video determines the level of compression that H.264 can achieve. A very high motion clip can require 2-3x the bitrate of a static clip. Depending by the type of video you have to encode, you can choose a different resolution-bandwidth mix.

- Static Resolutions–Bandwidth mix examples:
  - **1080p:** Full HD (1920x1080) *generic* video may require 2-3 Mbit/s
  - **720p:** HD (1280x720) *generic* video may require 1.5-2 Mbit/s
  - **576p:** HQ (1024x576) *generic* video may require 1-1.2 Mbit/s
  - **480p:** SD (848x480) *generic* video may require 0.8-1 Mbit/s
  - **360p:** MD (640x360) *generic* video may require 0.6-0.8 Mbit/s

Adobe

Encoding Best Practices

MAX CONNECT. DISCOVER. INSPIRE.

Choose the best resolution-bandwidth mix for your video

- Content Adaptive resolution-bandwidth Mix
  - Med motion (tv series, news):
    target bitrate and resolution
  - Low motion (talking head, interview, meetings):
    lower bitrate or higher resolution
  - High motion (sports, music clips, action movies):
    higher bitrate or lower resolution

Adobe

Choose the best resolution-bandwidth mix for your video

- How to reduce the bandwidth for high-motion clips?
  - **Encoding in anamorphic resolutions can save 20-25% of bandwidth**.

    For 1080p you can encode the video at 1440x1080 or 1280x1080 and then interpolate at the original resolution in the Flash Player.

    For 720p you can encode at 1024x720. Anamorphic videos remain HD compliant
  - **Note: The perceived loss in quality caused by the use of a lower resolution is minor than the perceived loss in quality caused by a too much high quantization.**

Encoding Best Practices

MAX CONNECT. DISCOVER. INSPIRE.

Pre-process video source

- Video Encoding has two further enemies: **video noise** and **interlacing**.
- **Video noise** (very frequently present in HD footage) lowers the efficiency of encoding. It is very important to reduce video noise with proper filters. The best are the “temporal denoise filters” or “3D denoise filters”.

  Note: *resizing (bilinear or bicubic) to a lower output resolution acts as a denoising filter, this is why is always better to encode from a HD source than from a SD source even when the target resolution is SD*.

Adobe

Pre-process video source

- **Interlacing:** SD sources are very often interlaced, and 1080i sources are interlaced. It is very important to use the most professional deinterlacing routine (*motion compensated adaptive deinterlace routines*) to preserve frame resolution and detail because most de-int filters (wave, bob) simply cut a field or blend two fields producing a sensible loss in quality and detail.

  If you have only a simple deinterlacer, consider to encode the video at half the vertical resolution instead; This will produce almost the same quality with much lower bitrate. (i.e. a 720x576 interlaced source encoded at 720x288 and interpolated back in the player).

Optimize video playback in the Flash Player

- The optimum would be to exploit hardware acceleration because hardware scaling is very good for quality and performance. This can be done at full-screen (FP9) or directly in the browser (FP10). The software based scaling in FP9 is very good too, but slower.
- Use **video.smoothing=true** when video resolution is different from player resolution (and in our anamorphic proposition). Disable it when you go full-screen to exploit the faster hardware scaling.
- If you don't care about player interface distortion, go full-screen using :

```
Stage["fullScreenSourceRect"] = new Rectangle(0, 0, Stage.width, Stage.height);
Stage["displayState"] = "fullScreen";
```

Optimize video playback in the Flash Player

- If you are encoding at very low bitrate, try experimenting with “details restoring filters”. You can use the standard filter object of Flash Player 8 or the most advanced Pixel Bender. A common “sharpen filter” can restore details lost during the encoding. Note: *This is only a “perceptual” restore and is CPU intensive*.

Multi-bitrate and FMS

MAX CONNECT. DISCOVER. INSPIRE.

How to improve user experience and QoS ?

- The answer is : **Dynamic bitrate switching**
  With Dynamic bitrate switching is possible to react to bandwidth fluctuation changing on the fly the bitrate that are streaming.

  It is very difficult to achieve that in Flash Player 9 + FMS3* and many implementation simply choose the right bitrate at the beginning of the stream using the native FMS3 bandwidth test.

  Flash Player 10 and the next update of FMS (FMS 3.5) support a native mode, called **Dynamic Streaming**. A new specific API is created for monitoring the QoS of the stream and choose when to switch to a different bitrate. The switch is performed in a seamless way.

  * = mail me if you are interested in such technology.

Adobe

Optimizing encoding for Multi-bitrate

- The new dynamic streaming feature of Flash Player 10 + FMS 3.5 requires specific attention in encoding to optimize the switching performance.
  - Encode the sound track with the same parameter for each bitrate
  - Encode preferably at CBR and with a fixed distance between IDRs (i.e. 100 = 4sec)
  - Use the lower bitrate stream as a fall-back for low-end PCs
  - Do not use too much bitrates. Balanced mix:

    - 1280x720 @ 1.5Mbit/s
    - 1024x576 @ 1.1Mbit/s
    - 800x480 @ 800Kbit/s
    - 640x360 @ 600Kbit/s

DEMO