White Paper: Digital Video – Encoder & Decoder Overview

Bill Reckwerdt, CTO
Video Clarity, Inc.

Simple Definition of Digital Video

Using the simplest definition, digital video is the representation or encoding of an analog video signals in digital bits for storage, transmission, and/or display. If you have rented a DVD, watched digital cable, DirecTV, or Dish, played a video game, then you have experienced digital video.

The Encoding Process

An important component in digital video is the pixel, or picture element, which represents the color in bits. The color is a blend of red, green, and blue and is represented in bits. The number of bits is termed the bit-depth (and is usually 8, 10, or 16-bits per component). More bits allow a more precise representation of hue from the red, green and blue colors. Graphics are usually represented in RGB (red-green-blue) format, and TV video is usually represented as Y'CbCr, where Y' is the luminance (or brightness) and CbCr represent the color (pure color with no brightness).

Video is basically a three-dimensional array of color pixels. Two dimensions serve as spatial (horizontal and vertical) directions, and one dimension represents the temporal (time) domain. A frame is a set of all pixels that correspond to a single point in time. Basically, a frame is the same as a still picture.

The number of pixels defines the spatial resolution. Standard television is displayed with 720x480 at 30Hz (576 at 25Hz for PAL) resolution. High Definition television is usually defined as 1280x720 at 60Hz (720P) or 1920x1080 at 30Hz (1080i). This means that standard definition requires 720*480*30fps*24color-bits = 250Mb/s. HDTV is 6 times more data. Thus, the data must be compressed.

Video sequences contain spatial and temporal redundancy. Similarities can thus be encoded by merely registering differences within a frame (spatial) and/or between frames (temporal). Spatial encoding is performed by taking advantage of the fact that the human eye is unable to distinguish small differences in color as easily as it can changes in brightness and so very similar areas of color can be "averaged out". With temporal compression only the changes from one frame to the next are encoded as often a large number of the pixels will be the same on a series of frames (About video compression).

The Decoding (Displaying) Process

Once the video is created, stored, and transmitted, a computer process must open the video and display the original video. This is termed a video decoder. It reads the encoded video file, decompresses and displays it.

The video data must be played in the correct order, with little or no packet loss, and with smooth, continuous timing or essential information will be missing. To ensure video quality, networking companies must provide a good transport mechanism, encoders must produce good picture quality, and the decoder must do a good just of displaying and compensating for errors.

Video Encoding Standards

MPEG-1 (ISO/IEC 11172)
The first digital video encoding standard was developed by the Moving Pictures Experts Groups and was termed MPEG-1. It was adopted in 1992 and provided VHS quality digital video for CD-ROM playback.

MPEG-1 employs intra-frame spatial compression on redundant color values using discrete cosine transforms (DCTs). The DCT is then further reduced by quantizing (basically reducing the scale) and converting 4:4:4 RGB data to 4:2:0 Y'CbCr, which reduces the amount of color information from 24-bits to 12-bits.

MPEG-1 relies on prediction, or more precisely motion-compensated prediction, for temporal compression between frames. It uses 3 frames to create temporal compression:
- I frames
- B frames
- P frames

An I-frame has no reference to the past or future frames. P-frames are forward predicted frames with reference to previous I or P frames. B-frames are encoded with reference

to previous and future I, P, or B frames. The smaller number of I frames compared to P and B frames reduces bit-rate even further.

MPEG-2 (ISO/IEC 13818-2)
MPEG-2 was designed to encompass, and to be backward compatible with MPEG-1. It includes support for interlaced video for broadcast TV.
A television broadcast frame is created with two separate fields, a top and bottom interlaced field, with the first line of the bottom field appearing immediately after the first line of the top field. Thus, 30 frames-per-second is actually sent as 60 fields-per-second.

In addition, MPEG-2 includes improved color-sub-sampling, error correction, and improved audio support.

MPEG-3
MPEG-3 was designed for high definition television (HDTV), but the MPEG-2 standard scaled to encompass the requirements so this initiative was withdrawn.

MPEG-4 (ISO/IEC 14496)
MPEG-4 arose from the need to have scalable support for low bit rate applications – streaming over the Internet – and the need for better compression at higher resolutions.

MPEG-4 includes support for object-oriented encoding – combining film, graphics, text, audio, and animation together – and digital rights management. Unfortunately, the standard has become quite confusing due to the high number of profiles.

MPEG-4 was designed with the hope to achieve a 50% reduction in bandwidth compared to MPEG-2, and to introduce object oriented solutions. It achieved the second part, but the reduction in bandwidth was only about 30%. To achieve the bandwidth goals, ISO combined with ITU-T H.26x to produce a new video standard, which is known as MPEG-4 part 10, H.264, and/or Advanced Video Coding (AVC). This new video standard is not compatible with the original MPEG-4 video standard, but uses the same system parameters to achieve objected oriented solutions and digital rights management. MPEG-4 part 10 is computational more complicated when compared to the original MPEG-4 (now known as MPEG-4 part 2 or MPEG-4 ASP) so advanced computer systems are needed to encode and decode this new format.

MPEG-4 is used by many satellite providers for HDTV, is one of the standards for HD DVDs, and is used for Video-on-Demand.

VC-1 (SMPTE 410M)
VC-1 originally developed by Microsoft and also known as Windows Media Video 3 or Windows Media CODEC (enCOder / DECoder) 9 is an evolution of the standard MPEG-1, -2 DCT based algorithms. It is an alternative to MPEG-4 part 10, and is used in similar applications.

VC-1 widely assumed to be a Microsoft product is actually owned by a pool of 15 companies, and was accepted in April of 2006 as a SMPTE standard.

Motion JPEG 2000
Motion JPEG 2000 (often referenced as MJ2 or MJP2) is the leading digital film standard currently supported by Digital Cinema Initiatives (a consortium of most major studios and vendors) for the storage, distribution and exhibition of motion pictures. It is an open ISO standard and an advanced update to MJPEG (or MJ), which was based on the legacy JPEG format. Unlike MPEG-2, MPEG-4 and VC-1, MJ2 does not employ temporal or inter-frame compression. Instead, each frame is an independent entity encoded by either a lossy or lossless variant of JPEG 2000. JPEG 2000 uses a wavelet approach to compression as opposed to a DCT.

One legal danger, with respect to MJ2, is that it does not have a clear license. The JPEG committee is working on a royalty free usage model, but it is not in place, yet.

What does this all mean?
Objective Video Quality Metrics that look for blockiness, tiling, spatial, and temporal effects need to understand that different coding techniques create different artifacts. Moreover, many broadcasters (or more generally, encoding / decoding customers) would like to

compare different coding techniques before making a buy decision.

To this end, Video Clarity defined ClearView to measure video quality in the uncompressed domain, and to allow multiple viewing modes of uncompressed video. You can measure the video quality of any video sequence regardless of coding technique and view video sequences encoded using different coding techniques.

Figure 1: Video Testing for Video Processing and STB Manufacturers and their Customers

Figure 2: Multiple Display Modes with Objective Scoring
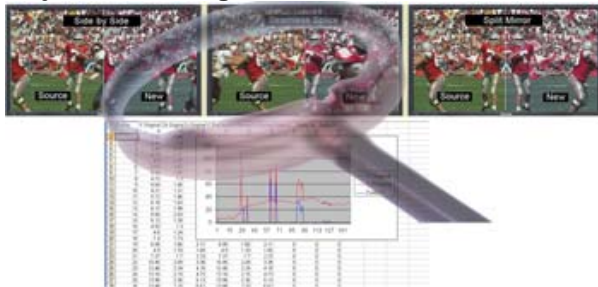
The Author

Bill Reckwerdt has been involved in digital video since the early 90's from digital compression, video on demand, to streaming servers. He received his MS specializing in Behavioral Modeling and Design Automation from the University of Illinois Urbana-Champaign.

He is currently the VP of Marketing and the CTO for Video Clarity, which makes quantitative, repeatable video quality testing tools. For more information about Video Clarity, please visit their website at http://www.videoclarity.com.