Video Technologies for the Web

Scott Bondurant Chandler Tom Hergert

Abstract

Visual information is moving to the Internet at a rapid pace, often making use of cutting-edge technologies such as streaming video. An understanding of competing technologies, bandwidths, and compression strategies is essential to creating web video in streaming, fast start, downloadable, LAN-based, and CD-ROM hybrid formats.

Overview

The creation and use of web video occurs in four discrete stages: creation of content, capture and compression, transmission, and reception. Each step requires a combination of art and science. By maximizing quality and understanding the available choices at each step, the best possible web video may be produced.

Comparing the common bandwidths

The historic television standard is analog based. Is is an inherently imprecise, continuously variable signal that can be displayed as an approximation of the visual image as recorded. It does not translate precisely into either the computer or digital applicance world. Much of the lexicon and common wisdom on broadcast television and video production must be modified in order to be applicable in a multimedia world. However, it is clear that the new video standards share as much with computer technologies as with their analog roots. Convergence of technology is no longer a hypothetical discussion, it is reality.

It is quite remarkable that video can be repurposed for multimedia applications. For example, the digital signal used in set-top DVD players is compressed 24 times the theoretical file size of "full resolution" quantization of analog. Although DVD technology does not match the quality and control found in most professional video, DVD playback does exceed the quality most consumers receive and record today.

The modern standard for CD-ROM playback is compressed even more dramatically. A common QuickTime data stream is compressed to 1/300th the size of the original signal. This is still modest compression compared to the compression applied to streaming video on the web. In order to provide real time playback, a web video stream must be compressed to 1/6,000th of the original bandwidth.

The "Trailer Effect"

The largest web video event to date was the release of the motion picture trailer for Star Wars Episode One: The Phantom Menace. Over 25 million people viewed or downloaded the preview. The vast majority of those users downloaded the highest bandwidth version. This version used CD-ROM quality bandwidth, resulting in a common download time of 25 times the clip's running time. The two-minute twenty-seven second clip took most users one hour and twelve minutes to download, with some download times exceeding two and one-half hours. We should consider these factors carefully. Clearly there are cases in which users have a strong preference for high quality even when wait times are substantial.

Content

Initial Quality (garbage in, garbage out)

The quality of the initial video can dramatically affect the outcome in the creation of web video. Having a pristine, technically clean image is critical to good compression and to maintaining maximum quality. Even though web video is typically viewed as a small image and the data stream is tremendously compressed, visual problems known as artifacts can be created/perpetuated during the process, polluting the final image. Ben Waggoner notes that, "Web video is surprisingly good at preserving composite noise throughout the compression process. Garbage in, Garbage out." (1999). Use the best original recording format possible. Digital Betacam and Betacam SP are standards, although the new DV format's purely digital stream can be a more cost effective alternative.

Shooting for the best image

Video intended for interactive multimedia or the Internet should be shot in a very different way than broadcast video. Because broadcast video does not use compression in the conventional sense, dramatic camera moves and zooms are common. These types of changes are not ideal for use on CD-ROM or for the web. Keep camera movements and subject movements to a minimum in order to maximize compression later in the process. The easiest video to compress is a "talking head" where a seated, single subject is talking. Use a tripod to reduce camera movement.

Web video clips should be short and succinct. Only the most critical aspects of a presentation or educational activity should be presented as video. Most educational and presentation objectives can be met using other, less bandwidth intensive, media. Video can add personality to an online experience and provide an invaluable method to convey certain difficult tasks. However, video should not be the de facto presentation method for your objective. Web users are concerned with immediacy and therefore with bandwidth. It is methodologically better to offer a series of short clips when needed as opposed to larger segments or complete lessons. As bandwidth improves, video will become less difficult to download; but for the near future video should be used judiciously.

Progressive vs. interlaced

Conventional video is created and displayed as an interlaced image, while computer visuals are created and viewed as progressively scanned (noninterlaced) images. If a given piece of video content will only be used as a compressed digital image and never played back as conventional video, shooting or creating clips and sequences as progressive scanned imagery may be a useful option.

Interlaced video images are drawn to the screen in a 30 (29.97) frame per second system as two 1/60 second images or fields. Field A consists of all the odd numbered horizontal lines in the image (typically 1-479), and field B is all the even numbered lines (2-480). This results in a temporal offset of the two fields which can result in a blurred or even vibrating image when both fields are displayed together as a single still image. One common precaution against this unwanted effect is "line doubling," wherein either the odd field (A) or the even field (B) is repeated, replacing its normal counterpart with itself, offset one line. This creates an image with an effective vertical resolution one half of what is available. Line doubling in itself violates the maxim that one should start with the best quality video available.

Some modern cameras and recording systems allow for progressive scanning of scenes during photography. In these systems, the image is recorded as single frame with the raster scanning the screen once, top to bottom. There are progressive scan systems that record 30 and/or 60 frames per second (fps), with some systems even allowing for slower frame rates such as 15 or 8 fps. Progressive scanning is believed to have an image quality that resembles film's 24 fps look. It will almost certainly be the basis of Advanced Television Systems such as HDTV. For the purposes of our discussion here, the value of progressive scan comes in the relative ease with which such images can be compressed. For compression of interlaced images, the first step is often to get rid of the interlacing, either by line doubling or by a mathematical extrapolation of one half of the image's horizontal lines. Either of these requires time and processing activity and degrades the image. Progressively scanned noninterlaced images are exempt from these problems.

Cropping and Playback Size

Traditional video is viewed on larger screens. Web video, in contrast, is usually viewed in a window at a fraction of the normal size. Most CD-ROM playback is created at 25% size, which equals 320 pixels wide by 240 pixels tall. Web video is generally even smaller in the range of 160 pixels by 120 pixels. A larger physical size (more pixels) is harder to compress and longer to download. It is often advisable to crop the image more dramatically than in the broadcast arena. Television screens typically show only the centermost 85% of the broadcast image called the "action safe" area. The unseen edges of the picture often contain noise or unnecessary imagery. Because web video does not bleed off the edges of the screen, this area should be cropped. Further, the small physical size means that cropping images much tighter may produce a more pleasing image. It can be effective to crop images even more dramatically. What would play well as a medium shot on a television screen is more powerful in a compressed window as a close up. Most compression algorithms allow almost any shape as long as the horizontal and vertical dimensions are evenly divisible by four.

Does video enhance content?

Because motion video is so bandwidth intensive, the decision to employ it in a given situation must be carefully considered. Motion pictures are a very common frame of reference for many cultures, so the temptation to always use video is strong. That couples with the impression that web video is "cool" and technologically impressive to push many clients, designers, and users to put web video into every project. Just as some messages in conventional media are better conveyed through print, still photography, audio, or some combination of those; web projects can make effective use of a variety of media. Motion video is an extremely complex and bandwidth intensive medium if images of reasonable quality are important.

Text in highly compressed video can be especially challenging. Video clips that are repurposed from existing productions, and some video created for web use, often include text information that becomes unintelligible once the video has been compressed. Detailed pictorial or diagrammatic information is almost always clearer in a still format instead of compressed video.

The audio portion of the compressed video signal uses a small fraction of the total bandwidth. A dedicated audio file, streaming or downloadable, will almost guarantee superior fidelity over the audio track of a video clip.

If the motion of a clip's subject matter is an important element, then compressed video may be the best medium. A dancer's fluid movements or the excitement of a horse race or other sporting event may not be perfectly rendered through the web, but without motion they could not be conveyed well. The talking head example cited earlier is especially frustrating in that although it compresses easily and well, there may be no need to have it at all. Streaming audio accompanied by a series of appropriate stills and illustrations might deliver the message more clearly than even the best compressed talking head. A question traditionally asked of clients with a message who approach informational film or video producers is, “Why do you need/want a motion picture production?” Careful consideration of this question has cost producers some business, but it has also prevented headaches and fiascos. Compressed video for web distribution warrants the same consideration, especially because alternative media can be delivered easily and sometimes more effectively via the same distribution system.

Capture and Compression

Capturing the video

Because of the immense disparity between the bandwidth of traditional analog and digital web signals, we must radically change the way the signal is handled, stored and sent.

There are a variety of ways to convert data from the playback device (a camera or VCR) into a signal that is usable in the digital realm. For highest quality, data can be digitized or captured from high-end playback devices using proprietary video editing workstations. Strong results can be obtained from less expensive digital players (DV format) using digital capture; poorer results will occur when analog signals are sampled into low-end AV capture systems.

YUV component inputs connect a Betacam SP player into a professional nonlinear editing system for digitizing. Serial Digital Interface connections used with the high end digital formats permit the cleanest, completely digital pathway when they are available.

With a DV bit stream format, a Firewire capture system maintains the digital file structure and its attendant quality. Also known as IEEE 1394 or Sony iLink, Firewire allows for hard drives, digital cameras and VCRs to be accessed at high speed and with digital data integrity. In the last 12 months, Firewire in and out has become available on many video cameras and is now found on many computers.

For analog formats, the video and audio signals leave many playback devices using RCA jacks (video, left audio, right audio) or the higher quality S-video cable (video only, still requires RCA cables for audio). The computer’s video card then handles the conversion from analog to digital.

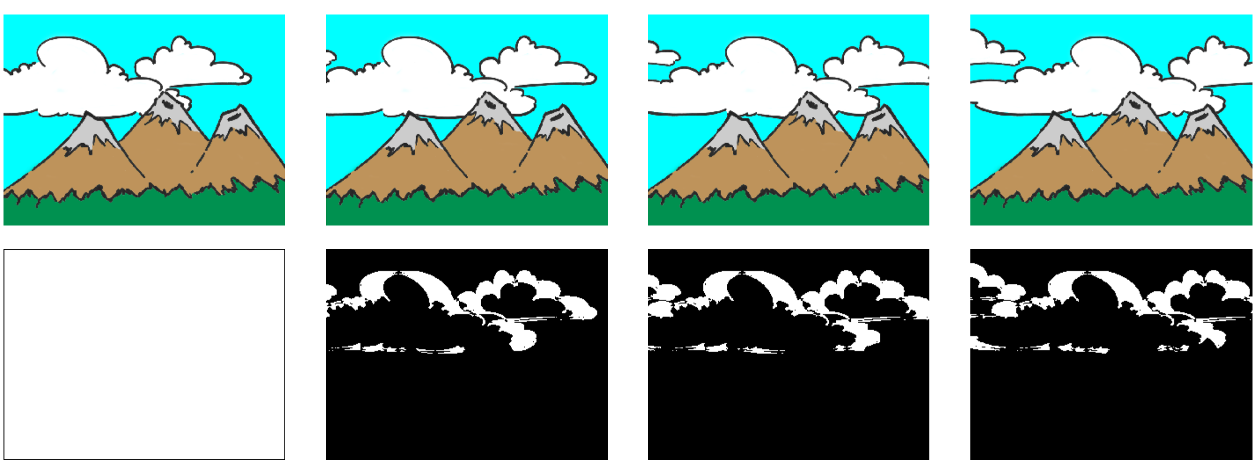
This capture phase often involves some level of compression that is not nearly enough to reach small file sizes or meet minimal bandwidth limitations. After the initial capture or digitization stage, video files must be processed further to make them usable in a web context.

Temporal compression

Video can be thought of as a sequence of still images presented at 29.97 frames per second. Many of these individual frames share the identical image data.

In Figure 1 the mountains are completely consistent across the animation. The clouds and sky are the only objects “in motion.” As such, it isn’t necessary to individually enumerate the pixels describing the mountains for each frame. This consistency of image across time makes temporal compression possible. The top row of Figure 1 shows a few frames of the animation. Note that the clouds move from left to right behind the mountains and in front of the sky. The lower row of images shows the temporal compression that takes place on the animation. White indicates pixels that have changed and need to be updated while black indicates pixels which are identical and therefore do not need to be

Figure 1
Example Of Temporal Compression



sent again. The first frame is a keyframe and includes the complete picture. In later frames, however, only small areas of the image change. Essentially, the edges of the clouds must be erased.

When video footage is compressed temporally, in most cases minimal instructions for the changes between frames are sent. These partial frames are called delta frames. At regular intervals, a complete frame (keyframe or “i” frame) is sent as a reference point for the frames around it. Keyframes are frames that include all of the information without reference to previous frames. This allows quality to be relatively consistent and for users to scrub (move freely through a movie timeline). Compression over time can make a huge difference in bandwidth with no or modest degradation in quality as playback progresses. Some compression algorithms require more keyframes to deliver acceptable image quality.

Spatial compression

In addition to being temporally compressed, most web video is spatially compressed. Spatial compression occurs when the image data is reorganized to make it easier to send. Spatial compression may be lossless or lossy.

Lossless compression looks for patterns in each image that can be repeated exactly or conveyed as instructions rather than images. Despite being reorganized, an image that has been compressed with lossless compression will match the original image exactly. The “animation” compression scheme often used for the development of special effects is an example of lossless spatial compression.

Lossy data compression works in a different way. Because lossy compression doesn’t attempt to recreate the image exactly, it can create much smaller files. It would be unfair to suggest that lossy compression is unattractive. Lossy data compression can be of exceptionally high quality.

Almost all of the expensive, proprietary nonlinear editing systems use spatial compression. In fact, the DV format is compressed 5:1, while DVD-video is compressed approximately 14:1. Some high end edit systems allow users to choose their compression ratios from as low as 2:1 to as high as 120:1, depending on their situational needs to balance hard drive space and playback bandwidth with image quality considerations.

Figure 2
Corporate Logo At Varying Amounts Of Compression

10 quality (100 max) JPEG, 16 k size — 50 quality (100 max) JPEG, 20 k size — uncompressed TIFF, 28 k size

To reach the extreme compression ratios of web video, extreme spatial compression is needed. All of the web compression techniques degrade the image to the point where the lossy nature of this type of compression is visible. The effect is especially noticeable as artifacts around the edges of objects. See Figure 2 for examples of spatial compression.

CODEC choice

Within each video architecture, multiple CODECs are available. CODECs are different systems that can be used to COmpress and DECompress video or audio. In general, video is compressed using one CODEC and audio is compressed using another CODEC.

Each CODEC has its own advantages and disadvantages. Under QuickTime, Sorenson has become the most common CODEC for video compression at low, medium and CD-ROM bandwidths. Sorenson does an excellent job with live action video and scenes that contain motion. Sorenson is available in version 1.0 (included with QuickTime 3.0), version 2.0 (included with QuickTime 4.0) and higher quality professional versions. All users with QuickTime 3.0 or newer (74% as of this writing, Statmarket, 1999) can view Sorenson compressed clips. In order to create Sorenson compressed clips—QuickTime Pro, MoviePlayer 2.5, MediaCleaner Pro or another video editing tool is needed. The commercial application MediaCleaner Pro is the industry leader for generating compression video destined for CD-ROM or the Internet.

To compress audio tracks for use on CD-ROM and the web there are two excellent choices. The Q-Design CODEC does a very nice job with music and voice at low and medium bandwidths (audio rarely needs to be compressed at high bandwidths). When voice is the primary element, the Qualcomm PureVoice system offers good quality at extremely low bandwidth.

Another common web compression choice is cinepak. Cinepak compression is one of the older compression schemes. Many webmasters use cinepak because it can offer high-quality video “slideshows.” If the image needs to change only once every few seconds, cinepak compression should be considered. Cinepak is also ideal for unusual sizes. Because of the inherent flexibility of the format, it deals well with extreme aspect ratios and extremely large windows. For instance, cinepak is the only format which can maintain high frame rates at 800 pixels wide by 600 pixels tall.

Another common architecture is available for video delivered via the web. RealMedia offers capabilities similar to those of QuickTime, although it can be more expensive to set up dedicated servers. RealMedia also has a smaller installed user base with 24% of users employing the newest version G2 (Statmarket, 1999).

It is often prudent not to use the latest CODEC versions. For instance, the Sorenson version 1 playback is enabled on more than four times as many browsers as Sorenson version 2. The latest CODECs also require significantly more processor time than earlier versions. When developing multimedia, it is common to set a "floor machine." This floor machine represents the minimum hardware level for which playback must be acceptable. For many applications, a low end machine will not be able to uncompress the latest CODECs with acceptable results.

In many cases, the institution hosting the web video will specify the appropriate formats. For instance, many colleges already have powerful Real servers in place. Sometimes pragmatism determines the compression of choice.

Transmission

The Pipe!

All other things being equal, the size of the data stream is the best indicator of web video quality. The stream of information (data stream) in any digital system is often compared to flowing water. Just as water flow is constricted by the size of the pipe, hose or straw through which it flows, data can be restricted by attributes of the pathways it must follow. The size of the data "pipes" that viewers have can vary radically. Most users are using 56 kbps (1,024 bits per second) modems to connect, although connections commonly run the gamut from half this speed to 50 times this speed.

Options exist to support users at different bandwidths within this range. In fact, it has become common to support users at 28.8 kbps (3.6 k/sec), 56 kbps (7 k/sec), dual ISDN (14 k/sec), medium compression levels (30 k/sec) and CD-ROM quality (105-180 k/sec). Commercial compression software can generate multiple versions of video files automatically for all possible bandwidths. With earlier versions of RealMedia, users were generally given a choice of which speed to access. Newer, commercial versions of RealMedia are capable of auto-detecting the connection speed of a particular connection and also adjusting to network conditions. QuickTime developers generally take a different approach: users select their connection speed globally. The QuickTime system is then able to detect that choice and send the appropriate quality clip. One advantage of this system is that users may select a bandwidth larger than they actually use; this allows them to access higher quality clips with additional download time.

Most networks in use today are packet based. The network breaks messages into a large number of pieces and transmits each piece individually. Packets are regularly lost or arrive in the wrong order. Most network communication can simply request lost information again, but streaming media is especially unforgiving. Isochronous networks that can reserve bandwidth for time sensitive applications are becoming more popular. Asynchronous Transfer Mode (ATM) is a high bandwidth network which reserves part of the spectrum for certain uses. ATM is widely used for video conferencing at speeds 180 times faster than web video. However, dedicated data paths will become more common. It is probable that data pipes not only will become larger but smarter over time.

Web Video Delivery Methods

Web video is usually delivered in one of four ways: streaming, fast start, download, via local area network or hybrid.

Streaming video refers to video clips that are provided to the web in near real time. Feeding live video into a server, which translates that signal into packet streams for immediate playback, creates these "live" clips. When watching a live clip, there is no choice to back up or jump forward. Streaming users simply watch the show at the bandwidth available to them. Streaming content is often of poor quality because of the number of simultaneous users (synchronous access) and because of the compression symmetry. Streaming video must be compressed into multiple bandwidths in real time. This means that the CODEC must generate the best results possible each fraction of a second. Compare this to asymmetrical compression. Asymmetrical compression can take as much time as needed to create the best image. The benefits of immediacy are dubious. In many applications, users would be better served by higher quality instead of watching an event "live." Streaming has historically been RealMedia's strength, although QuickTime now offers similar streaming capabilities.

Another option is to create fast start clips. These segments do not begin to play when downloading starts. The clips are cached in a buffer until the length of the clip exceeds the remaining download time. Once the machine believes that it can stay ahead of playback, it will begin to play the clip. Fast start clips are usually of substantially higher quality than streaming quality. In exchange for "live" playback, higher quality is obtained. Most QuickTime media is provided in a fast start format at the bandwidth (quality) requested by the user.

Users also can be allowed to download content. Downloadable movies can be acquired now and watched later. Downloadable content is usually used for large, high quality pieces. Clips that have been downloaded can be watched repeatedly without maintaining a network connection. Although downloadable video is being displaced by other formats, it represents a powerful alternative and conserves bandwidth if users are likely to view content more than once.

If the audience viewing your web video is well known, Internet delivery may not be your best option. CD-ROM quality is significantly higher than is achievable with most connections. Although web video may appear clean and smooth to the developers, remote users on slow connections will suffer. Consider providing users higher quality video on CD-ROM or CD-R (CD recordable) media as an alternative. The cost per user is generally a few dollars. If the purpose of web video is to communicate complex topics, isn't it worth the effort to maximize that quality? Technologies are readily available to combine web sites and video on CD with Internet service to create a robust hybrid experience.

Users on local area networks (LAN) can view streaming, fast start and downloadable content. These users are unique because their connection speeds are up to 1,500 times faster than modem users. Video on a LAN can be exceedingly high quality and even match CD-ROM quality.

Reception

The last step in the process is for the video to be received and played to the user. As discussed, not all users have the speed and power to play back web video compressed with the most advanced techniques.

There are several playback choices depending on the way the files are created and the way the client machine is configured. Historically, once a web video file or its proxy file was downloaded, the playback would be handled by a separate helper application. Under QuickTime, playback and basic video editing is provided by MoviePlayer 2.5.1. However, compression and save capability have been removed from later versions of MoviePlayer unless the user upgrades to QuickTime Pro. Many clients still access files through helper applications although these statistics are difficult to track.

With the addition of a browser plug-in, users can watch web video within their web browser and even within a web page. This allows for web video to be part of a larger, richer experience for the user. For educators, incorporating video directly into an instructional module can be especially powerful.

In fact, web video continues to become better integrated into the multimedia landscape. Both QuickTime and RealMedia now offer SMIL (pronounced smile) for temporal synchronicity and Flash to provide professional grade interactive environments inside web video.

Summary

There are a variety of lessons to be learned when dealing with web video. Remember that web video is massively compressed. Figure 3 demonstrates by cropping a photograph the magnitude of this compression. The image has been cropped to convey the raw ratios (by

Figure 3
Compression Ratios Expressed As Area

area) of compression for live web video (top), CD-ROM video, DVD-Video and as a full frame (bottom, not cropped). Although web video is actually compressed in terms of physical size, frame rate, temporally and spatially—the magnitude of the compression can be seen in these examples.

Provide web video in a variety of bandwidths. As connection speeds improve, users will migrate to the higher levels. It is also wise to archive source footage. Content may be repurposed under newer compression schemes or at higher speeds later.

Let the user choose which bandwidth to download. The trailer effect suggests that users may be predisposed to quality even when download time is longer. Empower the user to make a choice on the web page or use technology that automatically detects the user's preference.

Use professional tools when possible. The professional tools are not only more powerful but also create a distinct look and feel in the product. There is no replacement for video professionals and their experiences. Don't rely on inferior practitioners or inferior applications. The industry standard MediaCleaner Pro, for instance, will save countless hours compressing and recompressing clips at multiple resolutions.

A variety of resources are available to help in the creation of web video. Terran Interactive, the maker of MediaCleaner Pro, has a comprehensive web site at http://www.codeccentral.com/ and offer an excellent, free tutorial called "How to produce great QuickTime." Real offers a guide to their specific products called "Real encoding guide" at http://www.real.com/.

Remember, web video will improve. During the last year—we've seen a maturation of compression schemes, better compression tools, acceptance of digital video on the desktop and standardization on slightly higher connection speeds. No doubt this process will continue to improve in quality and support. It will never be worse than it is today and will almost certainly exceed our quality expectations in the future.

REFERENCES:

Waggoner, B. (1999). Making great web video. *DV Magazine*. 7(10), 46-56. October 1999.

WebSideStory (1999). Web Browsers: Top Plug-ins. *StatMarket*. [On-line]. URL: http://www.statmarket.com/SM?c=Plugins

Terran Interactive (1999). *How to produce great QuickTime*. [On-line]. URL: http://www.terran.com/QuickTime/Article/Welcome.html