USE OF DIGITAL VIDEO TECHNOLOGY FOR REAL-TIME EXERCISE MONITORING AND DEBRIEF OF COLLECTIVE TRAINING APPLICATIONS

George Gazzam
L-3 Communications Link Training and Simulation
Arlington, Texas

Sam Knight
L-3 Communications Link Training and Simulation
Orlando, Florida

ABSTRACT

As simulation evolves from single cockpit trainers to integrated multiple cockpit configurations, the need to provide real-time exercise monitoring and debrief capabilities becomes critical to a complete training environment. This paper examines the current state of the art in video compression and storage to solve problems related to monitoring an on-going training session in real-time and provide playback capabilities for debrief. A case study of the Sensor Video Recording System (SVRS) developed by L-3 Communications Link Simulation and Training in the U.S. Army s Aviation Combat Arms Tactical Trainer — Aviation Reconfigurable Manned Simulator(AVCATT-A) is discussed. The SVRS uses digital video compression and network streaming to solve monitoring/debrief requirements. The current resolution limitations, image artifacts, and storage requirements of video compression techniques are analyzed. Video regeneration is also examined as an alternative to video compression and storage. The limitations and latencies involved in network streaming are summarized along with lessons learned in developing a multichannel digital video system.

ABOUT THE AUTHORS

George W. Gazzam is a Principal Software Engineer with L-3 Communication Corporation, Link Simulation and Training Division, in Arlington, TX. He is currently developing the Sensor Video Recording System for the Training Environment component of the AVCATT-A training system. His experience covers a wide range of database tools and the development of integrated systems to support correlated visual, radar, and sensor simulation. Major development efforts include the Mission Support system for the Taiwan F-16 full mission trainer and the GeoTx image analysis tool for the B-2 simulator. His interests include digital image analysis, automated feature extraction, interactive editing environments, and the use of object oriented development methodologies. Mr. Gazzam holds a B.S. in Computer Science from the University of Pittsburgh, a M.S. in Computer Science from Binghamton University, Binghamton, NY, and is currently completing a M.B.A. from the University of Texas at Arlington.

Samuel N. Knight is a Manager of Engineering Programs for the Orlando component of the Link Simulation and Training Division of L-3 Communications. He has over 31 years of modeling and simulation engineering development experience including Army, Air Force, Navy and foreign programs. He is currently the Training Environment IPT Lead for AVCATT-A. His primary areas of expertise include tactical systems and simulations, tactical mission environments and semi-automated forces, visual and sensor simulations, Advanced Distributed Simulation (ADS), High Level Architecture (HLA), Distributed Interactive Simulation (DIS), team training, mission rehearsal requirements and supporting technologies, embedded training systems, integration techniques, testing and the use of simulators for systems and tactics evaluations. He is a committee member for the Simulation Interoperability Standards Organization (SISO) and is the President of SISO Inc. He was previously a Steering Committee member and Chairman of the Emissions Subgroup for the ARPA/STRICOM/UCF Working Group for the Interoperability of Defense Simulations. Mr. Knight has published works in the areas of tactical simulation, simulator networking, simulator fidelity, team training and mission rehearsal. Mr. Knight holds a BSEE degree from Clarkson University and a MEE degree from Cornell University.

USE OF DIGITAL VIDEO TECHNOLOGY FOR REAL-TIME EXERCISE MONITORING AND DEBRIEF OF COLLECTIVE TRAINING APPLICATIONS

George Gazzam
L-3 Communications Link Training and Simulation
Arlington, Texas

Sam Knight
L-3 Communications Link Training and Simulation
Orlando, Florida

INTRODUCTION

Multi-channel video recording and playback has long been a simulation challenge. The most cost effective method in the past was to use a cluster of VHS or other analog type video recorders. This approach became costly as simulators emerged with more video channels. Multiple channels required the use of time sync generators and required users to interrupt training activities frequently to replace the system s video cassettes both during the mission and replay. In addition, access to desired points in the recorded mission for analysis or replay was cumbersome for users. Fast forward features existed but the imagery was not smooth and clear in this mode.

Emerging digital video technology has recently matured and become cost effective to the point that it can be applied to address challenging simulation requirements. This paper discusses the current state-of-the-art in video compression and storage and describes the application of today s technology to the Sensor Video Recording System (SVRS) developed for the U.S. Army s Aviation Combat Arms Tactical Trainer — Aviation Reconfigurable Manned Simulator (AVCATT-A)

DESIGN APPROACHES

In any simulation system, there is a host computer and an image generator. The image generator is responsible for rendering or generating the computer generated imagery (i.e., sensor, out-the-window view, etc.). Monitoring is the act of observing the output of the image generator in real-time. The goal is to see what the student sees, in a remote location, at the same time. Debrief is an activity where the output of the image generator is reproduced as an instructional aide for post mission analysis or debrief purposes. Monitoring and debrief systems can be divided into three classes: Tape Systems, Regeneration, and Digital Video. Figure 1 shows how the various systems interact with the host and image generator.

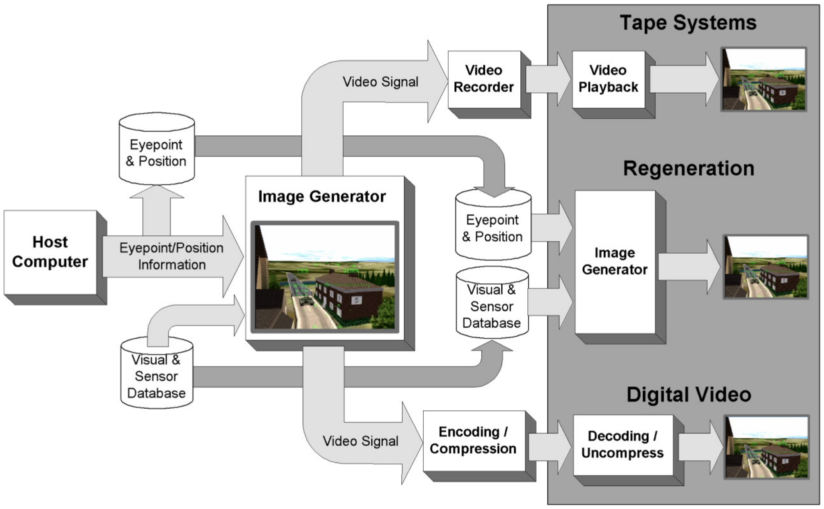

Figure 1: Monitoring and Debrief System design approaches.

Tape systems use video recording equipment to capture the imagery as it is produced to support debrief. For monitoring, the live video signal must be transmitted/routed for observation. As transmission distances increase, the quality of the video degrades. Traditionally, tape systems use analog recording techniques (VHS, SVHS, and Beta). These analog systems suffer from many limitations. These include reduced resolution relative to the source image, limited recording duration, and poor greater-than real-time playback quality (i.e., fast forward). When fast forward is engaged, the video is generally degraded with crawling lines, static, and warped images. Newer digital formats (DV, D9, Digital Betacam, etc.) provide greater picture quality, resolution and color representation but still suffer fast forward problems and have reduced recording times in high resolution (bit rate) modes.

Digital Video systems represent many new technologies that employ video compression. MPEG-2 is the most widely used compression technology and serves as the focus for this paper. While there are many video compression techniques, MPEG has been standardized and is used in many commercial and consumer products (e.g., DVD). In any compression technology there is an encoding stage and a decoding stage. The encoding stage converts the incoming video signal into a stream of bits (i.e., bit stream). The decoding stage simply converts the bit stream back into a video signal for display. One advantage of Digital Video is its ability to decode at greater-than real-time speeds without distracting artifacts.

Once the video is converted into a bit stream, the stream can be stored or transmitted using any digital transport medium (Ethernet, ATM, Firewire, etc.). Observation can now be supported at great distance with no loss in video quality using low cost computer networks. In addition, the same stream can be multicast to several observation locations simultaneously with no reduction in quality. The encoding, transport, and decoding stages of a Digital Video system each add latency during monitoring. In most systems, a 0.5 sec to 2.0 sec latency can occur from the time the image is generated until it is remotely displayed.

Both Tape Systems and Digital Video Systems must convert the output of the image generator into a compatible signal format. This process involves a change in both signal format and resolution. Since the resolution of most image generators is significantly greater than current tape and digital video systems, there is a loss in image quality. As high definition (HD) video recording and encoding hardware become available, the resolution problem is reduced. This topic is discussed in greater detail in later sections of this paper.

Regeneration is complex to implement, but has the potential to perfectly reproduce the output of the image generator. Using the information that flows from the host computer to the image generator, regeneration uses a second image generator to reproduce the display. During a simulation exercise, the host computer continuously calculates the ownship s current position, altitude, and attitude as it interacts with the simulated environment. This position information defines an eyepoint from which the image generator renders a view of simulated environment. As the simulation progresses, the position information is stored for debrief while being transmitted for monitoring. In addition to eyepoint data, other environmental information must be stored and transmitted. This includes current visibility/fog levels, time of day, weather conditions, and other environmental parameters.

Regeneration has the added cost of a second image generator. To combat this problem, a lower cost monitor/debrief image generator is often used. By reducing frame rate, field of view, or database complexity, monitoring and debrief can occur with minimal instructional loss. Ideally, both image generators use the same visual and sensor database. Otherwise, a complex database conversion process may be necessary. In the conversion process, the structure and contents of the database are modified to support regeneration. Conversion may introduce correlation problems when filtering reduces the polygon content. Additional complexity is added when semi-automated forces (SAF) interact with the simulation system. SAF entities and interactions must be replayed or reproduced during regeneration.

CHARACTERISTICS OF DIGITAL VIDEO

The term digital video is a somewhat generic term used to describe various production formats, tape formats, and as source for any of the Advanced Television Standards Committee (ATSC) standard definition (SD) and high definition (HD) video formats.

Color Space Conversion is a key aspect of digital video. Most computer-generated imagery is generated as a component RGB signal. Each component represents an intensity map of the image to be displayed for each of the primary colors (Red, Green, and Blue). When all three intensity maps are combined together, a full color picture is produced. Component RGB is ideal for maintaining high quality, full color video over very short transmission distances. Component RGB is difficult to compress for low bandwidth transmission. If a lossy compression technique is applied differently to each color component, color changes and color shifting results.

Sampling	Luminance Resolution	Chrominance Resolution		Reduction from 4:4:4	Notes
Source: 720x480	Y	U	V		
4:4:4	720x480	720x480	720x480	0%	Equivalent to component RGB
4:2:2	720x480	360x480	360x480	33%	MPEG Studio Production Quality
4:2:0	720x480	360x240	360x240	50%	MPEG Distribution Quality (DVD, DSB)
4:1:1	720x480	180x480	180x480	50%	DV (IEEE-1394) 525/60 Standard
4:1:0	720x480	180x120	180x120	63%	°

Figure 2: Chrominance Sampling and YUV resolution

For digital video applications the component RGB signal is color space converted into a luminance and two color difference signals (chrominance). This color space is often referred to as YUV or YCbCr. The Y represents an intensity value for each pixel element. The Cr (R-Y) and Cb (B-Y) represent red and blue color difference signals. There is no loss in picture quality or color fidelity in the color space conversion process. A component RGB signal has a single value per pixel for each color component. In YUV, one luminance value and two chrominance values are generated per pixel. The total amount of data is the same for both color spaces. The standard nomenclature for this level of data representation is 4:4:4. When the analog RGB waveform is sampled four Y, four U, and four V samples are taken per pixel period. The YUV color space provides compatibility for monochrome and legacy black and white televisions through the luminance signal (Y component). The UV components are not needed which can result in a 67% bandwidth reduction for applications needing monochrome video

Chroma (chrominance) sampling is the process of reducing the resolution of the chrominance (U and V) components. The eye is relatively insensitive to high frequency color changes. This allows for sub-sampling of the U and V components with little perceptual loss. A common chroma sampling used by MPEG compression is 4:2:0. A 4:2:0 chroma sampling maintains full luminance resolution, while the color difference signals are reduced to half of their horizontal and vertical resolutions. For example, a 720x480 4:4:4 signal is broken into a 720x480 luminance buffer while U and V are reduced to 360x240. A 50% bandwidth reduction is realized with minimal visual impact. Figure 2 lists many of the common chroma samplings and the resulting YUV resolution. The luminance (Y) component is never sub-sambled due to the human eye s sensitivity to changes in intensity.

Convergence between computer generated imagery and digital television standards is going to take time. Digital television standards are divided between standard definition (SD) and high definition (HD). Standard definition covers 640x480 and 720x480 resolutions for both 4:3 and 16:9 aspect ratios. The most common standard definition signals are 525/60 (NTSC) in North America and 625/60 in other areas of the world. The 525 and 625 referrer to the total number of lines per frame in the video signal. High definition includes 1280x720 and 1920x1080 for 16:9 aspect ratios. Computer imagery is generally produced in 4:3 aspect, in the following progressive resolutions: 640x480 (VGA), 800x600 (SVGA), 1024x768 (XGA), 1280x1024 (SXGA) and 1600x1200 (UXGA). While HD can handle SXGA resolution without loss, the availability of low cost hardware to support recording, distribution, and display is limited. Presently, MPEG technology is limited to SD resolutions.

The most common digital representation of the 525/60 signal is CCIR601 (ITU-R 601). CCIR601 is a studio quality, interlaced, component YUV signal using 4:2:2 chroma sampling and 10 or 8 bits per sample. The number of active samples is 720x485 (usually rounded to 720x480) which generates 270Mbit/sec (858 total horizontal samples over 525 lines).

For computer generated imagery CCIR601 has resolution and interlaced limitations. For interlaced video only 70% (Kell factor) of the active lines are perceivable. For a signal with 485 active lines (525/60) only 339 lines are perceivable. Scan line converters and video scalers can convert higher resolutions (i.e., SVGA, XGA, and SXGA) into CCIR601 format. Empirical examination has found that CCIR601can sufficiently represent VGA level resolution up to SXGA. For XGA and SXGA, small font text becomes increasing difficult to read. Overall, the loss in video quality is acceptable for most mission monitoring and mission playback and far superior to analog methods.

VIDEO COMPRESSION AND FORMATS

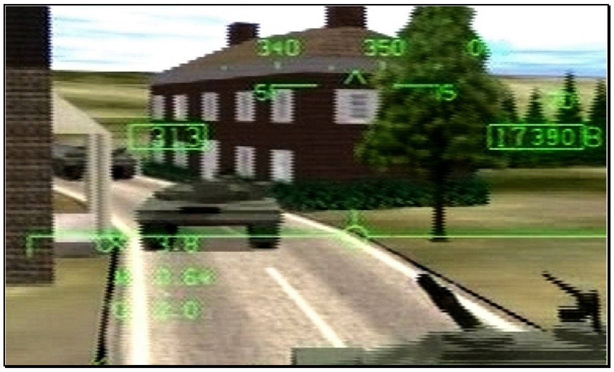

Figure 3: OTW scene with HUD overlay.

The goal of any compression technique is to maintain image quality and resolution while reducing the storage and transmission bandwidth requirements (Moreton, 1996). While there are many video compression techniques available, this paper is focusing on the MPEG-2 standard. MPEG was developed by the Motion Picture Experts Group to provide a standard for video and audio compression. The MPEG standard defines the syntax and structure of the video bitstream. To maintain interoperability, all MPEG video encoders must generate a syntactically compliant bitstream from which any MPEG decoder can regenerate the video image. While the MPEG standard is in the public domain, the manner in which MPEG encoders generate a compliant bitstream remains proprietary. The quality and bit rate of the resultant bitstream allow for a great deal of market differentiation between MPEG hardware vendors.

The MPEG compression process works by removing spatial, temporal, and perceptual redundancies from the source video image (Moreton, 1996). This section walks through the compression process and explains how each redundancy is systematically removed and how the resultant image is affected. The scope of this paper is to explain how MPEG compression can be used to effectively record computer generated imagery for simulation purposes. To maintain this focus many of the implementation details are not covered, but emphasis is placed on those issues that affect the quality of the recorded bitstream. Computer generated imagery represents a special challenge to MPEG compression. MPEG compression is optimized for natural imagery as seen in motion pictures and television. Natural imagery contains very little high frequency information, a factor that MPEG exploits heavily. Computer generated imagery, has the potential to contain high frequency information. This information comes in the form of flight, weapon, and other symbology and textual information which is seen as a image overlay. Figure 3 shows a computer generated scene of an out the window (OTW) image with overlaid HUD symbology (altitude, air speed, horizon line, compass, etc.). The OTW (background image) represents the natural information of this simulated world. In contrast, the overlaid symbology contains very high frequency information relative to the OTW image which introduces an unnatural component to the aggregate image.

As shown in figure 4, the starting point of the MPEG compression process is color space conversion and chroma sampling. Perceptual redundancy is removed from the compressed video signal as a result of 4:2:0 chroma sampling. 4:2:0 chroma sampling is used by most (non-production studio) MPEG encoders. By down-sampling from 4:4:4 (RGB) to 4:2:0 in YUV color space, 50% of the original RGB signal is lost with minimal impact to the human eye. The human eye is very sensitive to changes in luminance (Y), for this reason the luminance component is preserved at full resolution. The human eye is much less sensitive to color changes that allows for the half-resolution color difference signals (U and V) (Moreton, 1996).

The human eye is fooled quite easily when natural imagery is compressed. When unnatural overlay graphics are added to the scene visual artifacts start to appear. Referring back to figure 3, observe the artificial horizon line in the center of the picture. While the full resolution luminance sampling correctly represents the 1 pixel wide line, a color artifact is present. Upon close examination, the high intensity green bleeds into the surrounding pixels. This is caused by the half-resolution color-difference sampling. Empirical studies have shown that this is not a problem for most computer-generated imagery unless multiple colors are competing for the same U and V sample. For example, if a red line were to be displayed 1 pixel below the horizon line the resulting compressed image would be blurred. This becomes a serious problem for text displayed against a non-black background. When designing a full color text display (e.g., MFD, MPD, IOS, etc.) the use of color should be analyzed if the display is to be MPEG encoded. Upgrading to a 4:2:2 chroma sampling solves many of these problems but comes at a much greater cost along with addition storage requirements.

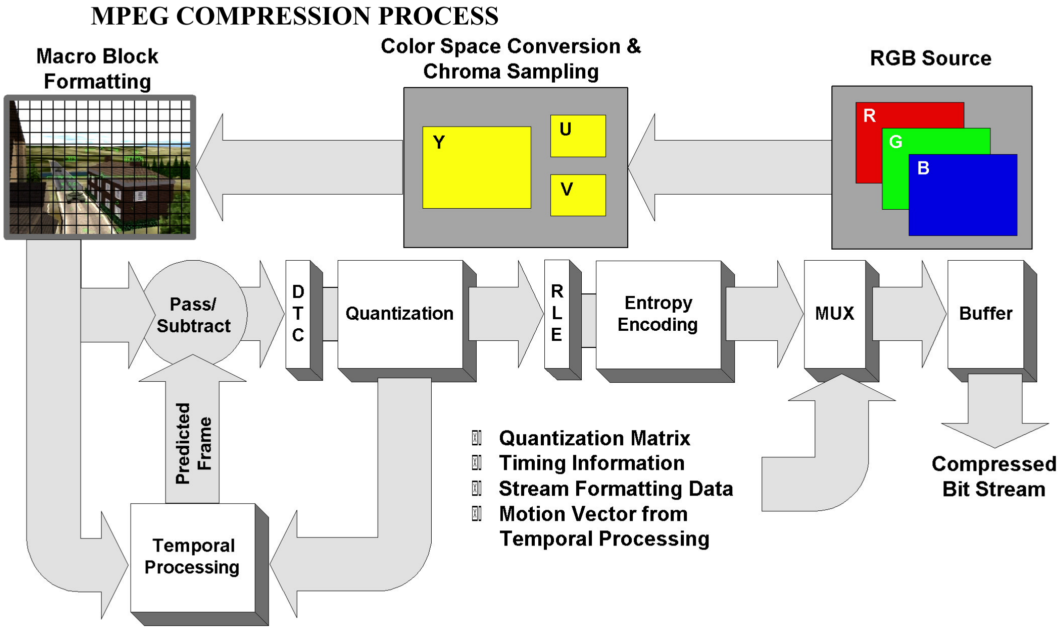

Figure 4 MPEG Encoding Process

Macro block formatting is the next stage in the compression process. Each macro block represents a 16x16 pixel area. This area is comprised of four 8x8 luminance blocks (Y) and two 8x8 chrominance blocks (U and V). The block serves as the primary tool used in removing temporal redundancy. When video is tested on a frame by frame basis, very little changes between adjacent frames (Moreton, 1996). By removing temporal redundancy only the parts of the frame that have changed from the previous frame are compressed and added to the bitstream. Areas of the frame that have not changed are reused from previous or future frames.

MPEG uses three frames types to remove temporal redundancy. These frames types are I, P, and B. The I, or intraframe, can be considered to be a full frame snapshot. The P, or predicted frame, uses information from previous I or P frames to predict the location of a block in the future. The B, or bi-directional frame, uses information from previous and future I and P frames. By buffering frames the encoder is able to exploit redundancies from previous and future frames. While buffering increases encoder latency, it helps to minimize the resultant bitstream. To generate a P or B frame the encoder searches surrounding blocks to find a match. When a match is found, a motion vector is generated to allow the block to be correctly positioned in the frame. The encoder must choose when to generate an I, P, or B frame. This selection process and the efficiency of the block search algorithm often define the quality of the encoder and of the resultant bit stream.

Sequences of I,P,B frames are assembled to reconstruct the image. An I frame serves as the starting point since it represents a fully compressed frame. It is followed by a series of P and B frames. After a period of time, another I frame is issued and the process repeats. It is important to have a periodic I frame (usually twice a second) otherwise small errors accumulate resulting in an electronic form of entropic heat death . When this occurs, the picture quickly degenerates. Refer to Figure 5 for a sample sequence of I,P,B frames. The I frame severs as the basis from which subsequent frames are created. Note how the sky, which changes very little from frame to frame, is stored in the I frame and used by the B and P frames which follow. A group of pictures (GOP) is defined as an I frame and subsequent P and B frames up to the next I frame.

Converting from the spatial domain to the frequency domain is performed by discrete cosine transformation (DCT). The 8x8 block of pixels is represented in the frequency domain by an 8x8 matrix of frequency coefficients. The upper left corner of the 8x8 matrix represents the lowest common frequency in the block. It is referred to as the DC component and represents the average value for the block (Moreton, 1996). The lower right corner represents the highest frequency in the block. Horizontal frequencies increase as each row is traversed from left to right. Vertical frequencies increase as each column is traversed from top to bottom.

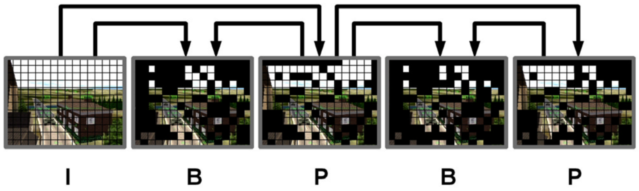

Figure 5: Sample MPEG sequence of I, P, B frames

Quantization is the process of removing high frequency information. While the DCT process is without loss and completely reversible, the quantization process is (very) lossy. High frequencies are removed by applying a scale factor to the DCT coefficients. As shown in Figure 6 the DCT coefficients enter the quantization process from the left. By applying different scale factors more high frequency information is removed. In the figure three very simplistic scale factors are represented. Note how the higher frequencies are scaled more severely in each of the scale factors. The goal is to force as many of the high frequency coefficients to zero. The quantized result can be seen on the third column. The quantization scale factor in the first row is less severe thus preserving more of the high frequencies. The scale factor in the bottom row is very severe forcing many of the coefficients to zero. The resultant matrix is then RLE and Entropy encoded which is explained later in this section.

Figure 7 is a small study in the effects of quantization. The decoded MPEG-2 source image is shown in the lower left corner. The white box near the center of the source image is the focus of this study. Arranged in a clockwise order around the source image are six thumbnails representing various output stream bit rates from 1Mbit/sec to 15Mbits/sec. The 1Mbit image represents an extremely quantized image. The only DCT coefficient remaining is the DC component that preserves the average color in the decoded thumbnail. The quantization scaling progressively decreases, preserving more of the high frequency information as you proceed from the 3Mbit to the 15Mbit thumbnail.

Figure 7 shows the effects of MPEG compression on overlay graphics. As discussed earlier, the overlay graphics on the HUD symbology introduce many high frequencies into the image. If the quantization factors are too high the symbology quickly becomes unreadable. Experimentation has shown a point of diminishing returns around the 5-7Mbit range. In this range the overall bitstream is minimized and the symbology remains readable. At the 15Mbit stream rate the resultant image exhibits no visual artifacts from quantization. At lower bit rates regions of the screen which are moving quickly may appear to be more grainy . This is a result of a high quantization factor being applied to the moving region. Note: when decoding the MPEG stream at 30 frames per second the eye tends to integrate successive frames together. This process improves the overall image quality making lower bit rates more acceptable to the observer.

The amount of quantization is controllable by the encoding hardware. It is becoming more common for MPEG encoders to support both constant bit rate (CBR) and variable bit rate (VBR) encoding. In CBR the output bit stream maintains a constant level of output. For example, CBR at 5Mbit/sec tends to generate 4.5 to 5.5Mbits/sec. In VBR, an average bit rate (desired) and the maximum bit rate are defined. The encoder uses of process of adaptive quantization to attempt to maintain the average bit rate. In video sequences with sustained high speed movement the bit rate may reach the maximum level. The goal with adaptive quantization is to maintain a level of quality. In CBR the goal is to maintain a level of bandwidth.

The final stages of the compression process involve run length encoding (RLE) and entropy encoding. RLE takes advantage of spatial redundancies in the image. A solid black image serves as a good example. The bitstream emerging from the quantization stage contains a 10-bit sample for each luminance and chroma sample. Since the image is black all 10 bits are zero. RLE would convert this bitstream into a single 10-bit sample and a count indicating the number of total samples. Once the stream is RLE encoded it is then entropy encoded. Entropy encoding attempts to reduce the bitstream to its minimal size (i.e., lowest energy). MPEG used Huffman encoding for this process. Huffman encoding attempts to find patterns in the bitstream. Patterns which occur frequently are encoding with fewer number of bits. Patters that are less frequent are encoded with more bits. A good example of a Huffman- like encoding is Morse code. Frequently occurring characters are represented with minimal dot/dashes where less frequent characters (x, z, q) are encoded with longer dot/dashes sequences (Moreton, 1996). RLE and Huffman encoding are lossless algorithms that help to minimize the bitstream.

The output of the compression process is a packetized elementary stream. The raw compressed data emerging from the RLE and entropy encoding stage is packed with MPEG header information. This information

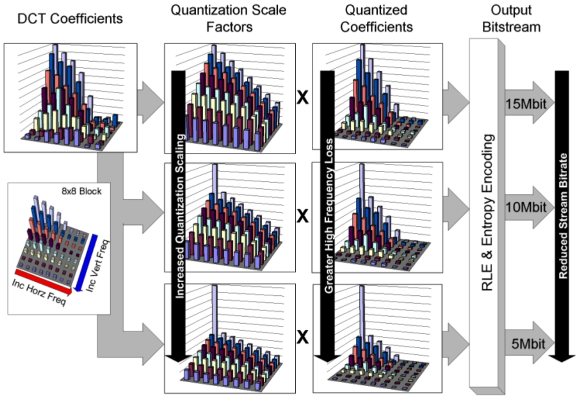

Figure 6: Quantization of DCT Coefficients

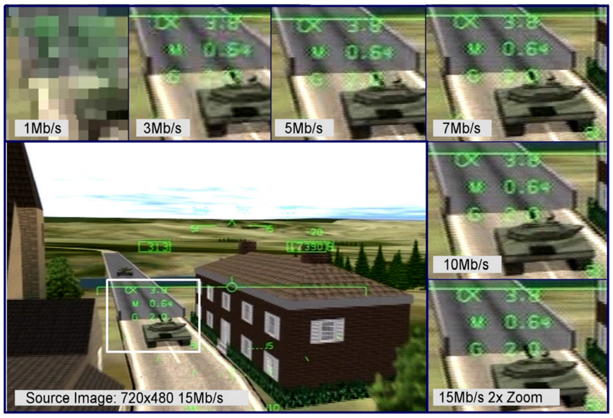

Figure 7: Effects of Quantization on imagery

includes time stamps, start codes, quantization tables, and other information needed to generate a syntactically correct MPEG bit stream.

Currently MPEG is available in three formats: MPEG-1, MPEG-2 and MPEG-4. MPEG-1 was originally designed for the compression of progressive scan video while MPEG-2 was primarily designed for interlaced video. Both formats have the capacity for very high resolutions: 4095x4095 at 60 frames per second for MPEG-1 and 16383x16383 for MPEG-2. When MPEG-1was first introduced a set of constraints were established to limit computation complexity, buffer size, and memory bandwidth (Fogg, 1996). These constraints were termed constrained parameter bitstreams (CPB). Due to technological limitations, MPEG-1 was established at a CPB resolution of 352x240 @30Hz. Later as computation speeds and capacities increased higher resolutions could be achieved. Higher resolutions are now available using the updated and enhanced syntax in the MPEG-2 standard.

As processing speeds increase, MPEG-2 will become available in higher resolutions. Presently, MPEG-2 is available in what is referred to as main profile at main level (denoted MP @ML) The level of an encoder measures its resolution, while the profile refers to the chroma sampling and serves as a measure of decoder complexity. Main level resolution is 720 samples x 576 lines. Based on earlier discussion this gives an effective working resolution of 640x480 (VGA). Main level resolution provides an important intersection point between computer generated imagery and the use of MPEG encoding hardware for video recording. For high end production environments, 4:2:2 chroma sampling at main level (4:2:2@ML) is available. Although 4:2:2 chroma sampling provides greater color representation and minimizes some visual artifacts the additional cost is generally prohibitive.

The performance characteristics of the storage device limit the maximum sustainable recording bit rate. When choosing a storage device the number of streams to be simultaneously recorded and the bit rate of each stream should be carefully considered. For lower bit rates (5-7Mbits/sec) multiple simultaneous streams may be recorded on low cost 7200rpm IDE ATA/100 hard drives. For multiple simultaneous streams at 15Mbit/sec and above, high-speed SCSI drives should be employed. When analyzing a storage device s performance characteristics consider the sustained data transfer rate and the worst case seek time (generally full track). While many drives claim performance characteristics that would imply the easy application of multiple 80Mbit/sec streams, the actual performance tends to fall short. Having sufficient operating headroom becomes particularly important as the storage device reaches capacity. At capacity, performance generally degrades to worst-case. This may jeopardize successful stream recording. To combat this problem, some storage media vendors have developed products specifically aimed at streaming media storage. These specialized devices and systems are designed to maintain a stated performance to full capacity. For safety, a recording system should utilize only 10% of the storage device s stated performance (when using general purpose devices). For example, a storage device may specify a sustained transfer rate of 36MB/sec; only 3.6MB/sec should be assumed for actual performance. This rule of thumb becomes particularly important when a system records multiple streams to the same storage device.

SVRS USE IN THE AVCATT-A

This section examines the Sensor Video Recording System (SVRS) developed for the U.S. Army s Aviation Combat Arms Tactical Trainer — Aviation Reconfigurable Manned Simulator (AVCATT-A). AVCATT-A provides collective, unit-level training for Army aviation reconnaissance, attack, assault, and support units via six networked, reconfigurable cockpits interacting in a rich synthetic battlespace housed in a mobile facility. AVCATT-A provides training for up to twelve aviators in up to seven rotary wing configurations. The configurations include AH-64 Apache, AH-64D Longbow, OH-58D Kiowa Warrior, UH-60 A & L Blackhawk, and CH-47D Chinook and RAH Comanche. Each AVCATT-A suite includes six manned-modules (MM) which can be individually configured to any of the aircraft configurations.

The AVCATT-A suite is comprised of two mobile semi-trailers. The first trailer contains three manned-modules and the Battle Master Control (BMC) area. The second trailer contains three additional manned-modules and the After Action Review (AAR) area. The BMC serves as the central control for the AVCATT-A suite. BMC functionality includes initialization, control, recording, and monitoring. The AAR provides real-time monitoring and after action debriefing for an audience of twenty personnel.

The Sensor Video Recording System (SVRS) provides video recording, monitoring, and playback capabilities for the AVCATT-A system. Depending on configuration, up to three channels of sensor video are recorded per manned module. These sensors include RADAR, Forward Looking Infrared (FLIR), Day Television (DTV), Night Vision Goggles (NVG) and Direct View Optics (DVO).

The SVRS has three modes of operation: recording, monitoring, and playback. During record, each sensor video channel is MPEG-2 compressed to maintain high image quality while minimizing storage. During playback, any of the recorded sensor channels are viewable to support debrief activities. Monitoring provides near real-time observation of the sensor during the training session. Recording, playback and monitoring can occur simultaneously. To support these operations, the SVRS has storage capacity for up to two training sessions. While the on-going mission is being recorded, playback occurs from previous mission recordings to support debrief activities.

The SVRS provides a sensor video display for the AAR and the BMC. The sensor display provides six simultaneous sensor views. The AAR and BMC can each choose which sensors are displayed. The ability to switch between monitoring of the on-going training session, and playback of previously recorded sensors from a prior session is provided. In AVCATT-A operation, the SVRS provides monitoring capabilities for the BMC and the ability to switch between monitoring and playback for the AAR. The SVRS receives control information from the AAR and BMC simultaneously during training and debriefing activities.

The SVRS meets all AVCATT-A recording/playback requirements including: record 3 channels per manned module, 18 total channels per AVCATT-A suite; support 8 hours of recording for each channel; a capacity to store current (on-going) and previous training sessions (18 channels x 8 hours per channel x 2 sessions = 288 hours of recording); support two independent 6-channel sensor displays in the AAR and BMC (sensor display contains 1 full resolution image and 5 reduced/thumbnail images): support ability to select 6 of 18 channels per sensor display; switch between monitoring and playback for AAR; support playback speeds from _x (slow motion) to 10x; maintain readability of sensor overlay graphics; provide random access video on demand playback services to support timestamp access; insure network utilization is limited to 50% of maximum bandwidth.

The SVRS is a video server. From a functional fidelity perspective it is best considered to be an 18 channel VCR with the ability to view 12 of the 18 channels concurrently. The SVRS is designed to provide recording, monitoring, and playback capabilities using encoding stations, decoding stations, and video effects hardware. Each station functions independently, but collectively (as a system) provide the required services using commercially available hardware. Figure 8 provides a high level view of the SVRS. The encoding and decoding stations are grouped together with half of the encoding/decoding stations located in the AAR trailer and the other half in the BMC trailer.

Encoding stations provide recording and video streaming functionality. Each sensor video channel is converted into a digital MPEG-2 data stream and stored locally in the encoding station. Each encoding station is equipped with 2 storage drives capable of eight hours of recording per channel. This provides storage for the current/ongoing training session (drive 1) while providing playback capability from the previous session (drive 2). The encoding station is capable of streaming 6 channels of video simultaneously while recording. In monitor mode, the MPEG-2 data is streamed to the decoding stations while being recorded. In playback mode, the MPEG-2 data is retrieved from disk and streamed to the decoding stations. Six streams can support 3 channels of monitoring (BMC) while providing 3 channels of playback (AAR). Alternately, 3 monitor channels to the BMC and 3 monitor channels to the AAR are provided.

The decoding stations convert the MPEG-2 data stream back into a video signal. The decoding station is the

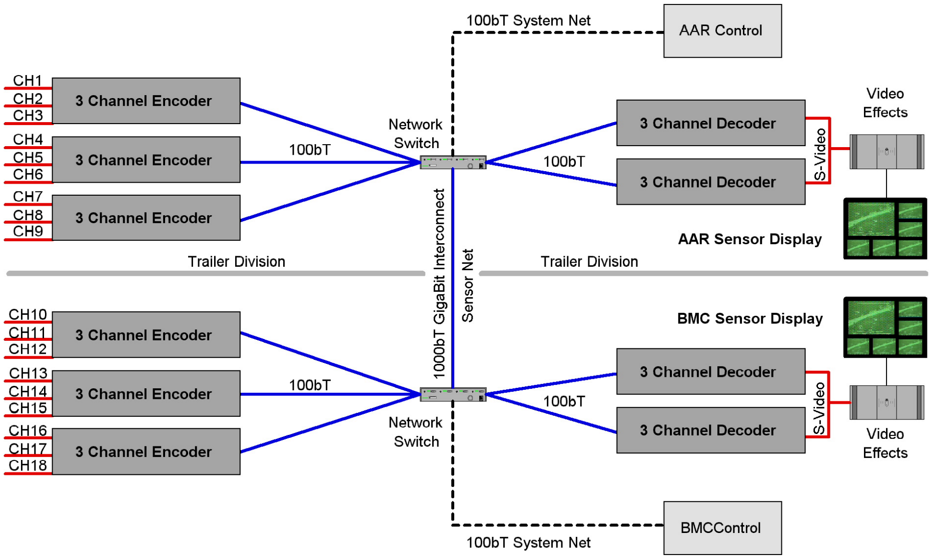

Figure 8: SVRS Architecture

receiver of the MPEG-2 data stream. The AAR and BMC each contain two decoding stations to support 6 display channels. Based on the channels selected by the users in the AAR/BMC, the decoding stations auto-configure themselves to receive the MPEG-2 stream from the encoding stations.

The video effects box is responsible for arranging and combining the 6 independent video signals from the decoding stations. The video effects box generates a single video output with 1 of the 6 input signals shown as full size while the other 5 are shown as thumbnails. Each video effects box supplies a signal to an overhead projector in the AAR and BMC.

The network topology isolates the SVRS from all network traffic except for commands received from control stations. Video data is streamed from the encoding stations to the decoding stations. Each channel of video utilizes approximately 5% (5Mbits/sec) of the 100bT sensor network. An individual encoding station can stream a maximum of 6 channels simultaneously; 3 to the AAR and 3 to the BMC. The 50% network utilization is maintained with only 30+% utilization for 6 channels. The two semi-trailers are interconnected via 1 Gbit (1000bT). The interconnection allows the sharing of recorded streams between trailers. The worst case scenario occurs when the AAR selects channels 10 through 18 and the BMC selects 1 through 9. At this point a total of 12 streams (6 streams in each direction) are being sent across the gigabit interconnection.

The AAR and BMC control are connected to the SVRS via a 100bT system network connection. This connection represents the only link with the AVCATT-A system. Through this connection, command packets are broadcast to all encoder/decoder nodes. Status information is returned to the control stations via the same connection. The AAR and BMC control provide the user interface to the SVRS. The interface provides recording control, channel selection, playback control, video effects control, and system status.

The software architecture of the encoding and decoding stations in the SVRS allows for simultaneous, multichannel recording and playback streaming. By dedicating a separate thread of execution for each task, a single encoding PC is able to handle recording, playback, and streaming functionality.

Playback speed changes are supported for _ speed (slow motion) to 10x (fast forward) without distracting visual artifacts. MPEG-2 video compression was chosen for its ability to maintain image quality for a 640x480 source (standard NTSC video) image at low bit rates. For the SVRS, a 5Mbit/sec constant bit rate encoding was employed. Subjective testing revealed 5Mbit/sec provides VHS image quality while maintaining high compression (270Mbit/sec CCIR601 source to 5Mbit/sec MPEG-2 stream). At 5Mbit/sec, 8 hours of compressed video can be stored on a 25GB hard drive. Using two 75GB hard drives per encoding station provided capacity for two, 8 hour recording sessions for 3 channels. CCIR601 digital video was used as the encoder input format. The encoding and decoding stations work together but function autonomously. Coordination and synchronization is achieved by all of the stations working together at the same time. To make this happen, a command must be received and executed by all stations simultaneously.

Video streaming applications must balance latency with system stability. Latency is amount of time to compress the video signal into a MPEG-2 bit stream (encoder latency), transmit the bit stream across the network (transport latency), and convert the bit stream back into a video signal for display (decoder latency). The encoder generates and the decoder consumes data at the same rate. If problems occur in the data handling or transmission process (i.e., lost or late packets) the decoder is unable to generate a video signal for display. The system quickly becomes unstable and results in a lost image.

Data buffering at the encoding station (server) and the decoding station (client) ensures stability, but increases latency. The SVRS streams data at 5Mbit/sec. This translates to 640Kbytes per second of video (on average). The number of channels processed and streamed per encoding station and the total number of channels simultaneously streaming are critical parameters to the buffer tuning process. As more channels are added to each encoding station, the system processor (CPU) is forced to service more threads. More buffering is needed to balance the flow of data between channels. As more channels are streamed across the network, buffering is needed to reduce network congestion. With buffering the SVRS latency is 1.3 seconds.

Minimizing latency is important for real-time monitoring and playback speed switching. When playback speed is switched (e.g., normal speed to 10x speed) the buffers must first be exhausted before the change is speed becomes evident. For example, a 64Kbyte server buffer, 0.3 second transport, 64Kbyte client buffer, and 0.4 second decoder latency results in 0.9 seconds of latency. Upon a user requested speed change, the change will not appear for 0.9 seconds and must last a minimum of 0.9 seconds. As a result, quick speed changes are not possible in a network streaming application. Monitor mode, where live video is compressed and streamed simultaneously is also subject to the same delays.

LESSIONS LEARNED / SUMMARY

The use of digital video technology has many training advantages over analog recording. MPEG-2 image compression provides superior image quality in an industry standardized digital format. As the MPEG standard evolves higher resolutions at lower bit rates will be achieved. While current MPEG-2 technology is limited to NTSC (640x480) resolutions the use of scan line conversion provides effective recording capability for higher resolutions. MPEG provides enhanced capabilities for mission monitoring and debriefing. MPEG provides greater than real time playback without the distracting artifacts associated with analog media. Speeds from x0.5 to 10x can easily be supported to enhance mission debriefing. Random access provides virtually instant video on demand playback capabilities when jumping between mission mark points. Long duration missions can be seamlessly recorded without the need to change recording media. Remote observation can be supported via network streaming. Using a standard computer network the encoded video stream can be transmitted over long distances without loss in video quality. The use of broadcast/multicast technologies allows for multiple viewing locations. Depending on the encoding parameters, 10 to 20 channels of video can be streamed on a single 100bT network. This provides scalability for multiple simulators linked in a collaborative environment.

In summary digital video is an intriguing and complex technology that has recently matured and become cost effective to the point that it can be applied to address challenging simulation requirements. While the full potentials of this technology are still emerging, today s capabilities can be used to design simulation applications that were prohibitively costly only a few years ago. The AVCATT-A Sensor Video Recording System (SVRS) is such an application. The state-of-the-art SVRS in AVCATT-A facilitates enhanced Manned Module sensor video selection and monitoring at the Battle Master Control (BMC) station and advanced record/replay capabilities in the After Action Review (AAR) area. Aviation units are provided with an AAR that supports eight hour missions with an affordable number of components, no need to setup and use time sync generators and no need to segment an AAR to allow the replacement of video cassettes for each of the eighteen channels being recorded.

REFERENCES:

Fogg, C. What is MPEG. (1996). Berkeley Multimedia Research Center, University of California at Berkeley. http://bmrc.berkeley.edu/research/mpeg/faq/mpeg2-v38/faq_v38.html. version 3.8, (June 4, 2000).

Moreton, H. (1996, Jul/Aug & 1996 Nov/Dec), Understanding Compression. Developer News. Mountain View, CA: Silicon Graphics Inc.

MPEG Specification: Information Technology — Generic Coding of Moving Pictures and Associated Audio, ISO/IEC 13818 Part 2, Video, ISO/IEC 13818-2:1996 AMD5:2000, ITU-T Rec H.262