COMPARISON OF VIDEO COMPRESSION STANDARDS IN A TSPI APPLICATION

Michail Stolhand
JT3 UTTR/Hill AFB

D. Shane Woolridge
Apogee Labs Inc.

ABSTRACT

The need to compress video for the purpose of transport and recording plays a large part in most telemetry applications. Throughput latency requirements, bandwidth limitations, quality, and full motion characteristics are all considerations when deciding on which compression algorithm to use. Another deciding factor is the environment in (or conditions under) which the video will be used when decompressed. This paper will evaluate three standard types of video compression, MPEG2, MPEG2 using I-frame only, and MJPEG. These compression techniques will be compared to a legacy Wide Band System (ALS) for TSPI (Time-Space-Position-Information) error analysis and subjective quality assessment.

KEY WORDS

Video Compression, Latency, MPEG2, Motion JPEG, Quantization, Skewness, Kurtosis

INTRODUCTION

One of the many services provided by UTTR (Utah Test and Training Range) is the capability of delivering TSPI data to their customers. Due to the nature of the tests being conducted, this video data is often gathered at remote unmanned facilities. In the past, raw analog video tape recordings were made, and then transported to a video processing facility several hours away. This process could take several days before video engineers could start processing the recordings. Therefore, a SONNET network was added to UTTR for the purpose of moving this video from the remote test sites to the main facility in near real-time. In order to get the video signals across the network, as in most data transport applications, they must first be converted from analog to digital signals. The main obstacle in the transportation of these digitized video streams is the aggregate bandwidth that is a result of the digitizing process. It is impractical to transport these streams without compression.

For example, in order to determine the data rate needed to reproduce analog video in a straight digital format, a calculation can be made based on the number of pixels found on an NTSC (720 pixels per line, 525 lines per frame, 30 frames per second) formatted video screen. The analog signal that represents color video consists of three components. Luminosity is the primary component that represents the black and white image. Chrominance represents the red and blue components of the image, with green being mathematically derived from the values of red and blue. The quantization levels for each are: 8-bits to luminosity, 4-bits to Cr (Chrominance red), and 4-bits to Cb (Chrominance blue). This structure is often referred to by the ratio of 4:2:2 and results in the need for 16 bits per pixel. At 13.5 M samples per second, which is a common sample rate for commercially available digitizers, the result is a serial data stream of 216Mbps. If multiple video streams need to be transported simultaneously, most networks will be overflowed by the aggregate bit rate.

Several video compression techniques have been developed to reduce the digitized video bit rate. One common method of compression is Motion JPEG, or MJPEG. To briefly describe this compression method, we start by performing two-dimensional Discrete Cosine Transform (DCT) coding on small blocks (8 pixels by 8 lines) of the picture to produce blocks of coefficients. Applying some level of quantization to each coefficient filters the results. The greatest reduction of information is achieved by reducing the amount of high frequency components, which are not easily distinguishable by the human eye. It is the quantization that is introduced that makes the compression process ‘lossy’.

Finally, entropy coding, which is a form of lossless data compression, is applied to the blocks. This process groups components with similar frequencies together and creates variable length code (VLC) words to represent the components. This process is called intra-frame coding because each frame is independently coded. The throughput delay for this type of coding can be kept well below 100 ms. The delay through the encoder and decoder is equal to the buffer size divided by the channel bit rate.

MPEG2 uses this same intra-frame coding technique, but also employs motion-compensated inter-frame prediction where some frames are predicted based on reference frames. Three frame types are defined. Intra frames (I-frames) are coded with no reference to other frames, Predictive frames (P-frames) can use previous I or P-frames for motion compensation, and Bidirectionally-predictive frames (B-frames) can use previous and next I or P-frames for motion compensation. B-frames also allow for the highest degree of compression as well as the highest degree of quality. A repeating Group Of Pictures (GOP) can be displayed in any order of frame types, with any number of frames in the group. A typical pattern can be seen in Figure 1.

B_1 B_2 I_3 B_4 B_5 P_6 B_7 B_8 P_9 B_{10} B_{11} P_{12}

Figure 1 – Typical GOP Pattern

When measured, an I-frame is about three times larger in size than a P-frame, which is about 50% larger than a B-frame. This is why a typical GOP will only contain one I-frame, several P-frames, and a majority of B-frames. However, because the B-fames require the GOP to reorder the sequence of frames (the previous and next I and P reference frames are transmitted first) a significant delay is added to the coding/decoding process. B-frames can add well over 100 ms to the throughput delay.

Because MPEG2 is flexible regarding the amount and type of frames used in a GOP, it can run in an I-frame only mode, which is similar to MJPEG where every frame is independently compressed. This is important in a high-speed scoring TSPI application where accurate frame-by-frame measurements must be made, and predicted images cannot be trusted.

Another feature worth mentioning with the MPEG2 compression device used in this testing is a unique approach of synchronizing the output video stream to the input video stream by phase locking the output encoder to the input encoder reference. In a typical encode/decode process, a 27 MHz reference is phase-locked to an incoming video stream and is used during the sampling of the video. During the decoding process, a separate 27 MHz reference is used. Because these references are not exact, as they drift the decoder will eventually and periodically skip a frame or repeat a frame. In a scoring application, this anomaly could happen during a critical frame and needs to be corrected. Apogee Labs has improved upon this process by measuring the 27 MHz reference at the input encoder and inserting this measurement in the video packet header before being transported over the link where it is exactly replicated on the output decoder.

SYSTEM ARCHITECTURE

The test site contains a target that is surrounded by several calibration poles. These poles provide a known reference to the field of view, and allow for accurate measurements during frame-by-frame, and field-by-field evaluation. Several cameras are positioned outside the circle of calibration poles. If three or more cameras are used, a good three-dimensional awareness can be created during post-processing. The video streams from the cameras are time stamped and sent through a distribution amp. This enables the raw analog video to be recorded directly out of the cameras while the same video is simultaneously compressed, multiplexed and transported over DS-3 using the Apogee Labs Model 430 MITC *FALCON*. This multiplexed stream is received near real-time at the video-processing center, where it is demultiplexed and the individual video streams are recorded. Please refer to the system diagram below.

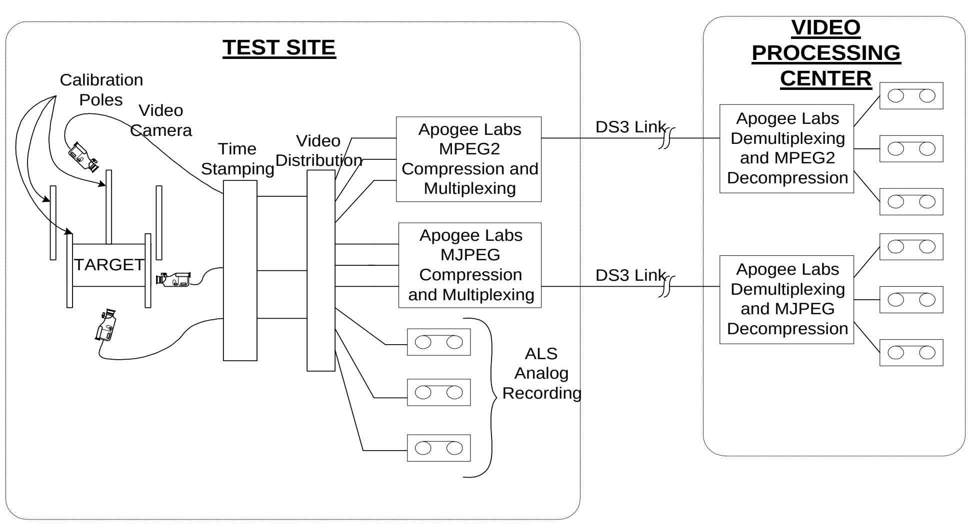

Figure 2: Hill AFB Configuration

TESTS PERFORMED

For testing purposes, an 8" target was mounted on a vehicle and driven across a 220' field of view at a relatively high rate of speed. This vehicle moves about 2 feet per video field, which can be compared to an actual small 450 kts missile where it is completely moved in successive fields with no overlap. Two sets of three cameras were used in a total of six data passes, which were processed independently by two data analysts. ALS is full stream, uncompressed video. The only adjustable setting on the MPEG2 compression module is the data rate since the resolution is auto-set based on the compression rate. This MPEG2 video was captured and analysis was done at bit rates of 6 Mbps and 14 Mbps compression rates. The following user settings apply to the JPEG data in this test:

- 13.333 Mbps
- Mode: NTSC
- Horizontal Resolution: 560 pixels
- Quantization: High
- Interlace: On

The final two data passes contain three cameras of ALS full stream, uncompressed video. It was compared in one test to JPEG compression using the settings in the bullet list above. In the second test it was compared with two MPEG cameras running at 14 Mbps and one JPEG camera using the settings in the bullet list above.

ANALYSIS

The data was brought into spreadsheets similar to the ones shown in table 1. The first step in analysis was to difference the raw data by data pass, separated into categories of camera and data compression speed, between the three file compression types, ALS minus JPEG, ALS minus MPEG and MPEG minus JPEG. Then the mean, standard deviation, maximum, minimum, skew[1] and kurtosis[2] of these differences were computed as illustrated in table 2.

Table 1: Example of Data Transferred to Spreadsheet

Camera 2 ALS Raw Data

Time Sec	Delta Time Sec	Az Degrees	El Degrees	Horizontal Pix/Read	Vertical Pix/Read
57.6340		264.0697	0.0444	304.2500	299.2500
57.6507	0.0167	263.9919	0.0295	299.7500	300.0000
57.6673	0.0167	263.8580	0.0149	292.0000	300.7500
57.6840	0.0167	263.8494	0.0250	291.5000	300.2500

---

[1] Skewness characterizes the degree of asymmetry of a distribution around its mean. Positive skewness indicates a distribution with an asymmetric tail extending toward more positive values. Negative skewness indicates a distribution with an asymmetric tail extending toward more negative values.

[2] Kurtosis characterizes the relative peakedness or flatness of a distribution compared with the normal distribution. Positive kurtosis indicates a relatively peaked distribution. Negative kurtosis indicates a relatively flat distribution.

Camera 2 JPEG Raw Data

Time Sec	Delta Time Sec	Az Degrees	El Degrees	Horizontal Pix/Read	Vertical Pix/Read
57.6340		264.1607	0.0816	304.0000	294.0000
57.6507	0.0167	264.0305	0.0282	296.3330	296.6670
57.6673	0.0167	263.9547	0.0554	292.0000	295.3330
57.6840	0.0167	263.9210	0.0353	290.0000	296.3330

Camera 2 6.001 MBS MPEG Raw Data

Time Sec	Delta Time Sec	Az Degrees	El Degrees	Horizontal Pix/Read	Vertical Pix/Read
57.63400		264.15210	0.04830	304.33300	299.33300
57.65067	0.01667	264.12880	0.05510	303.00000	299.00000
57.66733	0.01667	264.04350	0.02170	298.00000	300.66700
57.68400	0.01667	263.96140	0.05580	293.33300	299.00000

Table 2: Sample of Computations & Statistics

Camera 2 ALS-JPEG Difference Data

	Az Degrees	El Degrees	Horizontal Pix/Read	Vertical Pix/Read
	-0.0910	-0.0372	0.2500	5.2500
	-0.0386	0.0013	3.4170	3.3330
	-0.0967	-0.0405	0.0000	5.4170
	-0.0514	0.0202	2.6670	2.4000
Average:	-0.087585	0.007825	0.531800	3.015850
Std Dev:	0.024632	0.019082	1.438378	0.947785
Max:	-0.038600	0.028300	3.417000	5.417000
Min:	-0.124000	-0.040500	-1.600000	2.000000
Skew:	0.365227	-1.577437	0.376790	1.576484
Kurtosis:	-0.694943	1.992912	-0.695715	1.988802

The first comparison ran was between operators. This was done as a base line. For a system to introduce significant bias it must introduce more bias than is present simply by changing operators. The first step in processing VCAS data is to have the raw data “read” by an operator. The data is read by digitizing each separate image of the video data to allow it to be displayed by the VCAP program. The VCAP program then displays an image, which allows the operator to put a cursor on the point to be read and clicks the mouse button. This process continues for all images from each camera. Because this operator is a human being, reading introduces biases based upon the operator’s vision, judgment, and experience.

Table 3 shows the RMS difference, in pixels, computed from the statistics mentioned above. First the RMS of the average difference and of the standard deviation of that difference for both the horizontal (u) and the vertical (v) counts for each operator is computed by camera. Then those values are combined into a total RMS difference and standard deviation. These small differences will form a baseline for our other comparisons.

Table 3: Operator Comparison

Operator 1		**Operator 2**	
In pixels		***In pixels***	
RMS Difference		RMS Difference	
Camera 2:	1.90822997	Camera 2:	2.0170699
Camera 7:	1.51113157	Camera 7:	1.398292219
Camera 9:	1.61306051	Camera 9:	1.294371381
Average:	1.67747401	Average:	1.569911167
RMS Standard Deviation		RMS Standard Deviation	
Camera 2:	1.36218249	Camera 2:	1.734125322
Camera 7:	1.31412209	Camera 7:	1.010302621
Camera 9:	1.18591313	Camera 9:	0.752976948
Average:	1.287405904	Average:	1.16580163

Table 3

The second comparison was between the three different file compression algorithms. In this test no "truth" data was provided so the three file compression algorithms (ALS, MPEG, and JPEG) were compared against each other. The data was differenced: ALS minus JPEG, ALS minus MPEG, and MPEG minus JPEG. These statistics were computed and are compared in table 4. Again, the RMS of the horizontal and vertical difference and standard deviation, in pixels, for each camera was computed. Then the total RMS difference was computed.

Table 4: File Compression Difference Comparison

ALS - JPEG RMS Difference		**ALS - MPEG Difference**		**MPEG - JPEG Difference**	
RMS Difference		***RMS Difference***		***RMS Difference***	
Camera 2:	2.28016	Camera 2:	1.26694	Camera 2:	2.20691
Camera 7:	1.59107	Camera 7:	1.08917	Camera 7:	1.62490
Camera 9:	1.59481	Camera 9:	0.97779	Camera 9:	1.70779
Average:	1.82201	Average:	1.11130	Average:	1.84654
RMS Standard Deviation		***RMS Standard Deviation***		***RMS Standard Deviation***	
Camera 2:	1.64424	Camera 2:	1.73349	Camera 2:	1.20843
Camera 7:	1.17548	Camera 7:	1.28548	Camera 7:	1.04270
Camera 9:	1.05024	Camera 9:	1.09638	Camera 9:	0.80932
Average:	1.28998	Average:	1.37178	Average:	1.02015

Additional analysis was performed on the best estimate of trajectory (BET) data. A spreadsheet was presented that appeared to show an increase in noise and mean radial error (MRE) in data provided by analysis after the installation of the compression hardware. After reviewing this spread sheet, similar analysis was performed on the data provided by this experiment in an attempt to mimic the presented output. The following table shows the results. This data is shown in feet on Table 5.

Table 5 – Apogee Data

Data from Apogee Experiment Dec 2003 / Jan 2004

	MRE	Dev MRE	RMS Sigma	Mean Sigma
ALS	1.9317	0.3864	1.1138	1.0053
JPEG	1.6547	0.4912	0.9479	0.8573
MPEG	1.7063	0.4812	0.9813	0.8862

Finally the data from the last two runs was analyzed. This comparison was done with the same process as described above with the results shown in Table 6.

Table 6 – JPEG and MPEG Comparison

ALS - JPEG RMS Difference		ALS - MPEG Difference	
RMS Difference		***RMS Difference***	
Camera 5:	0.95682	Camera 5:	1.01454
Camera 9:	0.13248	Camera 9:	1.24421
Camera 11:	0.25173	Camera 11:	0.30873
Average:	0.44701	Average:	0.85583
RMS Standard Deviation		***RMS Standard Deviation***	
Camera 5:	0.93863	Camera 5:	0.96306
Camera 9:	0.40789	Camera 9:	0.51681
Camera 11:	0.45977	Camera 11:	0.49513
Average:	0.60209	Average:	0.65833

The averages listed above were then compared to the averages listed in table 4. Comparing the ALS – JPEG data there was an improvement of 408% in average error size and of 214% in standard deviation. Comparing the ALS – MPEG data there was an improvement of 130% in average error size and of 208% in standard deviation.

CONCLUSION

As can be seen in the testing results, there is no significant difference, that is, a difference greater by more than an order of magnitude than the one seen between operators, in the RMS difference or RMS deviation in any of this data. The differences observed in the raw data are on the order tenths of pixels for both RMS difference and RMS deviation. In the BET the differences for MRE, deviation in MRE and both methods of computing average sigma are on the order of tenths of feet. Both analysts expressed that it was difficult to impossible to tell the difference between these file formats while viewing and scoring the data. Thus, based upon the data from this experiment, there is no measurable difference between video stored and transmitted in ALS, JPEG or MPEG formats.

REFERENCES

Bergeson Sr., Guy N., "Apogee Test Analysis," Associate Analyst, JT3

Tudor, P.N., "MPEG-2 Video Compression," Electronics & Communication Engineering Journal, Volume 7, Issue 6, December 1995, pp. 257-264.

Woolridge, G. Jesse, "Video Digitizing," Apogee Labs, Inc., North Wales, PA, April 2004.

Philips SAA7114H Data Sheet, PAL/NTSC/SECAM Video Decoder.

Philips SAA6752HS Data Sheet, MPEG2 Video and MPEG Audio Encoder with Multiplexer.