MISB Motion Imagery Standards Board	**MISB EG 1011**
Engineering Guideline	
LVSD Video Streaming	**2 August 2010**

1 Scope

This Engineering Guideline (EG) provides guidelines for the encoding and streaming of Large Volume Streaming Data (LVSD) video sub-windows (regions-of-interest) to FMV clients. This scope is denoted by the large pink ellipse in Figure 1.

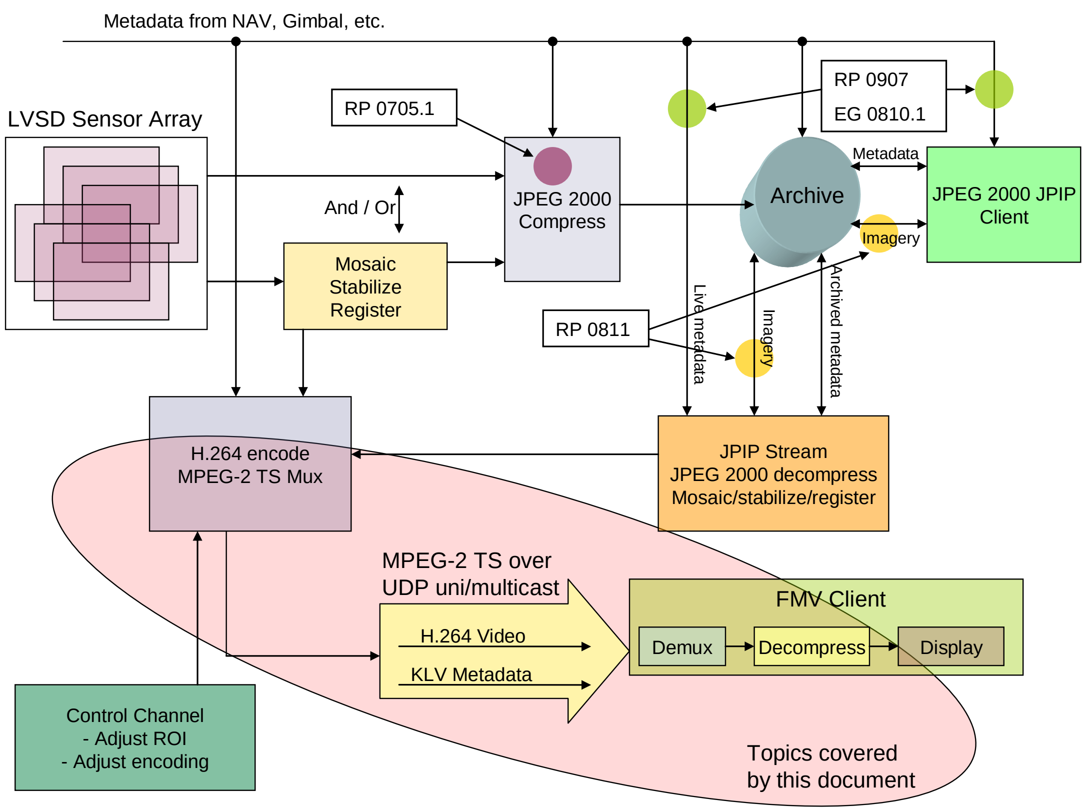

Figure 1: LVSD Data and Protocols Standards Landscape

Other LVSD-specific MISB documents are indicated in Figure 1. RP 0705 [1] defines a profile for the JPEG-2000 encoding of archived imagery. RP 0705 makes recommendations that are slightly different from the prior NPJE and EPJE profiles in order to enable more efficient streaming. RP 0811 [2] defines a profile for JPIP (JPEG-2000 Interactive Protocol per ISO 15444-9 and ITU-T T.808) streaming of such imagery. RP 0811 specifies that JPIP servers on LVSD systems must support full featured JPP (precinct based, as opposed to JPT, tile based) streaming. RP 0907 [3] will define a mechanism (RTP) for live streaming of the metadata required by a JPIP client while EG 0810 [4] defines the content of that metadata. Live metadata delivery to a JPIP client is to be distinguished from both non-live metadata delivery and the metadata embedded within a video sub-window stream.

There are currently gaps in the LVSD standards landscape. MISB plans on gradually addressing each of them. For example, a control channel for control of FMV is drawn in Figure 1, exists in functioning LVSD systems, and is assumed in this document – but there is no standard definition of this channel. To avoid overloading use of the word channel, we will from this point refer to this as the VCI, video control interface. MISB will work to define a VCI in the near future. Another gap is a standard for retrieval of LVSD metadata from an archive, where a query-response data flow is needed rather than a live stream.

Figure 1 shows only the flow of LVSD pixel data and metadata describing the pixel data. Information derived from analytics on these data is not shown. MISB is currently undertaking to define higher level derived information such as tracks and activities. Standards relating to these are expected.

Figure 1 is abstracted to represent multiple different LVSD system designs. A given LVSD system would likely implement a subset of the data flows indicated in the diagram. For example, a system may choose to archive raw sensor data versus a mosaiced product. The archive in Figure 1 could be any post-encode storage, whether DRAM, solid state, or disk. In may be either on-platform or on-ground.

This document focuses on the encoding of pixel data. Metadata within an FMV stream is, for the most part, addressed by existing MISB documents.

2 References

[1] MISB RP 0705.1 *LVSD Compression Profile*, November 2009

[2] MISB RP 0811 *JPIP Profile (Client/Server Functions*), September 2009

[3] MISP RP 0907 *LVSD Metadata Streaming, Draft*

[4] MISP EG 0810.1 *Profile2: KLV for LVSD Applications*, June 2010

[5] MISB EG 0902, *MISB Minimum Metadata Set*, 14 May 2009

[6] MISB EG 0801, *Profile 1: Photogrammetry Metadata Set for Digital Motion Imagery*, 14 May 2009

[7] MISB STD 9601 *Standard Definition Digital Motion Imagery, Compression Systems*

[8] MISB MISP 5.5, *Motion Imagery Standards Profile Version 5.5*, May 2010

[9] MISB EG 0802.1, *H.26 4/AVC Coding and Multiplexing, September 2010*

[10] MISB STANDARD 0604.1, *Time Stamping Compressed Motion Imagery,* September 2010

[11] MISB Standard 0404 *Compression for Infrared Motion Imagery,* September 2009

[12] Young, D., Bakir T., Butto R., Duffield C., Petitti F., "Loss of Interpretability due to Compression Effects as Measured by the New Video NIIRS", Proc SPIE 7529, May 2010

3 Definitions

Not Allowed: Any data representation or encoding that is *not allowed* in this profile *shall not* be implemented by *any* compliant implementation. Note that *not allowed* and *shall not* are equivalent.

Not Recommended: Any data representation or encoding that is *not recommended* in this profile *may* be implemented by any implementation. Implementations are *encouraged to not implement* data representations or encodings that are not recommended unless there is an operational need. Note that *not recommended* and *should not* are equivalent.

Optional: Any data representation or encoding that is *optional* in this profile *may* be implemented by any implementation. It is left to each specific implementation as to whether or not to use the particular data representation or encoding. Note that incorrect implementations are *never* allowed. Note that *optional* and *may* are equivalent.

Recommended: Any data representation or encoding that is *recommended* in this profile *may* be implemented by any implementation. Implementations are *encouraged to implement* the data representation or encoding but they are *not required* to do so. Note that *recommended* and *should* are equivalent.

Required: Any data representation or encoding that is *required* in this profile *shall* be implemented by all compliant implementations. Note that *required* and *shall* are equivalent.

4 Abbreviations and symbols

LVSD: Large Volume Streaming Data. This is a NATO term that applies to systems which collect very large imagery–on the order of 100 to 1000 Megapixels/frame. LVSD systems typically acquire imagery at 1 Hz or faster and may operate for hours at a time. Within the U.S. these systems are known as WALF (Wide Area Large Format), WAAS (Wide Area Aerial Surveillance), WAMI (Wide Area Motion Imagery), WAPS (Wide Area Persistent Surveillance), etc. A LVSD system may be comprised of a multitude of sensors. They are characterized by the large volumes of data they collect; far greater than available bandwidth typically allows to be transmitted in real time. LVSD systems also frequently engage in “non-directed” collection activities.

FMV: Full Motion Video. Often used synonymously, although incorrectly, with “traditional video”. FMV by definition within the MISP must contain metadata. An LVSD system is not an FMV system, but may produce FMV streams.

JPIP: JPEG-2000 Interactive Protocol

ES: Elementary Stream

GOP: Group of Pictures, defined pattern for repetition of I, P, B frames

ROI: Region of Interest, used interchangeably with the term “window”

SEI: Supplemental Enhancement Information, Annex D of H.264 spec, defines header elements

TS: Transport Stream

VUI: Video Usability Information, Annex E of H.264 spec, defines optional header elements

VCI: Video Control Interface – discussed in this document, controls a video sub window

5 Introduction

Streaming an MPEG-2 TS video stream from an LVDS system has a great deal of commonality with streaming from traditional FMV systems. In fact, this commonality is a driving factor in the desire to deliver LVSD data via MPEG-2 TS. Interoperability with the large installed base of MPEG-2 TS clients and other infrastructure, including existing MISB standards, provides a compelling reason to utilize MPEG-2 TS.

However, LVSD systems do introduce differences relative to traditional FMV. These distinctions, and recommendations related to them, form the body of this document. These differences, which will each be discussed in more detail, are the existence of a VCI (video control interface), low temporal frame rates, the potential for higher bit depth, and the likelihood that the FMV stream will be transported over a channel shared with other FMV streams.

The other MISB-recommended mechanism for streaming LVSD data is JPIP (ISO/IEC 15444-9). This is discussed in MISB recommended practice's RP 0811[2] and RP 0907[3]. JPIP and MPEG-2 TS are complementary, and an LVSD system may support either or both mechanisms. One key advantage in using MPEG-2 TS is interoperability with the already installed infrastructure for FMV. An advantage of FMV compression methods, like MPEG-2 and H.264, when carried in a MPEG-2 TS are reduced bitrates from the inter-frame compression as compared to the intra-frame only compression of JPEG-2000.

JPEG-2000 / JPIP also provide advantages. These include rapid random access to any frame, support for large frame sizes, support for high bit depths[1], and a well-defined image control channel. Since JPEG-2000 is the currently recommended archival format (JPEG DCT is also allowed), once the data is compressed for the archive no additional encoding or decoding is needed to produce a JPIP stream. In other words, a system supporting H.264 requires both JPEG-2000 encoders and H.264 encoders whereas a JPIP-only system requires only JPEG-2000 encoding. Beyond this brief comparison, JPIP will not be discussed in this document.

6 Video Control Interface (VCI)

Traditional live FMV systems have a dataflow similar to broadcast television. A user connects (tunes) to the desired channel and watches the data that is pushed through that channel. An ability to change to a different channel is the extent of the available control.

LVSD systems inherently require a sophisticated VCI. At a minimum, the ability to define a region of interest within the scene that is to be streamed is needed. Once such a control schema exists, it is logical to enrich the schema with control parameters that afford even greater

[1] There are profiles of H.264 that support high bit depths but they are currently not as well supported by the industry as JPEG2000

functionality. Parameters to adjust variations in the video encoding via the VCI are discussed here.

Currently, an LVSD VCI has not yet been standardized. There are no assumptions as to whom or what is operating the VCI. Depending on CONOPS and specific system design, the VCI might be operated by: the end user, software in a client device, a system administrator, software on the platform, and other possibilities. It is likely that the control can originate from multiple such sources. By way of example: the client software requests a stream encoded to match its decoding capability; the end user requests a pan left; an administrator sets a maximum allowed bit-rate; and software on the platform adjusts the ROI to auto-track a moving object.

An overall system design goal is to maintain interoperability with legacy devices while allowing improved capabilities for newer devices. A well architected VCI facilitates such opportunity and a more flexible system. Throughout this document recommendations are made that support a range of options rather than one single option.

7 Metadata

Metadata in LVSD video windows is to be populated as specified by MISB standards 0902[5] (Motion Imagery Sensor Minimum Metadata Set), 0801[6] (Photogrammetry Metadata Set), *etc*. The only LVSD-specific issue is that LVSD video sub-windows are mosaiced image products, rather than raw sensor data. Therefore, all sensor metadata fields are **required** to be filled with data elements from a virtual pinhole camera having the field of view of the video window. Metadata suitable for the raw camera data are **not allowed**. In other words, the metadata must match the pixel data.

8 Video Codec and Profile / Level

H.264/AVC (MISB STD 9601 [7] per MISP [8]) is the MISB recommended video encoding standard for low bandwidth FMV systems (less than 1 Mb/s), and is allowable for other FMV systems. EG 0802[9] provides guidance on its usage for FMV. Within the realm of H.264, the CABAC entropy coding of the main and high profiles provides an approximate 10-20% bit-rate improvement[2] over the CAVLC entropy coding of the baseline profile. However, an overwhelming number of client devices, particularly mobile devices, do not support CABAC. Therefore, support for both CABAC and CAVLC is **required.**

LVSD video region-of-interest encoders are **required** to provide the option to encode using H.264.

LVSD video region-of-interest encoders **may** provide the option to encode using MPEG-2–for backward compatibility with older devices. MPEG-2 encoding is not to be confused with MPEG-2 TS (transport stream) also called Xon2, the container format that MISB recommends for both MPEG-2 and H.264

H.264 LVSD video region-of-interest encoders are **strongly recommended** to provide the option to use CABAC entropy coding.

H.264 LVSD video region-of-interest encoders are **required** to provide the option to use CAVLC entropy coding.

[2] The improvement is content dependent.

The H.264 Baseline, Main, and Extended profiles only support 4:2:0 chroma sub-sampling and not the 4:0:0 needed for single channel imagery. All High profiles, and above, in the standard support 4:0:0, though not all actual implementations thereof do. Single channel (4:0:0) encoding is poorly supported by most H.264 implementations. The method for encoding single channel imagery is to zero out the chroma channels; keeping in mind that chroma is expressed as a signed integer. The bitrate / size penalty for doing this is miniscule, which may account for encoder and decoder vendors not supporting 4:0:0. Although not required, it is preferred that H.264 vendors implement 4:0:0 so that a client may distinguish between single and three-channel content.

H.264 LVSD video encoders are **required** to provide the option to use the default 4:2:0 chroma sub-sampling, even when encoding single channel imagery. H.264 LVSD video encoders **may optionally** support 4:0:0 chroma subsampling (no chroma). When encoding color imagery, H.264 LVSD encoders **may optionally** support 4:2:2 and 4:4:4 chroma subsampling.

Practicalities of actual implementations, such as memory size and processing power, impose limits on the maximum H.264 level that can be supported. Each level, defined in Annex A.3 of the H.264 standard, defines a maximum number of 16x16 macroblocks per frame and per second. Level 3.1, for instance, is the lowest level suitable for 720p. There is no particular level specified as a minimum that must be supported. However, as required by the H.264 standard, given some maximum level to be supported all lower levels below this maximum are required to be supported as well.

A MISM table (MISP 5.5 [6], Table 2.8) has been prepared for LVSD video region-of-interest coding. However, it can only be used as a starting point for deciding upon a suitable frame size for a particular bandwidth. Complicating factors are: firstly, the quality of the pre-processing for stabilization and parallax removal; the better these processes, the less bandwidth that is required for a given frame size. Secondly, whether the video region-of-interest is rapidly panning; a stationary video region-of-interest may be allocated less bandwidth than a moving one, depending on scene. Finally, frame size is constrained by the capabilities of the client devices. For example, it makes little sense to stream a frame size larger than 640x480 to a mobile device with only a 640x480 display.

8.1 Experimental Results: Impact of Rapid Panning

The MISB conducted an experiment to illustrate the impact of rapid panning on compression efficiency. Sample 2 Hz LVSD content was selected with a GSD (ground sample distance) of approximately 12 inches. Four horizontal pan speeds were selected: stationary; 44 pix/frame; 88 pix/frame; and 132 pix/frame. 44 pix/frame was chosen to approximate following a vehicle moving at 60 mph. 88 and 132 pix/frame were not chosen to simulate higher speed vehicles, but rather to simulate the same 60 mph vehicle moving across sensors with GSD's of 6 and 4 inches, although one could interpret it either way. For consistency of content, a single elongated rectangular region was selected in the scene. The 44 pix/frame video panned left to right and then right to left over the rectangle once. The 88 and 132 pix/frame videos panned back and forth over the same rectangle 2 and 3 times respectively. The stationary video was centered in the rectangle.

Figure 2 indicates greater H.264 bitrate increases at lower quality versus minimal increases at extremely high qualities. For "reasonably high" quality, QP28, a 100% bitrate increase is observed for the highest pan speed.

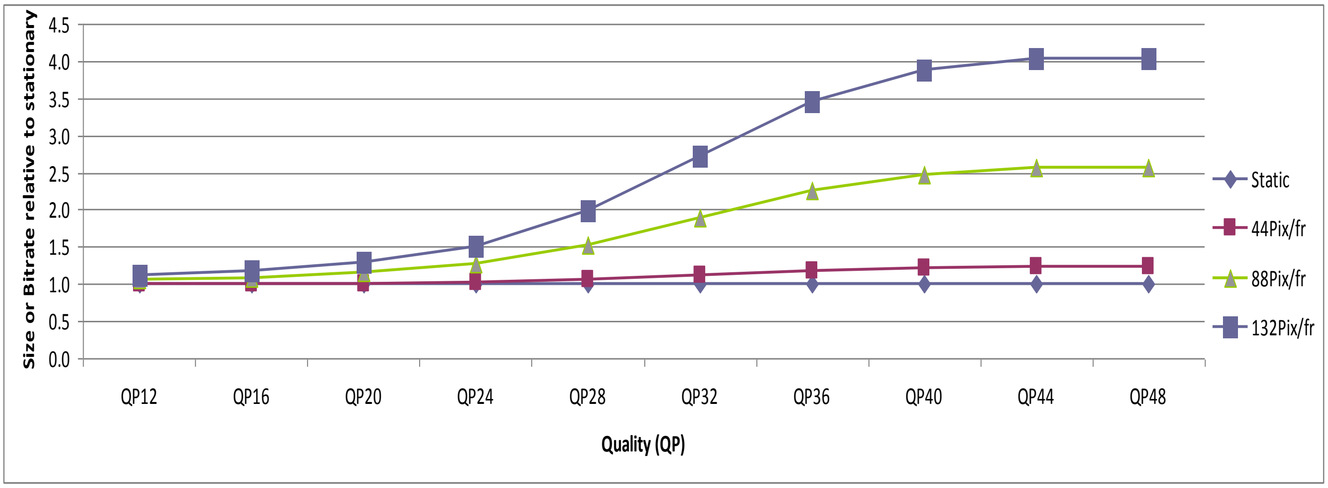

Figure 2: H.264 Bitrate versus Pan Speed

While this experiment used consistent scene content for all pan speeds, real content could mask the effect shown in Figure 2. For example, rapid panning over low complexity agricultural land might require fewer bits to achieve a given quality than a stationary view of a dense urban area would require. Figure 2 is intended to isolate pan speed as a factor in bitrate / quality, but pan speed is not the only factor.

Note on experimental methodology

The above experiment is one of several whose results appear in this document. Each of them was conducted with a set of constant qualities rather than constant bitrates. At each quality, the bitrate / file size is allowed to vary and bitrate, or a normalized version as in Figure 2, is plotted in the results. Charting of bitrate permits concrete statements such as "100% bitrate increase is observed" whereas if bitrates had been held constant and quality had varied, if would have been difficult to make similarly concrete statements about the more abstract concept of quality. Constant quality was achieved by using encoders capable of encoding with a constant quantization parameter (QP). Higher QP's denote lower quality. All experiments discussed in this document use an IPP… GOP structure; B-frames were not used. Unless otherwise stated, a GOP size of 8 frames (duration of 4 seconds) is used in each experiment. Unless otherwise stated, 2Hz 720p content was used for each experiment.

9 LVSD Frame Rate

Many LVSD systems produce imagery at non-standard frame rates; 2 Hz and 10 Hz are common. The H.264 standard is frame rate agnostic. In fact, an H.264 encoder is not even required (by the H.264 standard, MISB does require) to designate a frame rate within the encoded file / stream. Unfortunately, some real-world encoders, multiplexers, streamers, and decoders do make assumptions related to frame rate. More positively, many video products do work correctly at low frame rates. With care, testing, and updates from vendors, non-traditional frame rates can be made to work. Do not assume that an end-to-end system will work with low frame rates without testing it first.

Frame rate and timing may be signaled three different ways (VUI, SEI, and PTS) within an MPEG-2 TS containing an H.264 elementary stream (ES). MISB recommends that all three signaling methods be used simultaneously by encoders and multiplexers for maximum compatibility with downstream streamers, de-multiplexers, and decoders. This topic is discussed in more detail in [9] and [10]. The recommendations below reiterate some of the conclusions from these documents.

An H.264 LVSD encoder is **required** to correctly populate the VUI timing_info fields in the ES.

An H.264 LVSD encoder is **required** to be able to signal constant frame rate in the VUI timing_info_present_flag and fixed_frame_rate_flag flags in the ES. An H.264 LVSD encoder **may optionally** support variable frame rate.

An H.264 LVSD encoder is **required** to correctly populate the per frame SEI pic_timing field in the ES if variable frame rate is used. For constant frame rate, population of SEI pic_timing is **recommended**.

An H.264 LVSD MPEG-2 TS multiplexer is **required** to correctly populate the PTS (presentation time stamp) fields in the TS. (Note: multiplexers will often generate these time-stamps from timing information in the upstream ES, which is why filling in the VUI and SEI is important.)

9.1 GOP structure and size

Because frame rate affects GOP (Group of Pictures) sizing, it is helpful to think about GOP size in terms of seconds rather than frames. For instance, at 30 Hz, a 30-frame GOP size is only one second, but at 2 Hz the duration becomes 15 seconds. Selection of a GOP size is a tradeoff between compression efficiency and the delay in joining a session or channel, among other factors. Large GOP size also impacts video-on-demand functionality. Trick functions like fast forward, seeking, and reverse play all suffer performance degradation as GOP size increases.

Compression efficiency does improve as GOP size increases. This is a simple consequence of P-frames (Predictive) generally being smaller in size than I-frames (Intraframe). A smaller GOP implies more I-frames relative to a larger GOP over the course of the compressed sequence. Better pre-encode stabilization and parallax removal will produce smaller P-frames, thereby affording longer GOP sizes. MISB does **not recommend** the use of B-frames for live streaming because they increase overall sensor-to-display latency. B-frames **may** be considered for archival or other non latency-sensitive applications.

Figure 3 charts the impact of GOP size on bitrate / file size. Sample LVSD sensor Pan EO content of 2 Hz 720p with duration of 1 minute was used to generate this chart. As Figure 3 illustrates, the specific quality chosen impacts the level of the GOP size versus file size curve, but not its shape.

GOP sizes of 1 (I-frame only), 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, and 120 (infinite GOP) are plotted, although the chart is truncated at GOP 25. It is apparent that there is a rapid decrease in file size as the GOP structure transitions from I-frame only to including P-frames. After a GOP size of approximately 8, there are diminishing returns and file size levels off. A GOP size of 8 corresponds to 4 seconds (for 2 Hz imagery). Since the sample content is not perfectly stabilized and has significant parallax and mosaic seams, it is considered an expected worst case of a LVSD system. Content with better pre-processing would push out the point of diminishing returns to a number of frames larger than 8. However, even with well pre-processed imagery, if the video window is rapidly panned the compression efficiency will decrease (see § 8.1).

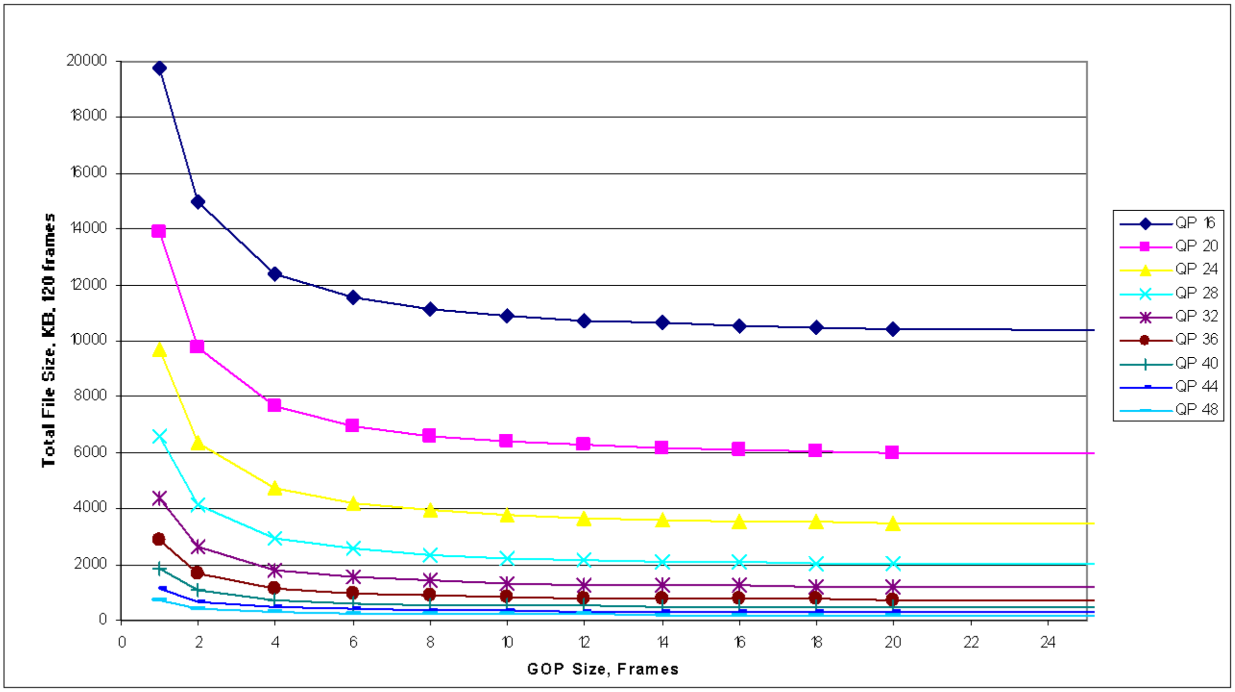

Figure 3: Impact of GOP Size on Compression

As the GOP size increases, the delay in joining a session or channel can become an issue. Channel joining delay is the amount of time from when a client begins to receive a video stream until it displayed. Receipt and decoding of an I-frame is required before a full frame can be displayed. In the 4 second GOP case, it may take up to 4 seconds to "tune" to the channel with an average delay of half that, or 2 seconds. This is comparable to the time it takes to change channels on a digital television or set-top box. Channel joining delay is to be distinguished from end-to-end latency – once a channel is successfully joined, the end-to-end latency is a function of the system stream processes and network, but should in general be low in comparison. For well stabilized imagery, it is possible that the point of diminishing bitrate improvement will occur past the point of acceptable joining delay, and a tradeoff between compression ratio and joining delay will have to be made.

Beyond a certain threshold, channel joining delay becomes intolerable to the user. Although this threshold cannot be dictated, awareness of the issue enables an educated choice. There are two suggested techniques to mitigate channel joining delay. First, LVSD FMV clients can provide feedback to the user in distinguishing between the delay to join a session and a possible communications failure. Feedback to indicate that reliable communications is established and that the stream will display momentarily can make the wait time to join more tolerable. Second, MISB **recommends** the use of intra-refresh, to be discussed next.

An H.264 LVSD encoder is **required** to support GOP durations of 2, 3, 4, 6, 8, and 10 seconds. Additional GOP durations/sizes **may** be supported.

9.2 Intra-refresh

Intra-refresh is a feature of H.264 whereby instead of sending discrete I-frames, the I-frames are in effect evenly distributed over the GOP length. In other words, for an 8-frame GOP, intra-refresh mode will send one eighth of an I-frame each frame rather than sending an entire I-frame as the first frame in the GOP. Note that in this case the encoder is **required** to set the **constrained_intra_pred_flag** to 1 so that the data from inter coded macroblocks is not used for prediction of intra coded macroblocks. This is needed to insure correct decoding of intra coded macroblocks when neighboring inter-coded macroblocks are corrupted. No full I-frames are generated (other than the very first), rather, at least an eighth of the macroblocks in each frame are guaranteed to be Intra-coded. From the point of view of mitigating joining delay, the client may immediately display, in this case, one eighth of a frame; therefore, the user receives quicker visual confirmation that the video is streaming. Intra-refresh also serves to help achieve a more constant bitrate (CBR), when CBR is desired.

Figure 4 depicts a single GOP from a video stream with and without Intra-Refresh. The top left diagram shows compressed frame sizes as produced by a hypothetical CBR rate control mechanism. A CBR rate control mechanism will guarantee that the aggregate size of the frames within a given duration is less than or equal to the target bitrate multiplied by the same duration. Often, rate control will use the GOP size as the duration. However, for long GOP's such as are possible with LVSD, the rate controller may choose a shorter duration for the enforcement of CBR. The top left of Figure 4 demonstrates the expected result of enforcing CBR over durations of 2 seconds, while having a 4 second GOP (2 Hz). The first 4 frames in the GOP, including the I-frame, are of the same aggregate size as the last 4 frames. CBR enforcement over less than GOP sized durations tends to result in the pattern of P-frame size drawn – P-frames immediately following an I-frame are smaller than those later in the GOP. This causes a possible quality fluctuation; the smaller P-frames will likely be of less quality than the larger P-frames. Quality fluctuations may be visually apparent if large enough.

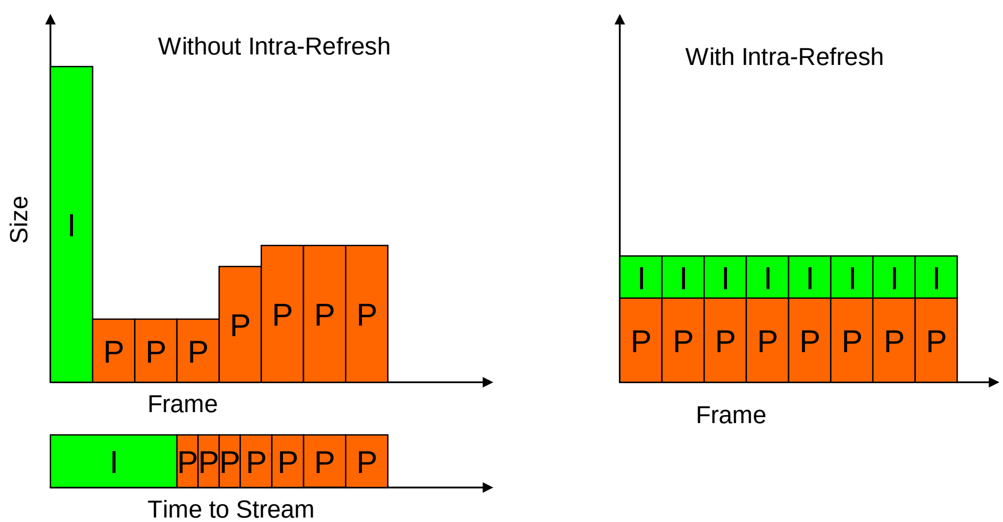

Figure 4: GOP with and without Intra-Refresh

Assuming a fixed bandwidth communication channel, the bottom left diagram shows that it takes a longer time to transmit the I-frame than the subsequent P-frames. This is irrespective of the CBR duration phenomenon discussed in the prior paragraph; it is simply a consequence of I-frames being larger than P-frames. As illustrated, the I-frame takes 3 frame times to stream while the P-frames arrive "bunched-up" at faster than real time. An increase in end-to-end latency is added by the I-frame taking longer than a half second to transmit– in this case an additional 2 frames worth of time; which for 2 Hz material would be a full second.

Both of these detrimental effects are mitigated by Intra-refresh. As shown in Figure 4 to the right, an *I-slice* is generated for each frame and now the frames can be very close to the same size. This reduces (implementation dependent) both end-to-end latency and the visibility of any quality fluctuations.

9.3 Frame duplication for mitigation of frame rate issues

MISB recommends that video content be encoded at its true, as captured, frame rate. However, if portions of the system downstream from the encoder and TS (Transport Stream) multiplexer are incapable of properly handling low frame rates; the H.264 LVSD encoder **may optionally** increase the apparent frame rate by duplicating frames (this is discouraged).

There are two reasons to prefer true frame rate encoding over frame duplication. First, there is no metadata that indicates that frame duplication has occurred. A client would not know that the 24 or 30 Hz stream it is receiving is in reality 2 or 10 Hz – which is the point since duplication of frames is intended to appear as a higher frame rate. A system may misinterpret timestamps in video with duplicated frames. Secondly, frame duplication results in a small but measurable bit rate increase or quality decrease.

Figure 5 shows the size increase of an H.264 stream at 10, 24, and 30 Hz, relative to 2 Hz. At 10 Hz each frame appears 5 times in the stream, *etc*. The increase is minimal at higher qualities (lower QP numbers correspond to higher quality) but for the lowest quality plotted is 41% larger than the unduplicated 2 Hz reference. For the sample content, at QP 32, the size increases by less than 10% at 30 Hz and corresponds to a PSNR of 35db.

All data points plotted in Figure 5 use a GOP duration of 4 seconds, which corresponds to 8, 40, 96, and 120 frames for 2, 10, 24, and 30 Hz respectively. These GOP sizes are even multiples of the number of frames duplicated. If you must duplicate frames, it is **not allowed** to place an I-frame in the middle of a run of duplicate frames. Although not shown here, in duplicating frames the same relation between GOP size and compression efficiency exists as discussed in the GOP structure section – that is, significant improvement up to a GOP size of 4 seconds and then diminishing benefit to extending GOP size further. MISB **recommends** maintaining a GOP duration of at least 4 seconds whether duplicating frames or not. This is consistent with the requirement at the end of § 9.1 that H.264 LVSD encoders support GOP durations of 2, 3,4,6,8, and 10 seconds.

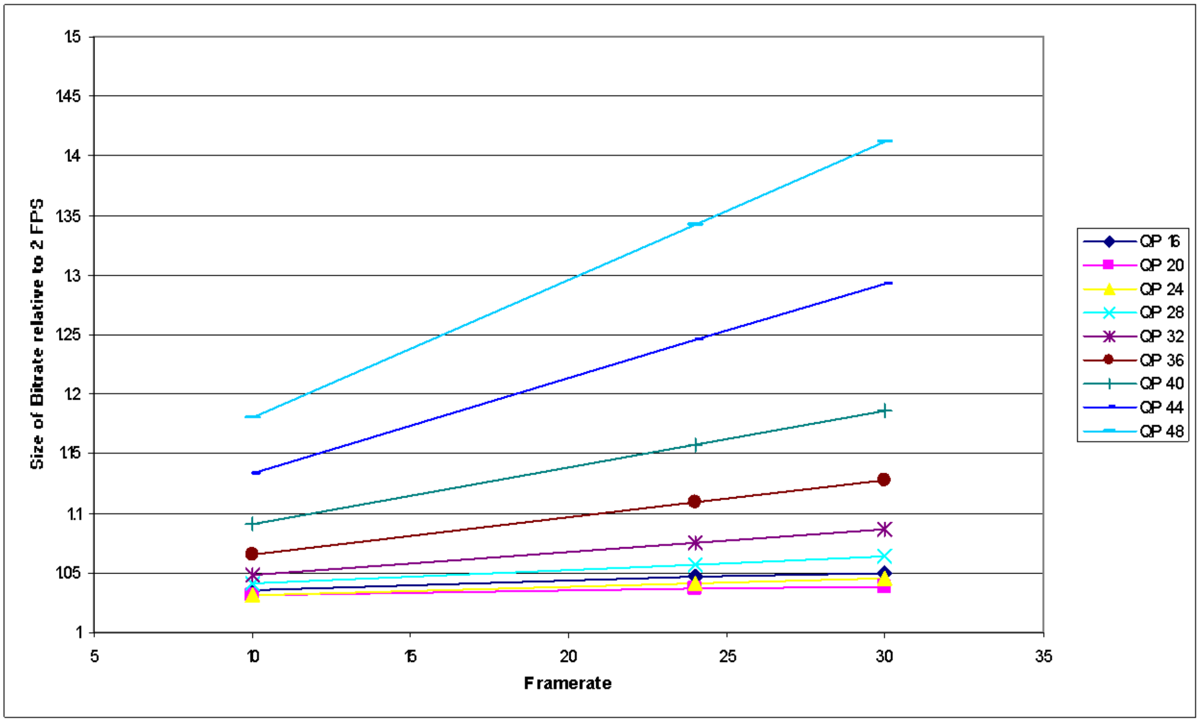

Figure 5: Bit rate increase due to frame duplication

9.4 Impact of frame rate choice on bit rate

Choice of frame rate has an impact on bit rate. An LVSD video window encoder **may** offer a choice of frame rates to clients via the VCI. If such a choice is not available at runtime, a frame rate choice must have been made at design time. Increasing frame rate improves interpretability, particularly the ability to track, but there is a bit rate penalty.

In the absence of sample high frame rate LVSD data, this experiment was conducted with 720p60 ordinary FMV data. The 60 Hz source was decimated to 2, 10, 15, 30, and 60 Hz, and each sequence was encoded at a range of quality levels with a GOP of 4 seconds. The results in Figure 6 are normalized to the 2 Hz version of the content.

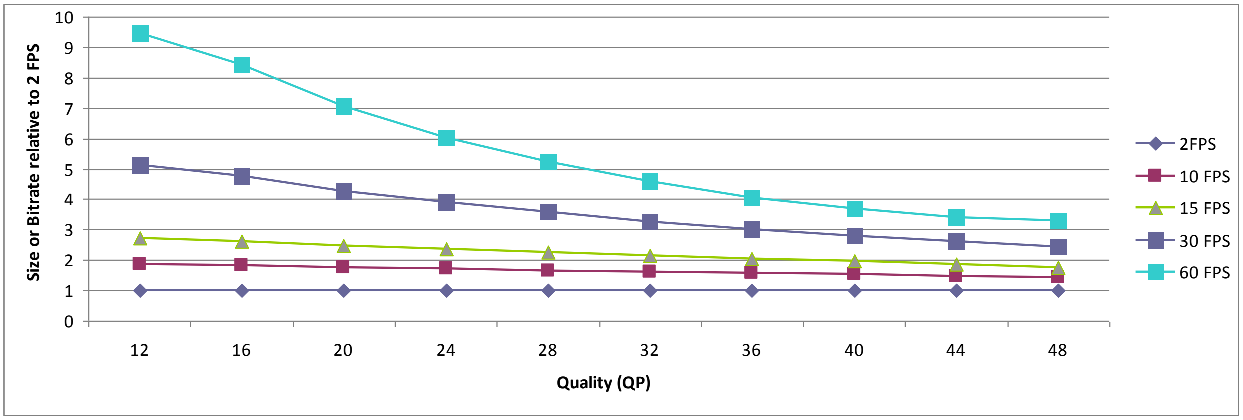

Figure 6: Frame rate versus H.264 Bit rate

Figure 6 indicates the expected result that bit rate increases in a less than linear manner with frame rate. Note the variation introduced by quality – the bit rate increase is more at higher qualities than lower. The high motion content sample data used for this experiment undoubtedly resulted in poorer scaling with frame rate than actual LVSD data would, especially with a stationary ROI. This result may be considered close to worst case. Nonetheless, at the “reasonably high” quality of QP 28, going from 2 to 60 Hz, a 30x increase in frame rate, results in a bit rate increase of only about five times.

10 Bit depth

Infrared LVSD sensors produce imagery with bit depth greater than 8 bits – typically 14. The high quality Pan EO sensors on LVSD systems also often produce greater bit depth imagery. Infrared imagery and high bit depth imagery are discussed in detail in MISB STD 0404 [11] – Compression for Infrared Motion Imagery, from which several main points are reviewed here, along with some LVSD-specific observations.

Whenever possible, it is desirable to maintain the original bit depth of imagery. When the client viewer has access to the imagery’s full bit depth, greater scope is available for operations such as increasing the contrast in an area of shadow. Bit depth reduction may cause important information in the shadow to be discarded. Unfortunately, most H.264 encoders and decoders do not support imagery with bit depths greater than eight bits despite the “fidelity range extensions” to the H.264 standard, which provide support for up to 14 bits, having been available for years.

Therefore, all H.264 LVSD video encoders are **required** to support 8 bit encoding, for compatibility with all decoders.

It is **strongly recommended** that H.264 LVSD video encoders also support 9 through 14 bit encoding, if the sensor is of high bit depth. As high bit depth processing becomes more widespread in the future, it is anticipated that this support will become required. Support for high bit depth may be expressed as support for certain H.264 profiles. The bit depth support of the H.264 profiles is shown below in Table 1.

H.264 LVSD video encoders are **not allowed** to encode more bits than are inherent to the sensor for bit depths greater than 8. In other words, don’t encode 8 or 9 bit content as if it were 10 bit. Odd bit depths such as 9 are permissible in H.264.

MISB does not currently make any recommendation for how to perform bit depth reduction or contrast enhancement. In STD 0404 a variant of histogram equalization is suggested (appendix) as one possibility. However, before implementing any algorithm more complex than simply dropping the least significant bits, be aware of a possible temporal issue. For imagery that is not well stabilized and corrected for non uniformity, there may be significant frame-to-frame differences. Applying a per-frame contrast stretch will tend to amplify these differences and hence negatively impact compression efficiency. Testing sample data with per-frame application of 0404 resulted in a 50% increase in bit rate at the same constant quality settings compared to simply dropping the least significant bits. Even with well pre-processed data, if the video window is undergoing rapid panning, the same issue may arise. It is suggested that temporally filtering any contrast stretch be considered. For example, the statistics for the 0404 algorithm could be gathered for the prior N frames rather than just the current frame. With N on the order of the GOP size, this would minimize frame-to-frame differences induced by the contrast stretch.

Profile Name	9-10 bit support	11-14 bit support
Constrained Baseline		
Baseline		
Main		
Extended		
High		
High 10	Yes	
High 4:2:2	Yes	
High 4:4:4 Predictive	Yes	Yes
High 10 Intra	Yes	
High 4:2:2 Intra	Yes	
High 4:4:4 Intra	Yes	Yes
CAVLC 4:4:4 Intra	Yes	Yes

Table 1: H.264 Profiles with high bit depth support

Table 2 segments decoder capability into three compliance classes. Class A decoders are **recommended**, particularly for clients that provide contrast enhancement capability. Class B decoders will likely be the most popular for mainstream clients.

Class A	Correctly decodes 9 → 14 bit imagery
Class B	Correctly decodes 8 most significant bits of 9 → 14 bit imagery
Class C	Fails to decode 9 → 14 bit imagery

Table 2: Decoder Compliance Classes

Some current H.264 decoders fall into Class C when presented with 10 bit data; that is, they completely fail either through outright crashing of the software or by outputting all-black frames. This is unacceptable since it discourages adoption of high bit depth support by encoders.

Therefore, Class C decoders and clients are **strongly discouraged**.

Preliminary testing with sample data indicates that there is little to no bit rate penalty for compressing all 10 bits of a 10 bit data set to a 10 bit H.264 versus compressing the 8 most significant bits of the same 10 bit data set to an 8 bit H.264. In fact, counter-intuitively, the tested 10 bit H.264 files were nominally smaller than 8 bit files generated from the same high bit depth source data. While this observation is unlikely to be true for all content, and needs further investigation, it is safe to assume that excessive bit rate / file size expansion is unlikely when compressing higher bit depths from low noise sensors[3].

[3] This expectation is also consistent with the results reported by a number of contributions to the JVT committee for high bit depth coding.

11 Shared Communication Channel

LVSD systems are intended to be shared across many users. Ten or more channels of video are anticipated from LVSD systems compared to the one or two channels of a typical FMV system. A channel as used here is a region-of-interest (ROI) taken from the LVSD data that is consistent with typical video formats, such as 1280x720 HD and 640x480 ED as examples. These ROI's are then transcoded into H.264 and packaged with metadata into a MPEG-2 Transport Stream container. All of these channels will share the available air-to-ground communications bandwidth.

When a channel is assigned a fixed amount of bandwidth, the sensible policy is to completely utilize the bandwidth of that channel to maximize video quality. With multiple channels sharing a total link bandwidth, there is a tradeoff between video quality per-channel and the number of channels possible.

For a detailed discussion of the quality needed to meet a given task, please refer to "Loss of Interpretability due to Compression Effects as Measured by the New Video NIIRS [12]" and other work from the NGA Interpretability, Quality, and Metrics Working Group (IQMWG).

The simplest bandwidth allocation strategy is to divide the total link capacity into an equal amount of bandwidth per video channel. However, since all channels may not be active simultaneously this is not the most efficient usage of the link bandwidth. MISB suggests allowing for dynamic allocation as the number of active channels changes rather than fixing the per-channel bandwidth based on the maximum number of channels. While dynamic bandwidth allocation and statistical multiplexing strategies optimize usage of the available bandwidth, it is recognized that they are more difficult to implement than fixed strategies. Possible dynamic strategies include: allowing a channel to temporarily exceed its maximum bandwidth for short bursts; allocating less bandwidth to clients with low resolution displays; and allocating less bandwidth to stationary ROI's monitoring rather static imagery versus actively panned ROI's.

Use of intra-refresh, discussed earlier, can help achieve a more constant bit rate per-channel with less quality variation.