MISB MOTION IMAGERY STANDARDS BOARD	MISB TRM 1008
Technical Reference Material	**2 December 2009**
ISO Base Media Format in the NSG	

Abstract

ISO Base Media Format (IBMF) has garnered a significant share of the commercial market for media distribution. This file format and its derivative relatives may be of use to the government. The purpose of this document is to portray the strengths and weaknesses of this technology, place it in context with other NSG-supported file formats, and provide recommendations for potential use in the NSG.

Overview of the IBMF Family

The ISO Base Media Format is a generalized media container file format. It sits at the center of a constellation of related formats:

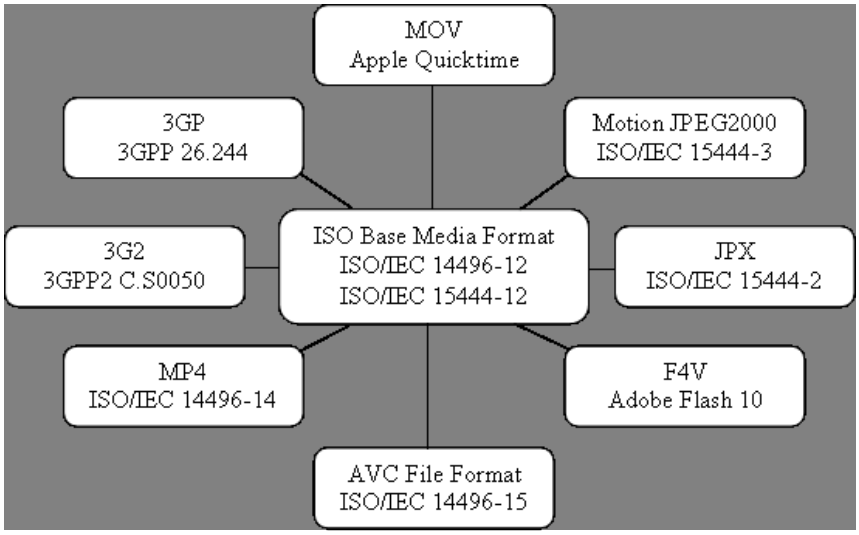

Figure 1 – Various IBMF-derive file formats

History

The ISO Base Media Format (IBMF) is a generalized file format and a descendant from Apple's QuickTime format. Both the ISO JPEG and MPEG subcommittees have agreed upon it as a standard (Part 12 of ISO/IEC 15444 and Part 12 of ISO/IEC 14496). IBMF and its derivatives have been adopted across the computer, mobile phone, and media industries. For example, Apple supports both MP4 files and the closely related QuickTime Movie files (MOV) within the QuickTime API. Adobe has chosen IBMF as the basis of their F4V specification, supported in Flash version 10 and higher. The Third Generation Partnership Project (3GPP) and 3GPP2 consortiums have derived their 3GP and 3G2 file formats from IBMF, with particular focus on the needs of video distribution for mobile phones. Motion JPEG2000 and the advanced JPX format are based on IBMF. IBMF has been adapted to suit the needs of these diverse media technology markets.

The Digital Video Broadcasting Project (DVB) has written several BlueBook documents featuring IBMF. These documents include DVB BlueBook A084 *Specification for the use of Video and Audio Coding in DVB services delivered directly over IP protocols*, DVB BlueBook A121, *The DVB File Format*. A084 provides guidance on streaming support via MP4 or 3GP files. A121 defines a closely related derivative file format (extension .dvb). These documents are currently on the standards track within the European Telecommunications Standardization Institute (ETSI).

Utility within NSG

The National System for Geospatial Intelligence (NSG) has needs in common with the commercial world, and some unique needs of its own. This section identifies areas where IBMF could find utility in the NSG, along with necessary adaptations.

Storage versus Delivery

To appreciate the potential utility of IBMF in the NSG, it is necessary to consider its relationship with the MPEG-2 Transport Stream (TS), Advanced Authoring Format (AAF) and Material eXchange Format (MXF) container formats. Each is designed to package video, audio, and metadata elementary streams, and to associate these streams temporally. As a pure storage vehicle all three serve this purpose. For file download, where a file is delivered over the guaranteed protocol TCP/IP, all serve similar roles.

MPEG-2 TS is a widely used protocol to carry media–most commonly video and audio in commercial broadcast cable delivery, mobile TV, and now Blu-ray high definition DVD. Within the NSG, MPEG-2 TS carries video, audio, metadata and other information.

MPEG2 TS is designed for constant delay networks and robustness in the presence of channel noise and interference, such as is typical in a broadcast application. Thus, for transmission mediums that are point-to-point circuit based, such as an RF link, MPEG2 TS is a preferred protocol. Neither MXF/AAF nor IBMF are designed for this purpose. This postures MPEG-2 TS as a delivery protocol rather than a storage protocol, which was its intent. So, the comparison of TS with AAF/MXF and IBMF is not an apples to apples one, and TS serves a function within the NSG that IBMF will not replace on its own.

AAF and MXF are used in certain studio editing environments (for example, the Public Broadcast Service). Commercial tools supporting MXF include those from Avid, Adobe, and IBM among others. Several major camera manufacturers (Sony and Panasonic) build products which produce MXF files. Within the NSG, AAF is used to archive content in its libraries. The ASPA browser provides the means to import content into AAF format for the NSG.

The ISO Base Media Format (IBMF) is enjoying wide support in the consumer markets, and more recently the commercial market as well. The design of IBMF was intended for both storage and delivery of multimedia content with the following goals in mind:

- Interchange – self contained, protocol independent
- Content Creation – component streams externally stored
- Streaming – interleaved media, hint tracks
- Local Presentation – CD or DVD with random access

As a storage mechanism, IBMF has been adopted by the DVB community to store MPEG-2 TS files. These files are accompanied with information called “hint tracks” that aid in serving out a particular TS file. More recently, Microsoft has adopted IBMF as a container for their adaptive streaming technology, where a file is broken into smaller sub-files at different bit rates. Any sub-file at any stored bit rate can be pulled from the IBMF container dependent on a client receiver’s bandwidth.

Metadata Carriage

Metadata and its handling may be the greatest point of departure between what is supported by commercial off-the-shelf products, and what is required for NSG compatible systems. Consumer media, such as an MP3 audio file, typically contains metadata fields such as author, copyright, and genre that globally apply to the whole file. On the other hand, NSG data generally carries precise geo-positioning information, timestamps, viewing angles, and other important metadata items that vary with time. Some efforts have been made in the consumer space to support synchronized timed metadata with media (for example, song lyrics as text along with the audio), but this is not yet widespread, nor is it clear that this will be sufficient for NSG use.

IBMF supports several types of metadata. Within an IBMF file, metadata may be included at the file, movie, track, or sample level. Generally, metadata at the file, movie or track level applies to the respective structure in an untimed way, while metadata at the sample level is timed. Within the 2008 version of ISO/IEC 14496-12, two specific mechanisms exist for sample level metadata: XMLMetadataSampleEntry, and TextMetaDataSampleEntry. These two are subtypes of the more general MetaDataSampleEntry box, which allows for future expansion.

MPEG-7: an ally of IBMF

MPEG-7, an XML based metadata encoding system standardized by ISO, enjoys first-class support within the IBMF family of standards. The principle means of metadata encoding within MPEG-7 include Binary-encoded XML (BiM) and Text-encoded XML (TeM). MPEG-7 also affords carriage of metadata schema (dictionaries) along with the actual values.

MPEG-7 use is still in its infancy. Several commercial products and research projects have been implemented using MPEG-7 to encode metadata. Particular attention has been devoted in the research community to the definition and use of low-level audio data descriptions. Such metadata has the potential to support novel multimedia search capabilities (for example, looking for sounds which are similar to a given sample). Though not yet widely used, MPEG-7 is notable for its innovative features and its close association and integration with IBMF.

KLV: a foundation of the NSG

Within the NSG, metadata is encoded as Key-Length-Value (KLV) binary data packets and packaged in an MPEG-2 transports stream, or stored within an AAF/MXF container. The US government, allied governments, and industrial partners have put significant effort into defining, implementing and testing standards to support interoperable KLV metadata across systems and manufacturers. Towards this end, in considering the inclusion of any new file format, it is necessary to understand how it supports KLV metadata.

IBMF support for KLV metadata has been proposed in a draft amendment to ISO/IEC 14496-12; however, this is not yet standardized. The planned mechanism is for inclusion using SMPTE URI labels consistent with IETF RFC 5119. This draft amendment proposes new box types (the URIBox, the URIInitBox, and the URIMetaSampleEntry) that can hold timed KLV metadata. Per RFC 5119 encoding, the 16-byte SMPTE key or Universal Label (UL) is encoded as a text

string[1] in the URIBox. Additional guidance regarding the use of the URIInitBox for storage of configuration information necessary to decode a given KLV snippet is given. For example, the Primer Pack construct used in MXF files could be encoded in the URIInitBox.

This proposed amendment offers a good degree of interoperability between IBMF and MPEG-2 TS for KLV metadata. The amendment, or an equivalent, must be approved before IBMF and KLV can be standardized and available to reference. The option of using MPEG-7 encoded metadata with IBMF instead, while possible today, would require more associated costs to convert KLV systems to use MPEG-7, and would offer few clear benefits. That said, if MPEG-7 becomes widely adopted in the commercial world and begins to make inroads into motion imagery workflows, it should be reconsidered for use within the NSG.

Streaming

Streaming is widely used in distributing video content within the commercial broadcast markets and the consumer internet. Various streaming mechanisms are currently in use–many proprietary. The real-time transport protocol (RTP) is a standard registered with the Internet Engineering Task Force (IETF) for streaming media over Internet protocol (IP). Profiles for carriage of different media types (for example video and audio compression) are defined and published for use by the IETF. Some RTP profiles are tailored for specific markets (e.g. MMS for cell phones).

Streaming is a central theme of the NSG Objective Video Architecture (NOVA). The NSG is attempting to move from traditional file-based methods for motion imagery to stream-based workflows that provide dynamic real-time deliveries characteristic of temporal motion imager sequences. Towards this end, infrastructure technologies that enable the management and delivery of streamed content are keys in providing a seamless efficient motion imagery workflow.

Some main features of streaming include:

- Delivery of live content, such as that from a UAV, as it happens
- Uses only network bandwidth that it needs
- Occupies no space on the client's disk
- Provides random access to content
- Allows for optional streaming tracks to be included
- Requires a streaming server over a standard web server
- Immediate viewing of content (minus infrastructure latencies)
- Can be broadcast or multi-cast

When streaming a live event the content is typically pushed from the source in a format that a receiver can understand. In the case of UAV video, for example, the video may be packaged together with metadata within a MPEG-2 Transport Stream at the encoder and subsequently streamed over UDP/IP on the ground. In this case, any transmission problems in the channel will result in dropped packets, which in turn will cause loss of quality (up to and including dropped

[1] Example: urn:smpte:ul:060E2B34.04010103.04010202.01011100

frames). There is no feedback to the source from the client to regulate the bit rate and affect the quality of service of the received stream.

In contrast to real-time event streaming, video-on-demand assumes that content has been previously stored on a streaming server, and that the client and server can now interoperate to determine the optimal way to deliver that content. Channel conditions and viewer interactivity can now direct how a stream is delivered. In these streaming situations, a file container is used to store the content for subsequent delivery, and so a file structure that facilitates access to the data when needed is critical to proper streaming operation.

RTP: an Open Standard

RTP (the protocol) is described in detail in IETF RFC 3550 / IETF Standard 64. RTP requires sending elementary streams (video, audio, etc) as individual RTP data streams over separate network ports. The multiplexing is handled by the network layer, and is not predetermined by storage in a given container format. Therefore, using a media file container with RTP requires that a streaming server, which is standard server with a special piece of software, separate out the individual component elementary streams, and packetize the data into RTP packets upon transmission.

RTP provides codec-specific guidance for forming elementary stream packets. Detailed guidance for video and audio encoding within RTP is provided in IETF RFC 3551. Specific details for H.264 video are found in IETF RFC 3984. The MISB has drafted an IETF RFC for carriage of KLV metadata over RTP, which is to be submitted to the IETF for approval.

IBMF supports RTP streaming with a 'hint track' construct. A hint track is a collection of track references that guide a streaming server through the process of repacketizing the elementary stream data. Among other purposes, hint tracks support trick-play functions, such as fast forward and rewind. The packet size is an input to the hint track. It is possible to create a number of hint tracks, each with different packet sizes. Further information (such as the SAP/SDP session announcement/description information) is stored in user data boxes within the hint track. Additionally, different timescales may be used (for example, 90 kHz for MPEG-2 system clock compatibility) with optional data fields present in the IBMF RTP support data structures.

HTTP: gaining in popularity

HTTP streaming is an alternative technology to RTP and is widely used for streaming video over IP. An advantage of HTTP streaming is simplicity – the server is a standard web server, while the client requires less "intelligence" than a RTP receiver. With HTTP streaming, only one network port is needed: the common port 80 that allows most web traffic through a firewall. The downsides of HTTP streaming include buffering delays, non-smooth playback, and non-standard communication protocols over HTTP.

A properly constructed IBMF file may be streamed via an HTTP 1.1 server, without further processing. To support this, it is necessary for the IBMF file to be self-contained (no external essence references), written in time order, and to have a movie atom at the beginning of the file.

Custom viewing clients will be necessary to support metadata, as well as features, such as random access when using HTTP streaming.

Software Availability

ISO Base Media Format and its derivatives are implemented in a variety of software libraries and products. Apple's QuickTime API and product line support creation, editing, streaming, and playback of IBMF files. Adobe's Flash 10 Player supports playback of F4V files (an IBMF sub-format).

Reference software implementations of IBMF are available. At ISO see: http://standards.iso.org/ittf/PubliclyAvailableStandards/index.html. Similarly, one may acquire reference software from MP4RA by sending email to mp4reg@group.apple.com with 'mp4reg' in the subject line.

Open source implementations are also available. The ffmpeg library supports IBMF; it is accessible via the URL http://ffmpeg.org/download.html. Also, http://code.google.com/p/mp4v2/ offers the mp4v2 library (which implements MP4 support).

Recommendations

Strengths and weaknesses are evident when comparing IBMF to MPEG-2 Transport Stream, AAF and MXF containers across the application uses discussed here–particularly, with respect to storage and streaming.

MPEG-2 TS is a well established transport protocol. As a line-of-sight broadcast protocol MPEG2-TS is a good choice for streaming live events. When coupled with RTP and FEC a high-quality delivery of content is possible even on congested or compromised networks. However, MPEG-2 TS offers less utility as a storage vehicle, and thus does not posture well against the more extensive capabilities offered by AAF/MXF and IBMF.

AAF/MXF has made inroads into the post-production world, and the content management and capture markets. While AAF is not a streaming format, MXF was designed to allow immediate viewing of a file in transport between locations; although this type of streaming should not be confused with video-on-demand streaming that supports interacting with the content in a TiVo-like mode, or adaptively changing the bit rate of the content based on channel limitations. So, while MXF can be called stream-able, it is not a streaming file format.

As a pure content container, IBMF offers many of the technical facilities as AAF/MXF; although for handling native video essence and SMPTE KLV metadata, IBMF is less mature. However, both the DVB community and Microsoft have endorsed IBMF as a container for MPEG-2 TS files and streams, which does suggest one possible use in the NSG. IBMF could serve as a generic container for TS streams and other related media streams–even other TS streams. Immersive video, where multiple videos must be synchronized and managed, is one such application where this could provide great advantage.

The ISO Base Media Format (IBMF) file container is designed to support streaming. Special "hinting' tracks help guide the server in delivering content to a client receiver. IBMF should be considered in streaming applications; particularly when used in conjunction with RTP or JPIP. In the consumer market, IBMF is widely supported with many available tools and technologies. The established size of the IBMF community of practice is driving the infrastructure and supporting technology costs downward making IBMF attractive for use.

At this point in time, the community is still evolving its strategies for content management and storage. Stove-pipe solutions are ever present, where each tends to manage their data repositories in non-uniform ways. Today MPEG-2 transport stream is well entrenched and provides a common "language" across platforms, and AAF is positioned as the storage standard for archival. What is lacking is a file format that offers good support for video streaming over networks. IBMF fills this niche.

Tools such as MPEG-7 may one day be a solution to managing our ever-expanding motion imagery data. Building upon well-established standards that are designed to work with our choices in compression, metadata, and transport protocols just makes good sense. IBMF has been adapted and adopted for many different application uses for one reason–it is a good technology to build upon. Microsoft's recent decision to support MP4 within Silverlight 3, and Adobe's choice of IBMF as the basis for F4V within Flash 10 signal a move within the industry towards IBMF as the common denominator format for streaming support. Not only will IBMF find utility in FMV for both content management and streaming, but also in wide area surveillance, immersive video, and other motion imagery sources.

Glossary

MPEG-2	Informal name for ISO/IEC 13818-x standards
MPEG-4	Informal name for ISO/IEC 14496-x standards
MPEG-7	Informal name for ISO/IEC 15938-x standards

Acronyms

3GPP	Third Generation Partnership Project
ASPA	Aerial Surveillance and Photogrammetry Application
AAF	Advanced Authoring Format
API	Application Programming Interface
HTTP	Hypertext Transfer Protocol
ISO	International Organization for Standardization
IEC	International Electrotechnical Commission
JPEG	Joint Photographic Experts Group
JPIP	JPEG2000 Interactive Protocol
JPX	JPEG2000 eXtensions file format
KLV	Key Length Value
MISB	Motion Imagery Standards Board
MPEG	Motion Picture Experts Group
MXF	Material eXchange Format
NOVA	NSG Objective Video Architecture
NSG	National System for Geospatial Intelligence
PBS	Public Broadcasting Service
RTP	Real-time Transfer Protocol
TS	Transport Stream