SMPTE STANDARD

Stereoscopic 3D Full Resolution Contribution Link Based on MPEG-2 TS

Table of Contents

Approved
January 12, 2012

Foreword

SMPTE (the Society of Motion Picture and Television Engineers) is an internationally-recognized standards developing organization. Headquartered and incorporated in the United States of America, SMPTE has members in over 80 countries on six continents. SMPTE's Engineering Documents, including Standards, Recommended Practices, and Engineering Guidelines, are prepared by SMPTE's Technology Committees. Participation in these Committees is open to all with a bona fide interest in their work. SMPTE cooperates closely with other standards-developing organizations, including ISO, IEC and ITU.

SMPTE Engineering Documents are drafted in accordance with the rules given in Part XIII of its Administrative Practices.

SMPTE ST 2063 was prepared by Technology Committee 32NF.

Intellectual Property

At the time of publication no notice had been received by SMPTE claiming patent rights essential to the implementation of this Standard. However, attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. SMPTE shall not be held responsible for identifying any or all such patent rights.

Introduction

This section is entirely informative and does not form an integral part of this Engineering Document.

This document supplies the necessary constraints on coding and transport for full resolution dual image stereoscopic 3D contribution systems relying upon the MPEG-2 Transport Stream (TS). Dual image stereoscopic 3D imaging systems deliver two images (left eye and right eye) that are arranged to be seen simultaneously, or near simultaneously, by the left and right eyes. Viewers then perceive increased depth in the picture, which becomes more like the natural binocular viewing experience.

Users of this document (who are expected to be either producers of content, producers of the equipment used in the stereoscopic 3D "ecosystem," or producers of the compression equipment constrained by this document) must understand a number of system design concerns which are (and must be) out of scope of this document. Among these are the types and rigging of the cameras used, the system timing necessary to ensure both eye's images are kept in time alignment, and many other items. Some of the key expectations are stated in Notes within the body of the document.

The reader should also be aware that this document is "codec agnostic," in that any video codec for which there is a defined mapping into the MPEG-2 TS may be used. These codecs might support 10-bit video or 8-bit video compression or support 4:2:2 video versus 4:2:0. Such choices are left to the implementers.

Since the importance of having precisely matched pairs of images is well documented, camera systems (camera bodies and lenses) should also be the same make and model (if separate units). Any signal processing required to invert images must be accounted for by the system design to maintain proper system timing.

Regardless of how the pair of images are created, it is assumed that they are properly aligned in space and time at the input of the compression chain. This standard therefore describes means by which the compression chain can be as transparent as possible to the time and space alignment of the image pair.

1 Scope

This document specifies how a stereoscopic 3D High Definition (“HDTV”) video contribution system based on the MPEG-2 Transport Stream (TS) performs coding, multiplexing, and decoding. It defines constraints for the input image pair, the bitstream, the multiplexing, timing synchronization, use of a single PCR PID, and signaling, as well as for the video coding and decoder behavior. This document is codec agnostic (i.e., any codec for which there are defined methods for transport via MPEG-2 TS is permitted). The input image pair needs to have the same image structure (horizontal and vertical pixel count, scanning system, colorimetry, and frame rate) and be coincident in time.

2 Conformance Notation

Normative text is text that describes elements of the design that are indispensable or contains the conformance language keywords: "shall", "should", or "may". Informative text is text that is potentially helpful to the user, but not indispensable, and can be removed, changed, or added editorially without affecting interoperability. Informative text does not contain any conformance keywords.

All text in this document is, by default, normative, except: the Introduction, any section explicitly labeled as "Informative" or individual paragraphs that start with "Note:”

The keywords "shall" and "shall not" indicate requirements strictly to be followed in order to conform to the document and from which no deviation is permitted.

The keywords, "should" and "should not" indicate that, among several possibilities, one is recommended as particularly suitable, without mentioning or excluding others; or that a certain course of action is preferred but not necessarily required; or that (in the negative form) a certain possibility or course of action is deprecated but not prohibited.

The keywords "may" and "need not" indicate courses of action permissible within the limits of the document.

The keyword “reserved” indicates a provision that is not defined at this time, shall not be used, and may be defined in the future. The keyword “forbidden” indicates “reserved” and in addition indicates that the provision will never be defined in the future.

A conformant implementation according to this document is one that includes all mandatory provisions ("shall") and, if implemented, all recommended provisions ("should") as described. A conformant implementation need not implement optional provisions ("may") and need not implement them as described.

Unless otherwise specified, the order of precedence of the types of normative information in this document shall be as follows: Normative prose shall be the authoritative definition; Tables shall be next; followed by formal languages; then figures; and then any other language forms.

3 Normative References

The following standards contain provisions which, through reference in this text, constitute provisions of this standard. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this standard are encouraged to investigate the possibility of applying the most recent edition of the standards indicated below.

ISO/IEC 13818-1:2007|ITU-T H.222.0-2006, Information Technology – Generic Coding of Moving Pictures and Associated Audio Information: Systems

SMPTE ST 292-2:2011, Dual 1.5 Gb/s Serial Digital Interface for Stereoscopic Image Transport

4 Definitions and Acronyms

4.1 Definitions

Coincident in Time: With respect to dual video signals for stereoscopic television, this means that not only are the two video signals "genlocked", but that they represent the same moments in time for the image displayed.

Contribution: The term “contribution” in the context of this document means a very high quality link typically between venue and studio where the material is expected to undergo post-production and eventual emission to viewers by other links.

EAV: An abbreviation for End of Active Video.

Genlock: Abbreviation of “[sync] Generator Lock.” Genlock is a technique for locking a device’s internal sync structures (and thus video structures) to a common external reference (a “sync generator”). This is especially important for stereoscopic 3D imaging systems where both eye images need to be aligned precisely both vertically and horizontally in respect to each other.

MPTS: An abbreviation for Multi-Program Transport Stream.

payload_id: Acronym for Payload Identifier. This is a structure defined by SMPTE 352 and used by SMPTE ST 292-2 to signal payload characteristics. Fields in the eye_identification_descriptor( ) are derived from the **payload_id** of the incoming video.

Program: The collection of all elements within the Transport Stream that have the same program number.

SAV: An abbreviation for Start of Active Video.

4.2 Acronyms from MPEG-2 Systems (ISO/IEC 13818-1)

DTS: An abbreviation for Decode Time Stamp.

PCR: An abbreviation for Program Clock Reference.

PES: An abbreviation for Packetized Elementary Stream.

PID: An abbreviation for packet identifier.

PMT: An abbreviation for Program Map Table.

PTS: An abbreviation for Presentation Time Stamp.

TS: An abbreviation for (MPEG-2) Transport Stream.

uimsbf: An abbreviation for “unsigned integer, most significant bit first.”

5 MPEG-2 Transport Stream Constraints

The coded images for both eyes, along with associated audio and data shall be carried in an MPEG-2 multi-program Transport Stream (MPTS), which shall have only two programs, one for the left eye images, and the other for the right eye images. Within the TS, PCR, PTS, and DTS clock sample timestamps shall be present and used by decoders to maintain the output images as coincident in time. There shall be a single PCR PID for both programs.

The coded video images for each eye shall be identified, as to which are left eye images and which are right eye images, by a descriptor as specified in Section 5.1 which shall be included in the descriptor loop immediately following the **ES_info_length** field in the PMT describing that video elementary stream.

Note: This descriptor is constructed per the guidance provided by Annex C.8.6 in ISO/IEC 13818-1.

An implementation conformant to this standard shall be identified by the presence in the PMT of the pair of descriptors specified in Section 5.1.

5.1 Eye Identification Descriptor

The PMT for a given eye's video stream shall be identified by the use of the eye_identification_descriptor( ) defined in this section.

Table 1 – eye_identification_descriptor( ) syntax

Syntax	No. of bits	Format
eye_identification_descriptor( ) {		
descriptor_tag	8	uimsbf
descriptor_length	8	uimsbf
eye_identifier	4	uimsbf
audio_status	4	uimsbf
}		

descriptor_tag – The value for the Eye Identification Descriptor tag is 0xCB.

descriptor_length – This is an 8-bit field specifying the number of bytes of the descriptor immediately following descriptor_length field and shall be1.

eye_identifier – This is an 4-bit field specifying the value 0x0 for left eye program or 0x1 for right eye program. All other values shall be reserved. The value should be derived from information in the payload Identifier of the input video to the encoder.

audio_status – This is an 4-bit field specifying the value 0x0 to indicated that the right eye program status is unknown or don't care, 0x1 to indicated that the right eye program carries a copy of left eye audio, or 0x2 to indicated that the right eye program carries additional channels. All other values shall be reserved. The value should be derived from the payload Identifier of the incoming video.

6 Video Coding Constraints

6.1 PES Alignment

Each video PES packet shall contain the start of only one access unit, as defined in Section 2.1.1 of ISO/IEC 13818-1. In the PES header, data_alignment_indicator shall be '1'. The first byte of a PES packet payload shall be the first byte of a video access unit. Each PES header shall contain a PTS and DTS if DTS differs from the PTS.

The values for the PTS clock samples for a corresponding left eye and right eye image shall be identical within 45 microseconds (which equals 4 clock periods of the PTS/DTS clock.)

Note: For assistance in conversion of MPEG-2 PTS/DTS to absolute time and/or frame time, see SMPTE EG 40.

6.2 PCR PID

As required by Section 5, a single PCR PID shall be referenced by the video streams for both eyes, as well as for associated audio and data services. The PCR PID shall be a unique PID value or the value of either video PID.

6.3 Encoder Constraints

It is recommended that the input video signal sent to the left eye and right eye encoders be compliant with SMPTE ST 292-2. In accordance with SMPTE ST 292-2, the timing difference between the serial digital clocks and EAV / SAV of the Left eye stream and the Right eye stream may differ by up to 400 ns at the source. Any such timing difference shall be removed prior to encoding. When two encoders are used it is recommended that a common synchronizing signal ("genlock") be used.

Coding decisions should be shared between left eye and right eye encoders, including such decisions as repeat-field removal, and picture coding-type decisions. If repeated frames of video are to be removed before compression (e.g., "3/2 pulldown" removal), the frame removal process should be the same for both left eye and right eye images. Pairs of coded pictures should have the same H and V structure (which permits the decoder to recover the input video image structure).

There are no special constrains for the placement of audio and ancillary data embedded in the video input stream beyond those given in SMPTE ST 292-2. The encoder/decoder system shall ensure that the placement of audio and ancillary data on output matches that of the input digital interface. Audio data is normally carried in one or more packetized elementary streams, using a uniquely assigned PID or PIDs values per other standards (see Annex A). Ancillary data may be carried in Transport Stream packets identified by a uniquely assigned PID value per other SMPTE standards.

Note 1: System designers need to assume that both sources will be genlocked, and the images will be coincident in time. This is especially important for stereoscopic 3D where both eye images need to be aligned precisely.

Note 2: Both video encoders need to also be the same make and model.

6.4 Visual Fidelity (Informative)

The visual fidelity of both image streams (both eyes) should match. To that end, the picture structure and picture type should match between eyes. Bitrate allocation for corresponding left eye and right eye coded pictures should be similar.

The coded video quality of the two streams should match between both eyes (with a minor difference between any measured values permissible).

7 Decoder Constraints

Decoders shall utilize PCR, PTS, DTS, and any other appropriate means, to maintain frame accurate match between both left eye and right eye decoded images. Both left eye and right eye decoders shall reference the single PCR specified in Section 5 and Section 6.2 of this document. Both decoders of a pair shall be genlocked and their outputs shall be coincident in time. Decoder video outputs shall match the original video input image structure.

Per the requirements of Section 6.1, the values for PTS and DTS for a given left eye and right eye image are expected to be identical within 45 microseconds (the decoder should present these pictures simultaneously). Designers should be aware that bitwise identical values of PTS or DTS from eye to eye are not expected.

The output video signal from the left eye and right eye decoders shall be compliant with SMPTE ST 292-2 and shall reconstruct the original Video Payload information.

Note 1: Designers should be aware that there is fielded equipment designed prior to the publication of this standard which may send a different PCR PID for each eye.

Note 2: Designers are urged to consult the error recovery discussions in CEB-20 for guidance in handling input TS errors and are urged to coordinate recovery decisions between both decoders of a pair.

Note 3: For assistance in conversion of MPEG-2 PTS/DTS to absolute time and/or frame time, see SMPTE EG 40.

Annex A Bibliography (Informative)

Note: All references in this document to other SMPTE documents use the current numbering style (e.g. SMPTE ST 302:2007) although, during a transitional phase, the document as published (printed or PDF) may bear an older designation (such as SMPTE 302M-2007). Documents with the same root number (e.g. 302) and publication year (e.g. 2007) are functionally identical.

SMPTE ST 302:2007, Mapping of AES3 Data into an MPEG-2 Transport Stream

SMPTE ST 337:2008, Format for Non-PCM Audio and Data in AES3 Serial Digital Audio Interface

SMPTE ST 352:2011, Video Payload Identification Codes for Serial Digital Interfaces

SMPTE ST 425-2:2012, Source Image Format and Ancillary Data Mapping for Stereoscopic Image Formats on a Single-Link 3 Gb/s Serial Interface

SMPTE ST 2038:2008, Carriage of Ancillary Data Packets in an MPEG-2 Transport Stream

SMPTE EG 40-2002, Conversion of Time Values Between SMPTE 12M-1 Time Code, MPEG-2 PCR Time Base and Absolute Time.

CEA CEB-20, A/V Synchronization Processing Recommended Practice, 2009.

Poynton, Charles, “Digital Video and HDTV, Algorithms and Interfaces”, 2003.