SMPTE STANDARD

VC-5 Video Essence – Part 5: Layers

Table of Contents

Approved October 23, 2015

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 its Standards Operations Manual. This SMPTE Engineering Document was prepared by Technology Committee 10E.

Intellectual Property

At the time of publication no notice had been received by SMPTE claiming patent rights essential to the implementation of this Engineering Document. 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.

SMPTE ST 2073-1, ST 2073-3, and ST 2973-4 define the representation of a single image in a bitstream.

One or more images compatible with SMPTE ST 2073-1, ST 2073-3, or ST 2073-4 can be represented in a single VC-5 bitstream using the extensions defined in this standard.

This standard enables multiple images to be represented in a VC-5 bitstream. Each individual image in the bitstream is called a layer. Each layer can be decoded into a separate image. Layers can be used to represent stereo images, bracketed exposures, interlaced fields, and tiled images, among other use cases. An image repacking process (for example as described in SMPTE ST 2073-1) can compose images from multiple layers into a single displayable image.

1 Scope

This standard defines extensions to ST 2073-1 to support the representation of multiple images in a single VC-5 bitstream. Each of the individual images is called a layer. All layers present in the bitstream have the same width, height, number of channels, and image format.

This standard defines mechanisms for representing stereo images, multiple exposure values, interlaced fields, and tiled images in a single bitstream.

This standard is backward compatible such that a decoder that does not support these extensions will decode the first layer in the bitstream into a single displayable image.

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; then 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 engineering document. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this engineering document are encouraged to investigate the possibility of applying the most recent edition of the standards indicated below.

SMPTE ST 2073-1:2014 VC-5 Video Essence. Part 1: Elementary Bitstream.

SMPTE ST 2073-3:2015 VC-5 Video Essence. Part 3: Image Formats.

SMPTE ST 2073-4:2015 VC-5 Video Essence. Part 4: Subsampled Color Difference Components.

4 Terms and Definitions

For the purposes of this document, the following terms and definitions apply.

4.1 Layers

Independently decodable images that share a common bitstream header.

5 Layer Syntax and Structure

5.1 Layer Overview

This standard extends other VC-5 standards to allow multiple images to be represented in a single bitstream by encoding each individual image as a separate layer (Figure 1). This standard is backward compatible with other VC-5 standards: A decoder that is not compliant with this standard will decode the first layer and ignore the remaining layers.

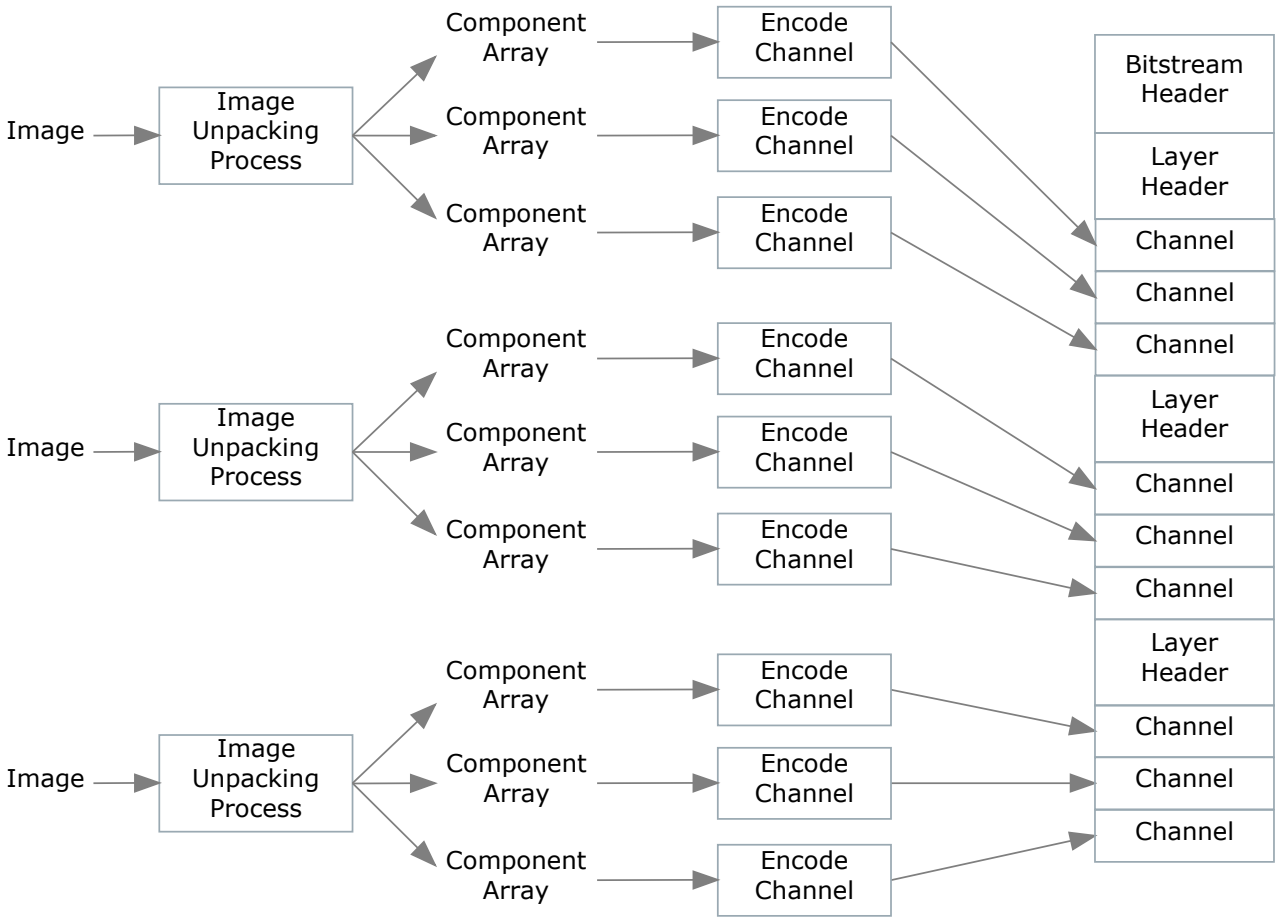

Figure 1. Diagram showing how multiple images are represented as layers in a bitstream.

5.2 Bitstream Header

All layers shall share a single set of bitstream header parameters as defined in SMPTE ST 2073-1, ST 2073-3, and ST 2073-4

NOTE A single set of bitstream header parameters, shared by all layers, implies that the images represented by each layer have the same width, height, and image format.

5.3 Layer Contents

A layer in the bitstream and the bitstream header parameters shared by all layers in the bitstream (see section 5.2) shall comprise the representation of a single image that is compliant with SMPTE ST 2073-1, ST 2073-3, or ST 2073-4.

5.4 Bitstream Layer Structure

Layers may be present in the bitstream in any numerical order by layer number, subject to the layer numbering constraints defined in section 6.2.

5.5 Layer Codeblocks

Layers shall not share codeblocks.

The syntax elements of a layer shall be contiguous. This excludes the bitstream header, as it is common to all layers.

6 Layer Codec State Parameters

6.1 Layer Count

The value of the **LayerCount** codec state parameter is the number of layers present in the bitstream as defined in annex A.2.1.

The **LayerCount** codec state parameter shall be a bitstream header parameter.

The **LayerCount** codec state parameter shall be present in the bitstream.

NOTE The presence of the LayerCount codec state parameter signals that the bitstream is compliant with this standard.

6.2 Layer Number

The value of the **LayerNumber** codec state parameter is the number of the layer starting with the next syntax element in the bitstream as defined in Annex A.2.2.

The value of the **LayerNumber** shall be in effect before the first codeblock of that layer.

Layers shall be numbered consecutively starting with zero.

Layers may be present in the bitstream in any numerical order by layer number, except that the first layer present in the bitstream shall be layer number zero.

Subsequent layers should be present in the bitstream in order by layer number.

6.3 Layer Pattern

The value of the **LayerPattern** codec state parameter specifies the arrangement of layers in the bitstream as one or more of the common layer patterns defined in section 7.

The **LayerPattern** codec state parameter shall be a bitstream header parameter.

6.4 Initial Codec State

After the codec state is initialized as described in SMPTE ST 2073-1, the value of the **LayerCount** codec state parameter shall be 1, the value of the **LayerNumber** codec state parameter shall be 0, and the value of the **LayerPattern** codec state parameter shall be 0.

6.5 Layer State Changes

Immediately after the last codeblock of a layer has been decoded, the **LayerNumber** codec state parameter shall be incremented by 1, the **ChannelNumber** codec state parameter shall be set to 0, and the **SubbandNumber** codec state parameter shall be set to 0.

The last codeblock of a layer shall be the codeblock that begins at the bitstream position that is larger than the bitstream position at the beginning of any other codeblock in that layer.

7 Common Layer Patterns

7.1 Layer Pattern Overview (Informative)

Layer patterns are conventions for predefined layer use cases.

7.2 Stereo Pairs

7.2.1 Organization of Stereo Layers

Layer patterns are conventions for predefined layer use cases

If the left and right images that comprise a stereo pair are encoded as separate layers, then the layer number for the left image of the stereo pair shall be less than the layer number for the right image of the stereo pair.

If a bitstream contains only two layers representing a stereo pair:

(1) Layer number zero shall contain the left image of the stereo pair,

(2) Layer number one shall contain the right image of the stereo pair.

7.3 Multiple Video Streams

7.3.1 Definition of Multiple Video Stream

Multiple video streams are video streams captured by more than one camera or viewpoint, excluding the case of two cameras or viewpoints for a stereo pair which is covered in section 7.2.

7.3.2 Organization of Layers for Multiple Video Streams

This standard does not specify the numbering of layers for multiple video streams.

7.4 Interlaced Video

7.4.1 Organization of Interlaced Fields

If the two fields that comprise an interlaced image are encoded as separate layers, then the layer number for the first field shall be less than the layer number for the second field.

If a bitstream contains only two layers representing the fields of an interlaced image:

(1) Layer number zero shall contain the first field,

(2) Layer number one shall contain the second field,

The first field shall be the top field.

7.5 Bracketed Exposure Images

7.5.1 Organization of Bracketed Exposure Images

Images with higher exposure values shall be assigned higher layer numbers in order of increasing exposure value with the image having the lowest exposure value assigned to layer number 0.

NOTE This requirement allows the decoder to refine a decoded image by decoding additional layers until sufficient detail has been obtained.

7.6 Tiled Images

7.6.1 Organization of Layers for Tiled Images

An image represented by tiles shall have each tile encoded as a separate layer. Tile rows are numbered in order from top to bottom; tile columns are numbered in order from left to right. Tile layers shall be numbered in raster-scan order as defined in SMPTE ST 2073-1.

7.6.2 Tiled Image Example (Informative)

An example of a tiled image with two rows, each row comprising three columns of tiles, is shown in Figure 2.

Figure 2. Diagram showing an example of how tiles are assigned layer numbers.

The arrangement of tiles is out of scope for this standard.

7.7 Chroma Key

7.7.1 Organization of Layers for Chroma Key

Multiple images that are intended to be composited together using an alpha channel or chroma key (such as green screen applications) can be represented as layers. Layers should be ordered from foreground to background in numerical order starting with 0.

7.7.2 Chroma Key Example (Informative)

Applications that use chroma keying typically use RGB or RGBA images. In the case of RGB images, the alpha channel is derived from the RGB components.

It is not possible to mix RGB and RGB(A) images in the same bitstream due to the layer constraints defined in section 5.3.

8 Multiple Layer Patterns

Layers are organized such that:

(1) The sequence of layers that comprise the images from one viewpoint shall not contain syntax elements from the layer for a different viewpoint,

(2) The sequence of layers that comprise one image in a tiled arrangement of images shall not contain syntax elements from the layer for a different image tile,

(3) The sequence of layers that comprise one exposure value shall not contain syntax elements from the layer for a different exposure value,

(4) The sequence of layers that comprise a field from an interlaced frame shall not contain syntax elements from the layer for the other field.

(5) The sequence of layers that comprise an image for chroma key applications shall not contain syntax elements from the layer for a different image for chroma key applications.

NOTE The above constraints apply when the bitstream contains more than one layer pattern.

These constraints on the assignment of layer numbers allow layers to be present in the bitstream in order by layer number and delineated by sections (SMPTE ST 2073-6).

Annex A Codec State Parameters

A.1 Bitstream Extensions

The codec state parameters defined in SMPTE ST 2073-1 Annex B.1 shall be extended by this standard to include the codec state parameters defined in Annex A.2 of this standard.

The tag-value pairs defined in SMPTE ST 2073-1 Annex B.2 shall be extended by this standard to include the tag-value pairs defined in Annex A.3 of this standard.

A.2 Parameter Descriptions

A.2.1 LayerCount

The **LayerCount** codec state parameter shall specify the number of layers present in the bitstream. The minimum value shall be 0. A tag-value pair that represents the **LayerCount** codec state parameter shall use the tag number listed in Table 2 and shall be an optional tag-value pair.

A.2.2 LayerNumber

The **LayerNumber** codec state parameter shall specify the number of the next layer in the bitstream. The value shall be in the range [0, **LayerCount**). A tag-value pair that represents the **LayerNumber** codec state parameter shall use the tag number listed in Table 2 and shall be an optional tag-value pair.

A.2.3 LayerPattern

The **LayerPattern** codec state parameter shall specify the arrangement of layers in the bitstream. The value shall be a bit mask as defined in Table A.1.

The bits in the mask shall be numbered from least significant bit to most significant bit in the value of the **LayerPattern** codec state parameter. Bit number 0 shall correspond to the least significant bit in the value of the **LayerPattern** codec state parameter. If and only if a layer pattern is present in the bitstream, then the corresponding bit in the **LayerPattern** codec state parameter shall be 1.

A tag-value pair that represents the **LayerPattern** codec state parameter shall use the tag number listed in Table A.2 and shall be an optional tag-value pair.

Table A.1. Correspondence between bits in the LayerPattern codec state parameter and layer patterns.

Bit Position	Layer Pattern
0	Stereo pair (section 7.2)
1	Multiple video streams (section 7.3)
2	Interlaced video (section 7.4)
3	High dynamic range (section 7.5)
4	Tiled image (section 7.6)
5	Chroma Key (section 7.7)

A.3 Tag-Value Pairs

The tag-value pairs that represent the codec state parameters defined by this standard in a VC-5 bitstream shall be as listed in Table A.2, with the tag number and default value as listed in that table. The tag-value pairs defined by this standard shall be bitstream header parameters if so specified in Table A.2.

Table A.2. Codec state parameters defined by this standard.

Parameter Name	Tag Number	Default Value	Bitstream Header Parameter?	Reference
LayerCount	120	0	Yes	A.2.1
LayerNumber	121	0	No	A.2.2
LayerPattern	122	0	Yes	A.2.3

Annex B Layer Compositing (Informative)

The individual images can be optionally composited into a single displayable image by an image repacking process.

Layers can be used to represent multiple images in the bitstream. Layers support workflows where separate decoded images are combined into a single image belonging to a single track of essence. This enables support for media frameworks where a single decoder instance can only access the essence in a single track. Multiplexing multiple streams of essence into a single track allows one decoder to composite multiple layers into a single displayable image.

The image compositing operation is not specified by this standard, but can be any operation that combines multiple images into a single image (generalized compositing).

Example uses for layers include:

(1) Multiple exposures can be represented in the bitstream as separate images that can be combined into a single high dynamic range (HDR) image by tone mapping,

(2) Fields from an interlaced image can be represented in the bitstream as separate images that can be composited into a single image by interleaving the rows,

(3) Two images that comprise a stereo pair can be combined into a displayable stereo image in any of the common image formats for stereo presentation including anaglyph, over-under, or side-by-side.

Bibliography (Informative)

SMPTE ST 2073-6:201x VC-5 Video Essence. Part 6: Sections.