SMPTE STANDARD

Depth Map Representation

Table of Contents

Page

Foreword 2
Intellectual Property 2
Introduction 2
1 Scope 3
2 Conformance Notation 3
3 Normative Reference 3
4 Definitions and Acronyms 4
4.1 Reference Camera 4
4.2 Depth Value 4
4.3 Relative Depth Value 4
4.4 Depth Map 4
5 Depth Map Representation 4
5.1 32-Bit Depth Map Representation 4
5.2 16-Bit Depth Map Representation 4
6 Conversion between Representations 5
6.1 Derivation of 16-Bit Relative Depth Value Representation from 32-Bit Depth Value Representation 5
6.2 Derivation of 32-Bit Depth Value Representation from 16-Bit Relative Depth Value Representation 5
7 Metadata 6
7.1 DepthScaleFactor 6
7.2 DepthOffset 6
7.3 Depth Source Type 6
7.4 Depth Remapping Type 6
Annex A Stereo Application (Informative) 7
Annex B Bibliography (Informative) 9

3 Barker Avenue, White Plains, NY 10601
(914) 761-1100

Approved
July 29, 2016

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.

SMPTE ST 2087 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.

Depth information can be useful for improving a number of production and post-production processes. For example, accurate depth information is necessary for the adjustment of the camera point of view during post-production. Depth information can also be used to assist in the compositing of multiple elements in a production that includes live action and CGI, for the proper placement of overlays on multi-view content, and to improve the ability to render stereoscopic content for a wide variety of viewing environments. Depth information could also be included as part of the distribution package for multi-view content where getting the information close to the source is desirable.

Depth information can be derived from a number of sources including animation rendering, multi-camera capture, and on-scene depth measurements. Each captured or synthesized view can have its own depth map. Disparity maps can be more conducive to some operations so a direct conversion between depth and disparity representations is desirable.

It is expected that this information will be carried within a file structure defined in companion document(s).

1 Scope

This standard provides a data representation for depth information. This information allows for simple interchange during production and post-production, and provides the essence for distribution of single-view and multi-view content. The standard specifies a 32-bit floating point representation and a 16-bit floating point representation for depth information.

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 Reference

The following standard contains provisions that, through reference in this text, constitute provisions of this standard. At the time of publication, the edition indicated was 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 standard indicated below.

IEEE 754-2008, IEEE Standard for Floating-Point Arithmetic

4 Definitions and Acronyms

The following definitions are used in this document.

4.1 Reference Camera

The Reference Camera is the camera that corresponds to the particular viewpoint of the image corresponding to the depth representation. The Reference Camera can be a virtual camera such that it might not match a camera used for capture.

4.2 Depth Value

A Depth Value is the distance in meters from the Reference Camera to a point on the surface of an object imaged by the camera. The distance is measured along a line parallel to the optical axis of the Reference Camera. The line originates from the plane that is both (a) perpendicular to the optical axis and (b) contains the center of perspective[1], and ends at the point on the object.

4.3 Relative Depth Value

A Relative Depth Value is an offset and scaled representation of the Depth Value.

4.4 Depth Map

A Depth Map is an array of values corresponding to the pixels in an image from the Reference Camera. Each value in the array represents a Depth Value for the corresponding pixel. Different representations for Depth Values are possible.

5 Depth Map Representation

A Depth Map shall contain one value for each pixel in the corresponding image.

5.1 32-Bit Depth Map Representation

A 32-bit Depth Map is an array of Depth Values represented in the IEEE 754 single-precision (32-bit) binary floating-point format.

A Depth Value of 1 meter is represented as 1.0. (0x3F800000).

A Depth Value of positive infinity is represented as +INF. (0x7F800000).

If the Depth Value is unknown, then it shall be represented as NaN (Not A Number). (0x7FC00000)[2].

For robustness, an implementation should use the specified value for NaN when writing, but should accept any NaN value as NaN when reading.

5.2 16-Bit Depth Map Representation

A 16-bit Depth Map is an array of Relative Depth Values in the IEEE 754 half-precision (16-bit) binary floating-point format.

[1] Also known as “entrance pupil”.

[2] Note: All NaN values have at least the bits corresponding to +INF set, plus at least one more bit set, other than the most significant bit (the msb is the sign bit and may be set for a NaN).

These values and associated metadata (DepthScaleFactor and DepthOffset) are necessary to determine the corresponding Depth Value.

A Relative Depth Value of positive infinity is represented as +INF. (0x7C00).

If the Relative Depth Value is unknown, then it shall be represented as NaN (Not A Number). (0x7E00)[2].

For robustness, an implementation should use the specified value for NaN when writing, but should accept any NaN value as NaN when reading.

If the Relative Depth Value is neither +INF, nor NaN, then its representation shall be in the range -65504.0 to 65504.0 inclusive.

6 Conversion between Representations

The derivations below should be performed in no less than single-precision floating point format.

6.1 Derivation of 16-Bit Relative Depth Value Representation from 32-bit Depth Value Representation

If the Depth Value is NaN, then the Relative Depth Value shall also be NaN.

If the Depth Value is +INF, then the Relative Depth Value shall also be +INF.

If the Depth Value is neither NaN, nor +INF, then compute an intermediate value, Z_{Int} as follows:

Equation 6.1

$$Z_{Int} = \frac{Z}{S} - \frac{Z_c}{S}$$

Where,

Z denotes a Depth Value,

S denotes the DepthScaleFactor,

Z_c denotes the DepthOffset.

If Z_{Int} is greater than 65504.0, then the Relative Depth Value shall be equal to 65504.0.

If Z_{Int} is less than -65504.0, then the Relative Depth Value shall be equal to -65504.0.

If Z_{Int} is in the range -65504.0 to 65504.0 inclusive, then the Relative Depth Value shall be equal to the half-precision floating point value nearest to Z_{Int}.

6.2 Derivation of 32-Bit Depth Value Representation from 16-Bit Relative Depth Value Representation

If the Relative Depth Value is NaN, then the Depth Value shall also be NaN.

If the Relative Depth Value is +INF, then the Depth Value shall also be +INF.

If the Relative Depth Value is neither NaN, nor +INF, then the Depth Value, denoted by Z, shall be derived from the Relative Depth Value, denoted by Z', as follows:

Equation 8.2

$$Z = Z_c + S * Z'$$

Where,

S denotes the DepthScaleFactor

Z_c denotes the DepthOffset.

7 Metadata

7.1 DepthScaleFactor

The DepthScaleFactor value shall be provided if the Depth Map contains 16-bit Relative Depth Values. The DepthScaleFactor value shall be a positive real number in the IEEE 754 single-precision (32-bit) binary floating-point format.

7.2 DepthOffset

The DepthOffset value shall be provided if the Depth Map contains 16-bit Relative Depth Values. The DepthOffset value shall be a non-negative real number in the IEEE 754 single-precision (32-bit) binary floating-point format.

7.3 DepthSourceType

DepthSourceType, if present, shall indicate the origin of the depth map as an 8-bit integer. The depth map may originate from a variety of sources. The source of the Depth Map may be annotated according to the following table:

DepthSourceType	Description
0	Unknown or undefined
1	Animation rendering
2	Movie production, prepared offline
3	Multi-view camera capture, prepared offline
4	Multi-view camera capture, from live action
5	On-scene depth measurements
>5	Reserved

7.4 DepthRemappingType

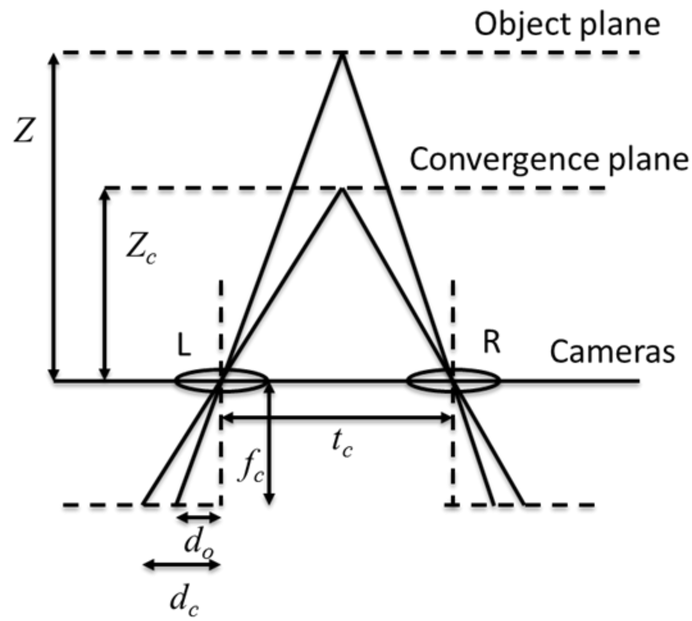
DepthRemappingType, if present, shall be an 8-bit integer that defines whether remapping (editing) has been applied to the Depth Values. The remapping may be annotated according to the following table:

DepthRemappingType	Description
0	Not remapped
1	Remapped
>1	Reserved

Annex A Stereo Application (Informative)

The relationship between the Depth Value of an object and the disparity that occurs when images of the object are captured using a stereo camera setup can be found by taking into account the interaxial distance, focal length, and effective pixel size of the cameras.

The following figure illustrates this relationship by assuming a shift sensor stereo camera setup.



The disparity, d, measured in pixels, caused by an object at a depth of Z from the cameras can be found as,

Equation B.1

$$d = 2(d_c - d_o) = \frac{f_c * t_c}{E_c}\left(\frac{1}{Z_C} - \frac{1}{Z}\right)$$

Where,

Z_c denotes the distance to the plane of convergence.

d_c denotes the required sensor shift in pixels for each camera in order to obtain a distance to convergence of Z_c

d_o denotes the horizontal distance in pixels from the optical axis of the camera to the projection of the object on the camera sensor,

f_c denotes the focal length of the two cameras assumed to be equal,

t_c denotes the interaxial distance,

E_c denotes the effective width of a pixel in the camera,

Conversely, the depth can be derived given the disparity as:

Equation B.2

$$Z = \frac{1}{\left(\frac{1}{Z_c} - \frac{d}{A}\right)}$$

Where,

A is equal to $f_c t_c / E_c$.

Alternatively, if a parallel camera setup with no sensor shift is assumed, then Z_c approaches infinity. In that case, the disparity can be computed as a function of the depth as,

Equation B.3

$$d = \frac{-f_c * t_c}{Z * E_c}$$

Conversely, the depth, Z, can be computed from the disparity as,

Equation B.4

$$Z = \frac{-f_c * t_c}{d * E_c}$$

Annex B Bibliography (Informative)

The following documents are among those consulted in relation to this standard:

SMPTE ST 2066-2012, Disparity Map Representation for Stereoscopic 3D

SMPTE EG 2061:2016, Stereoscopic Distribution Master – Glossary

In particular, these standards describe the image resolutions supported by this standard:

SMPTE ST 274:2008, Television — 1920 x 1080 Image Sample Structure, Digital Representation and Digital Timing Reference Sequences for Multiple Picture Rates

SMPTE ST 296:2012, 1280 x 720 Progressive Image 4:2:2 and 4:4:4 Sample Structure — Analog and Digital Representation and Analog Interface

SMPTE ST 2036-1:2014, Ultra High Definition Television — Image Parameter Values for Program Production

SMPTE ST 2048-1:2011, 2048 × 1080 and 4096 × 2160 Digital Cinematography Production Image Formats FS/709