Standard of the Camera & Imaging Products Association

CIPA DC-006-Translation- 2021

Stereo Still Image Format for Digital Cameras

This translation has been made based on the original Standard (CIPA DC-006-2021). In the event of any doubts arising as the contents, the original Standard is to be the final authority.

Established in August 2008
Revised in December 2021

Prepared by:
Standardization Committee

Published by:
Camera & Imaging Products Association

Disclaimer

1. Neither CIPA nor any of its members shall in any way warrant or take any responsibility for no-infringement of Intellectual Property Rights with respect to the use of CIPA Standards.
2. Neither CIPA nor any of its members shall give any warranty of any kind or take any responsibility for the scope, validity, and essentiality of the Essential Intellectual Property Rights with respect to CIPA Standards.
3. Neither CIPA nor any of its members which are not related to such licensing shall take any responsibility for the terms and conditions of licenses with owners of Intellectual Property Rights, or other licensing negotiations and the results of such negotiations with respect to CIPA Standards.
4. Neither CIPA nor any of its members shall give any warranty of any kind or take any responsibility, either expressed or implied, including warranties of merchantability and fitness for particular purpose, with respect to CIPA Standards.
5. Neither CIPA nor any of its members shall take any responsibility for any damages (meaning all damages including without limitation, loss of business profits, or other incidental or consequential damages) arising out of any use or inability to use the CIPA Standards. The same applies even if either CIPA or its members have been advised of the possibility of such damages.
6. Neither CIPA nor any of its members shall take any responsibility for any disputes that arise at an adopter of CIPA Standards that stem from or are in connection with CIPA standards or the use of CIPA standards.
7. In the event that a statement is not obtained from Sub-Working Group Participant Members to the effect that Essential Intellectual Property Rights are licensed under reasonable (or free) and nondiscriminatory terms, due to believing that Intellectual Property Rights will not be infringed by use of CIPA Standards even after the establishment, addition, or modification of Mandatory Provisions when enacting or revising CIPA Standards, neither CIPA nor any of its members shall give any warranty of any kind that Essential Intellectual Property Rights are not included in the CIPA Standards, and shall not take any responsibility for any disputes that arise as a result of such Intellectual Property Rights being included in the CIPA Standards.

Contents

Revision history

Aug, 2008 1st edition	CIPA DC-006-2008	Established Stereo Still Image Format for Digital Cameras
Dec, 2021 Revised 2nd edition	CIPA DC-006-2021	Updated [2. Normative references] ▸The standard number of Reference 1 has changed along with its revision. (The latest version of Reference 2 has no change.)

1. Scope

This standard specifies the formats to be used for images and metadata related to stereo image (Stim tags), in the case of recording stereo images as image files in digital still cameras and similar devices and systems.

* Note that stereo images in this standard are limited to the binocular (i.e. two viewpoints) type aligned viewpoint images.

2. Normative references

The following official standards contain provisions that constitute provisions of this standard through reference in this text

1) CIPA DC-009-2010 Design rule for Camera File system : DCF Version 2.0 (Edition 2010)

2) ISO/IEC 10918-1:1994 Information technology - Digital compression and coding of continuous-tone still images- Requirements and guidelines

3. Definition of Terms

The following definitions apply to terminology used in this document.

Term	Definition/meaning
Stereo image	An image containing information in the depth direction (including “sense of depth” information). Stereo image in this standard is an aligned viewpoint image produced placing two viewpoint images in side-by-side in the same plane to create an image for stereoscopic viewing that conveys the perception of depth.
Viewpoint image	Each image of a subject as viewed from different viewpoint.
Aligned viewpoint image	A single image created by combining viewpoint images in the same plane (see Section 6).
Parallax image	A collection of images of a subject as viewed from different viewpoints (i.e., viewpoint images).
First viewpoint image area	The left-hand side of the aligned viewpoint image area (see Section 6).
Second viewpoint image area	The right-hand side of the aligned viewpoint image area (see Section 6).

L viewpoint image	The image captured from the left-hand viewpoint on the assumption of taking by an imaging device (i.e., camera) (see Section 6).
R viewpoint image	The image captured from the right-hand viewpoint on the assumption of taking by an imaging device (i.e., camera) (see Section 6).
Stereo photographing adapter	An optical device that enables an ordinary camera to produce stereo images. An example of device has the structure that enables to capture both L and R viewpoint images optically by being attached to the front of the taking lens.
Stereo photographing	The act of producing stereo images consisting of two viewpoint images (or parallax images).
Parallel view	Observation method for a stereo image marked by viewing with substantially parallel (i.e., not intersecting) lines of sight. The image destined for the left eye is positioned on the left and the image for the right eye is positioned on the right.
Cross view	Observation method for a stereo image marked by viewing with intersecting lines of sight. The image destined for the left eye is positioned on the right and the image for the right eye is positioned on the left.
Parallel view alignment	Alignment of viewpoint images in an aligned viewpoint image in parallel view, i.e., with the L viewpoint image in the first viewpoint image area and the R viewpoint image in the second viewpoint image area (see Section 7.2.3).
Cross view alignment	Alignment of viewpoint images in an aligned viewpoint image in cross view, i.e., with the R viewpoint image in the first viewpoint image area and the L viewpoint image in the second viewpoint image area (see Section 7.2.3).
2D image	A flat plane image with no disparity or depth information. In this document, respective viewpoint images are 2D images.

3D display	A display/playback device equipped with a display/playback mode for stereoscopic viewing using binocular parallax.
2D view	Displaying a single viewpoint image without binocular parallax. Equivalent to 2D monocular display of initial display effect (see Section 7.2.17).
Convergence position	The position at which the lines of sight from the left and right eyes intersect in stereoscopic viewing using binocular parallax (i.e., stereo image capture and stereoscopic observation). See also: ConvergenceBaseImage (Section 7.2.12) ConvergenceDistance (Section 7.2.18)
Disparity	The differences between images taken from different perspectives corresponding to the left and right eyes. In this document, the disparity is expressed as the difference of the distances (in pixels) from the left-hand end of the points corresponding to the viewpoint images (see Detailed Explanation in Sections 7.2.15 and 7.2.16).
Representative image	One of the two viewpoint images consisting of the stereo body image file that has been designated by the file creator as representative of the total stereo image (see Section 7.2.11).
Representative image file	The image file containing the representative image as per this standard.
Body image file	The image file containing the stereo image (i.e. aligned viewpoint image) and associated metadata (Stim tags) as per this standard.

4. File structure

4.1. Purpose

A body image file for stereo images in this standard employs a unique file extension (see Section 5), which may affect compatibility with image viewers that are not compliant with this standard. For this reason, the DCF camera file system rule (Normative reference 1) is leveraged in order to prevent confusion and ensure efficient operation with ordinary image capture and viewing devices (such as digital cameras) by users.

Thus, the DCF object consists of the body image file and a representative image file containing the corresponding representative image (see below). This structure supports unified treatment at standard file operations such as copy, move and delete for the files. The representative image is substituted for viewing on DCF readers, while Exif tag information in the representative image file can also be accessed if required.

4.2. Body image file

The body image file is an image file having the unique file extension, defined as a DCF extended image file. The extension is defined as 'ssi'. It consists of the DCF object, with the DCF basic file (i.e., the representative image file) .
The format of the body image file is described in Section 5.

4.3. Representative image file

Recording a body image file shall involves recording the representative image specified by the representative image tag (see Section 7.2.11) as a DCF basic file, which consists of the DCF object, with the corresponding body image file.
The image size of the representative image may be any (i.e., resizable).

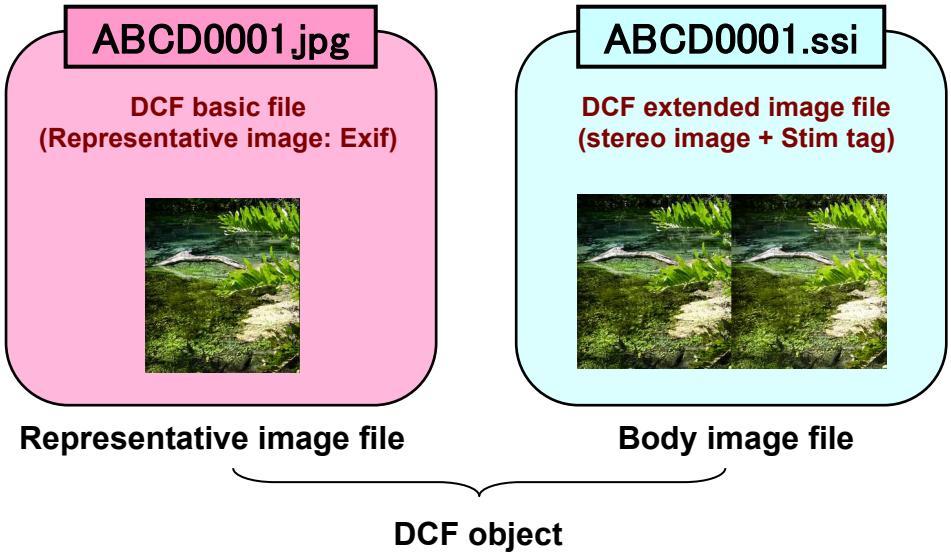


Figure 1 File structure

5. Body image file format

The body image file format for stereo images (Stim) specified in this standard conforms to the JPEG Baseline DCT format stipulated in ISO/IEC 10918-1 (Normative reference 2). An application marker segment for the Stim tag (APP3) is also inserted.

The APP3 marker segment consists of the APP3 marker, the Stim identifier code, the Stim header and the Stim tag (APP3 body), as shown in Figure 2.

The Stim identifier code indicates that the APP3 is an Stim tag. It shall be recorded as the 4-byte code followed by 2 bytes of 00.H, as shown in Figure 3.

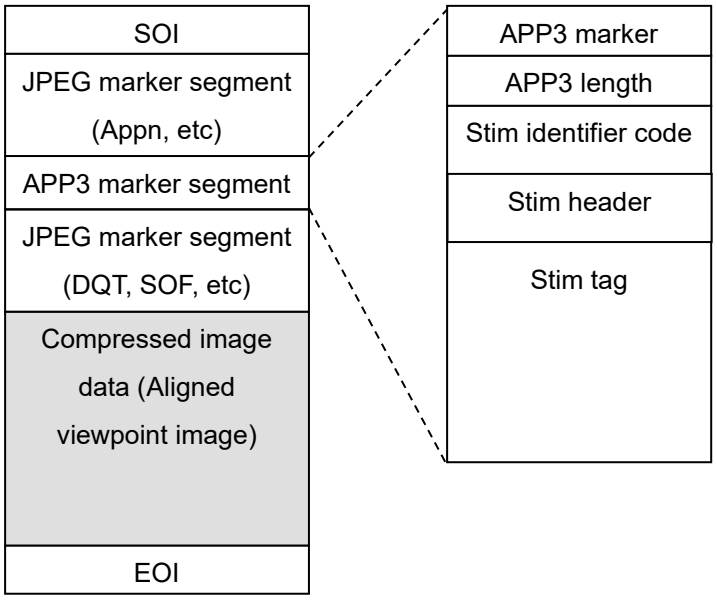


Figure 2 Structure of body image file incorporating Stim tag

Address offset (Hex)	Code (Hex)	Meaning
+00	FF	Marker Prefix
+01	E3	APP3
+02		APP3 length
+04	53	'S'
+05	74	't'
+06	69	'i'
+07	6D	'm'
+08	00	Null
+09	00	Pad
+0A		Stim header
+12		Stim tag

Figure 3 Basic structure of APP3 marker segment

The structure of Stim header is below.

Name	Size (byte)	Value
Byte Order	2	Either "II" (4949.H) (Little endian), or "MM" (4D4D.H) (Big endian). Select according to the recording CPU usually.
42	2	002A.H (Fixed)
Offset of IFD	4	Offset for Stim tag.(Byte number from the top of Stim header). Write 00000008.H.

The structure known as Image File Directory (IFD) is used for storing Stim tags (see Section 7.1.1).
Stim tags shall be recorded directly after Stim header.

6. Stereo image

In this standard, a stereo image refers to an aligned viewpoint image representing a side-by-side combination of two viewpoint images (monocular images corresponding to the left-eye and right-eye viewpoints) in the same plane. The viewpoint images are rectangular in shape.

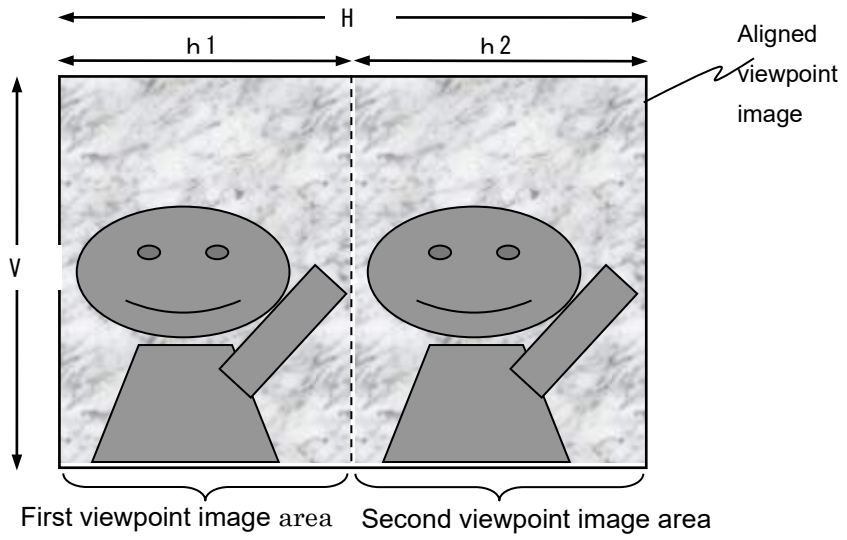


Figure 4 Structure of stereo image

In this standard, the term “aligned viewpoint image” is used to denote the overall image, while “first viewpoint image area” and “second viewpoint image area” denote the separate image areas. This convention, as illustrated in the

diagram above, helps to delineate the different image areas in the recorded image. The aligned viewpoint image consists of the first viewpoint image area on the left-hand side together with the second viewpoint image area on the right-hand side.

The first and second viewpoint image areas are defined based on the aligned viewpoint image as follows.

♦ Definition of Viewpoint image area:

If H is the size (in pixels) of the aligned viewpoint image in the horizontal direction and V the size in the vertical direction, then h, the value is H divided by two(the number of viewpoints in the horizontal direction), is the horizontal size of the respective viewpoint image areas. When H is an odd number of pixels, leftover pixels after division are allocated to the left-hand area (i.e., the first viewpoint image area). Thus, the respective horizontal sizes of the viewpoint image areas h1 and h2 can be expressed as follows:

$h1 = (H + \alpha) / 2$;

$h2 = (H - \alpha) / 2$;

*α = 0 when H is an even number, or 1 when H is an odd number.

The vertical size of both viewpoint image areas is V.

♦ Definition of viewpoint number

Viewpoint numbers are allocated to each viewpoint in accordance with the assumed alignment of shots by the image capture device (i.e., camera). This is independent of the definition for recorded image described above. The left-hand viewpoint is denoted by 0, and the right-hand viewpoint by 1. The images captured at the respective viewpoints are called “L viewpoint image” and “R viewpoint image”, as shown in Figure 5.

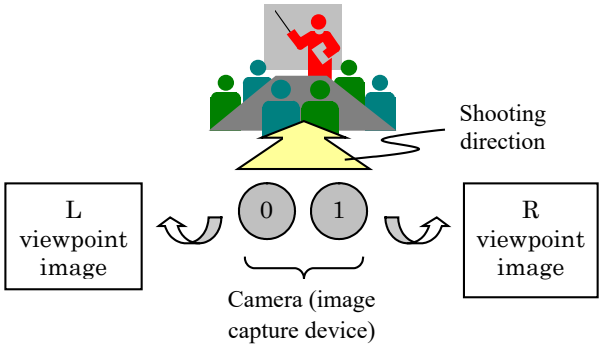


Figure 5 Viewpoint numbers

7. Stim tags

7.1. Stim tag structure and content

7.1.1. Stim tag structure

The IFD employed in this standard consists of a 2-byte count (number of fields), field entry array of 12 bytes each, and a 4-byte offset to the next IFD.
The structure is shown below.

Size (byte)	Contents
2	Count (Number of fields = n)
12	Field entry 1
12	Field entry 2
•	•
•	•
•	•
12	Field entry n
4	Offset to the next IFD*

Field entry array

*4-bytes of 00.H is recorded as the offset to the next IFD in this version.

The 12-byte field entries consist of the following four elements:

Bytes 0-1	Tag
Bytes 2-3	Type
Bytes 4-7	Count
Bytes 8-12	Value offset

The components are described in more detail below.

Tag:

Each tag is allocated a unique 2-byte number that identifies the fields.

Type:

Stim tags use the following types:

Value	Type	Meaning
1	BYTE	8-bit unsigned integer
2	-	reserved
3	SHORT	16-bit (2-byte) unsigned integer
4	LONG	32-bit (4-byte) unsigned integer
5	RATIONAL	LONG x 2. First LONG denotes numerator and second LONG denotes denominator
6	-	reserved
7	UNDEFINED	8-bit byte which can be any value as per field definition
8	-	reserved
9	SLONG	32-bit (4-byte) signed integer (complement notation for 2)

Count:

Number of values. Note that Count does not represent the total number of bytes. For example, a SHORT (16 bit) value has two bytes but the Count is 1.

Value Offset:

The offset from the top of the Stim tag to the recorded position of the value. Where the value fits into four bytes, the value itself is recorded. Where the value is smaller than four bytes, it is left-justified within the four-byte area; thus, the value is stored starting from the byte offset area with the lower offset. For example, in big endian format, 00010000.H shall be recorded if the type is SHORT and the value is 1.

Field entries must be recorded in order starting from the one with the lowest tag number. There are no stipulations regarding the order or position of recording tag values.

7.1.2. Stim tag elements

When stereo image data is recorded, the Stim tags listed below are stored in the prescribed locations within the recorded data. Table 1 lists the elements defined as Stim tags.

The Support level column in Table 1 indicates as follows:

M = Mandatory, R = Recommended, O = Optional

Table1 Stim tag elements

Tag name	Tag number		Type	Count	Support level
	DEC	HEX			
StimVersion	0	0	BYTE	4	M
ApplicationData	1	1	UNDEFINED	Any	O
ImageArrangement	2	2	BYTE	1	M
ImageRotation	3	3	BYTE	1	M
ScalingFactor	4	4	RATIONAL	1	M
CropSizeX	5	5	LONG	1	O
CropSizeY	6	6	LONG	1	O
CropOffsetX	7	7	UNDEFINED	7 or 12	O
CropOffsetY	8	8	UNDEFINED	7 or 12	O
ViewType	9	9	BYTE	1	O
RepresentativeImage	10	A	BYTE	1	M
ConvergenceBaseImage	11	B	BYTE	1	O
AssumedDisplaySize	12	C	LONG	1	R
AssumedViewDistance	13	D	LONG	1	R
RepresentativeDisparityNear	14	E	SLONG	1	R
RepresentativeDisparityFar	15	F	SLONG	1	R
InitialDisplayEffect	16	10	BYTE	1	O
ConvergenceDistance	17	11	LONG	1	O
CameraArrangementInterval	18	12	LONG	1	O
ShootingCount	19	13	BYTE	1	O

7.2. Implications and values of the tags

Stim tags are described in detail below. Note that in categories involving geometric numerical values such as distance and size, where the presence of the optical system creates a discrepancy between the actual value and the optically equivalent value for shooting and viewing (such as microscope images), the optically equivalent value shall be recorded instead of the actual value.

7.2.1. StimVersion

Description: Shows the version number of this standard, expressed in four bytes. The bytes are recorded as A1, A2, B1, B2, starting from the lowest address number, where A1 and A2 are the upper part of the standard version and B1 and B2 are the lower part.

<Tag> 0 (0000.H)
<Type> BYTE
<Count> 4
<Value> 0, 1, 0, 0 (fixed)

7.2.2. ApplicationData

Description: Available as an optional user data area. For instance, where a stereo photographing adaptor is used, the ApplicationData area can store information about the adaptor model.

<Tag> 1 (0001.H)
<Type> UNDEFINED
<Count> Any

7.2.3. ImageArrangement

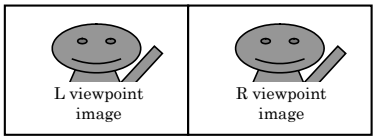
Description: Denotes the method used to arrange viewpoint images to produce the aligned viewpoint image. The two arrangement methods are:

(1) Parallel view alignment: the L viewpoint image is allocated to the first viewpoint image area and the R viewpoint image is allocated to the second viewpoint image area.

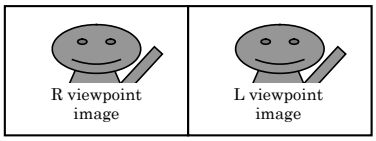
(2) Cross view alignment: the R viewpoint image is allocated to the first viewpoint image area and the L viewpoint image is allocated to the second viewpoint image area.

Figure 6 shows the alignment of viewpoint images in each arrangement method.

<Tag> 2 (0002.H)
<Type> BYTE
<Count> 1
<Value> 0 : parallel view alignment
1 : cross view alignment
other : reserved
<Default>None



(a) Parallel view alignment



(b) Cross view alignment

Figure 6 Stereo image arrangement

7.2.4. ImageRotation

Description: Denotes the method of rotation/reversal of viewpoint images.

<Tag> 3 (0003.H)
<Type> BYTE
<Count> 1
<Value> 1: No rotation/reversal
other: reserved

Remarks: For future expansion. Fixed at 1 in StimVersion = 0.1.0.0.

7.2.5. ScalingFactor

Description: Provides information on aspect ratio for reduction/expansion of horizontally aligned image.

<Tag> 4 (0004.H)
<Type> RATIONAL
<Count> 1
<Value> 1/1 (no change in aspect ratio), other = reserved

Remarks: For future expansion. Fixed at 1/1 in StimVersion = 0.1.0.0.

7.2.6. CropSizeX

Description: Denotes size of crop area in the horizontal direction. See also "♣ About crop area size and offset" below.

<Tag> 5 (0005.H)
<Type> LONG
<Count> 1
<Value> Unit: pixel
<Default> H/2 … rounded off (See Section 6)

7.2.7. CropSizeY

Description: Denotes size of crop area in the vertical direction. See also "♣ About crop area size and offset" below.

<Tag> 6 (0006.H)
<Type> LONG
<Count> 1
<Value> Unit: pixel
<Default> V (See Section 6)

7.2.8 CropOffsetX

Description: Denotes crop area offset in the horizontal direction. See also "♣ About crop area size and offset" below.

<Tag> 7 (0007.H)
<Type> UNDEFINED
<Count> 7 or 12
<Value> See below
<Default> 0, 0, 0 (See Table 2)

7.2.9. CropOffsetY

Description: Denotes crop area offset in the vertical direction. See also "♣ About crop area size and offset" below.

<Tag> 8 (0008.H)
<Type> UNDEFINED
<Count> 7 or 12
<Value> See below
<Default> 0, 0, 0 (See Table 2)

♣ About crop area size and offset

The four tag described in Sections 7.2.6 through 7.2.9 above are used to specify the expected areas associated with cropping as intended by the file creator. Specifying the location to be displayed is called image cropping. The specified area is called the cropping area. The cropping area is rectangular in shape, and the same size as the viewpoint images. Offset is specified as a set of coordinates relative to the origin at the upper left corner of the viewpoint images. Offset may be specified separately for each viewpoint image. The values of the tags that indicate offset (CropOffsetX and CropOffsetY) are expressed as shown in Table 2. Figure 7 shows an example of the cropping area.

Table 2　CropOffsetX and CropOffsetY description

Type	Size(byte)	Default	Meaning
SHORT	2	0	Indicates whether the offset value is the same for both viewpoint images (common setting) or different for each viewpoint image (individual setting). 0 : common offset setting 1 : individual offset setting other : reserved
BYTE	1	0	Viewpoint number (0 – 1) Note that Viewpoint number is 0 for common offset value
SLONG	4	0	Offset value Expressed as a coordinate (in pixels) relative to the pixel at the upper left corner of the viewpoint image
* Individual offset values are given for remaining viewpoint images.			
BYTE	1	—	Viewpoint number (0 – 1)
SLONG	4	—	Offset value Expressed as a coordinate (in pixels) relative to the pixel at the upper left corner of the viewpoint image

7 bytes

12 bytes

* The two offset tag values have a 7-byte structure if the value is common to both viewpoint images (common setting) or a 12-byte structure if the value is specified separately for each viewpoint image (individual setting) according to above.

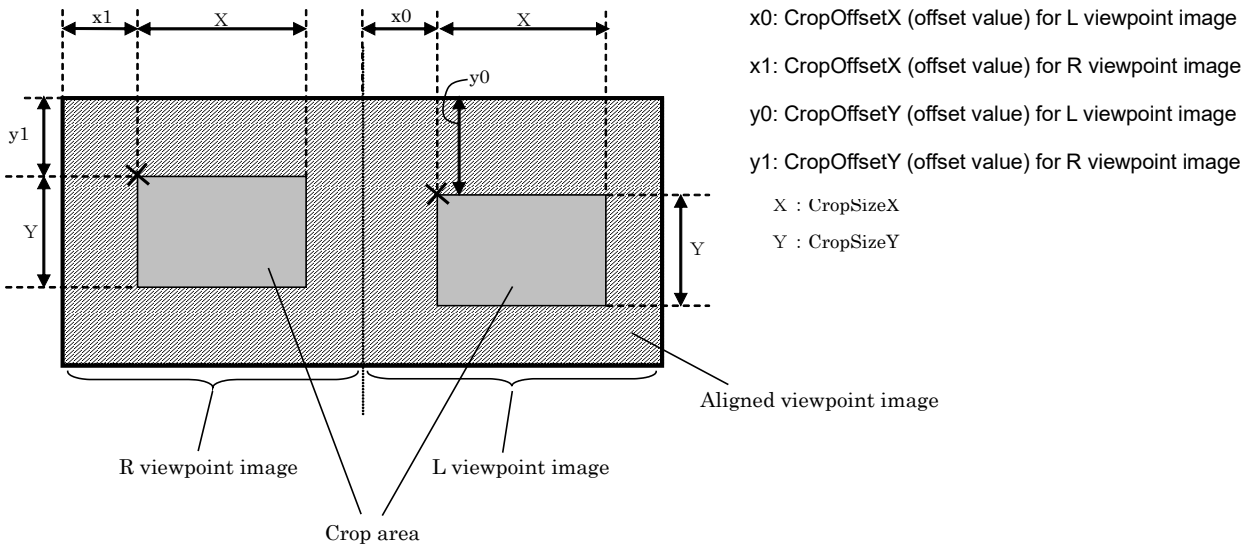


Figure 7 Crop area size offset (individual offset values, cross view alignment)

7.2.10. ViewType

Description: This tag indicates whether stereo image capture (or production) has been designed to make the subject of the image appear nearer to the viewer than the image plane, in other words, to intend the effect giving pop-up feeling by making the subject stand out from the rest of the image.

ViewType can be used in image searching and sorting. It is governed by shooting conditions such as the distance between the image capture device (i.e., the camera) and the main subject (see Explanation 3.3).

<Tag> 9 (0009.H)
<Type> BYTE
<Count> 1
<Value> 0 : no pop-up effect
1 : pop-up effect
other : reserved
<Default> 0

7.2.11. RepresentativeImage

Description: Specifies the representative image in the viewpoint number, denoting the intentions of the file creator.

It is available for printing as a 2-D picture with a standard printer, or for displaying in the case that 2-D showing mode selectable by the user is required even with 3-D display monitor, for example.

<Tag> 10 (000A.H)
<Type> BYTE
<Count> 1
<Value> 0 (L viewpoint image)
1 (R viewpoint image)
other: reserved
<Default> 0 (L viewpoint image)

7.2.12. ConvergenceBaseImage

Description: The convergence position can be adjusted during playback by shifting the horizontal display position of the viewpoint image. In certain situations, the file creator may wish to employ a viewpoint image with fixed display position for convergence adjustment. This is denoted with this tag. The viewpoint image is specified by the viewpoint number. 255 shall be specified for the case if the two viewpoint images will be shifted by equivalent amounts to the left and right (see Explanation 3.4).

<Tag> 11 (000B.H)
<Type> BYTE
<Count> 1
<Value> 0 (L viewpoint image)
1 (R viewpoint image)
255: equivalent for both viewpoints
other: reserved
<Default> none

7.2.13. AssumedDisplaySize

Description: Shows the display size of the stereo image as assumed/intended by the file creator, expressed in terms of the length in the horizontal direction (see Explanation 3.5).

Where the crop area (see Sections 7.2.6 – 7.2.9) is specified, the display size shall correspond to the crop area.

<Tag> 12 (000C.H)
<Type> LONG

<Count> 1
<Value> Unit mm
<Default> none

7.2.14. AssumedViewDistance

Description: Shows the distance from the viewing position to the display of the stereo image playback device, as assumed/intended by the file creator (see Explanation 3.5).

<Tag> 13 (000D.H)
<Type> LONG
<Count> 1
<Value> Unit mm
<Default> none

7.2.15. RepresentativeDisparityNear

Description: Indicates the disparity in the horizontal direction between the L and R viewpoint images for the nearest point from the viewer's perspective. It is an indicator of the stereoscopic effect of a stereo image. See "◆About Representative Disparity" below.

<Tag> 14 (000E.H)
<Type> SLONG
<Count> 1
<Value> Unit pixel
<Default> none

7.2.16. RepresentativeDisparityFar

Description: Indicates the disparity in the horizontal direction between the L and R viewpoint images for the farthest point from the viewer's perspective. It is an indicator of the stereoscopic effect of a stereo image. See"◆About Representative Disparity" below.

<Tag> 15 (000F.H)
<Type> SLONG
<Count> 1
<Value> Unit pixel
<Default> none

◆ About Representative Disparity

The two tags described in 7.2.15 and 7.2.16 above are used to express representative disparity, which is the maximum disparity between the L and R viewpoint images, as an indicator of the stereoscopic effect of a stereo image. The two tag values represent the disparity in the horizontal direction between the L and R viewpoint images for the nearest and farthest points in the image from the viewer's perspective respectively. They are defined as follows (see Explanation 3.6).

$$\text{RepresentativeDisparityNear} = (d_{nL} - d_{nR})$$
$$\text{RepresentativeDisparityFar} = (d_{fL} - d_{fR})$$

In the above expression, d_{nL} and d_{nR} are the distances on the respective viewpoint images (in pixels) from the left-hand edge of the images to the point which appears nearest to the viewer (i.e., the corresponding points with the maximum pop-up (forward) disparity between the L and R viewpoint images), while d_{fL} and d_{fR} are the distances on the respective viewpoint images (in pixels) from the left-hand edge of the images at the point which appears furthest from the viewer (i.e., the corresponding points with the maximum depth (backward) disparity between the L and R viewpoint images). Figure 8 illustrates this relationship.

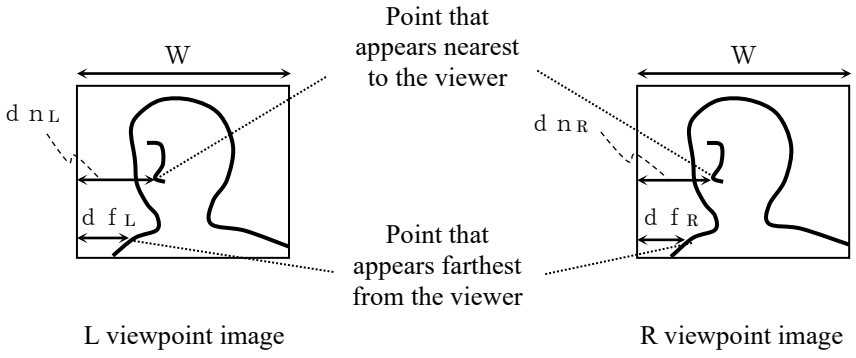


Figure 8 Parameters for calculating representative disparity

7.2.17. InitialDisplayEffect

Description: An effect available to the file creator on a dynamic (movie) display (or electronic display), whereby an image is initially displayed in 2-D (monocular) mode and is then switches to 3-D for added impact. The effect of impact may be judged by the file creator (see Explanation 3.7).

<Tag> 16 (0010.H)
<Type> BYTE
<Count> 1
<Value> 0 : initial display effect off
1 : initial display effect on
other : reserved
<Default> 0

7.2.18. ConvergenceDistance

Description: Indicates the actual distance from the centre of a line linking the principal points at the front of the taking lenses on the object side (which represents the camera position) to the convergence point. The convergence distance is shown as “d” in Figure 9. Where d is infinity (i.e., where the optical axes for both cameras are parallel), the ConvergenceDistance value is zero (see Explanation 3.8).

<Tag> 17 (0011.H)
<Type> LONG
<Count> 1
<Value> unit mm; value of 0 indicates d＝∞
<Default> none

7.2.19. CameraArrangementInterval

Description: Shows the distance between the principal points at the front of the taking lenses on the object side, shown as “e” in Figure 9 (see Explanation 3.8).

<Tag> 18 (0012.H)
<Type> LONG
<Count> 1
<Value> unit mm
<Default> none

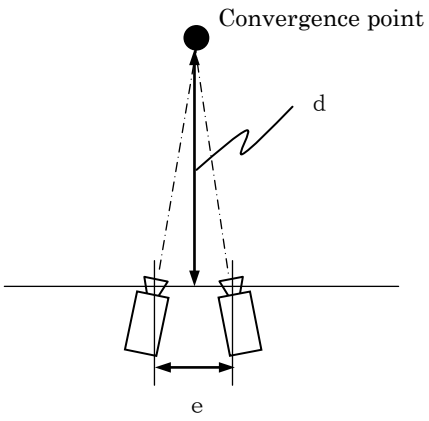


Figure 9 Camera arrangement (where d is distance to convergence point and e is interval)

7.2.20. ShootingCount

Description: Shows the number of shots taken in producing stereo images from two viewpoint images. Rather than a strict count of the number of shots, it classifies the shots in terms of degree of reliability with respect to homogeneity between viewpoint images as either “1-shot” or “2-shot” in accordance with the file creator’s judgment (see Explanation 3.9).

<Tag> 19 (0013.H)
<Type> BYTE
<Count> 1
<Value> 1 : 1-shot
2 : 2-shot
other : reserved
<Default> 1

Explanation

This explanation describes the items set out in this standard and its Annexes, as well as reference material and other associated matters. It does not constitute a part of the standard.

1. Background of the standardization

Stereo photography, which records stereo-viewable subject image by taking two complementary photographs together, leveraging the fact that human vision is getting a stereoscopic perception by the parallax between the left and right eyes, was born at the same time of the birth of photographic technology. The enduring popularity of stereo photography entertainment events indicates a very high level of untapped demand for systems that enable stereoscopic viewing of recorded images. However, apart from several boom cycles, stereo-photography has never really taken hold in the general consumer photography market. This has been attributed to a range of factors, but probably the most important factor is technical difficulty and equipment complexity involved in camera taking, printing and appreciation on the display, compared to general photographic technology.

Meanwhile the recent advent of certain models of mobile phones and computers equipped with 3D (i.e. stereoscopic viewable) displays suggests that 3D representation applicable equipments may well be embraced by consumer markets.

While binocular type 3D photographic images (called stereo images in this standard) have been used by certain users, the applications of stereo images seems to be spreading into diverse range now, because of the combination of recent rapidly growing popularization of digital still cameras and the advent of various displays mentioned above.

In order to make stereo images more appealing and accessible, it will be necessary to have the environment for easy and effective creating and viewing stereo images while also enhancing the viewing experience. In particular, it is important to construct the infrastructure for bridging the gap between the creation of stereo images on digital still cameras (which are enjoying spectacular growth at present) and the tools used to view them (such as display devices and printing systems).

However standard stereo image format has not been established. If this

situation is allowed, manufacturers would simply develop a number of different own formats. This would not only hinder the real popularization but also create unnecessary confusion with existing services such as photography (printing service) and media service possibly.

Considering the above-mentioned situation, this document that is the stereo still image format for digital still cameras designed to provide a standardized format for stereo images and also to promote effective utilization of stereo image data, has been established. Specifically, it sets out regulations for the recording of stereo images into image files together with associated data (Stim tags) used to coordinate the camera with the viewing device.

2. Objectives of this standard

- The target is a format compatible with the existing popular binocular stereo image typically captured by a digital still camera fitted with a stereo imaging optical adaptor, as shown in Figure 10. It could also potentially apply to data created in the same format with authoring tools or equivalent.
- The aim is to create a stereo image format that is independent of the camera type or model used to record the image.
- A framework for ensuring consistent display quality on a variety of different 3-D display devices will be provided.
- Attention should be paid to have no impact on the usage of incompatible devices.

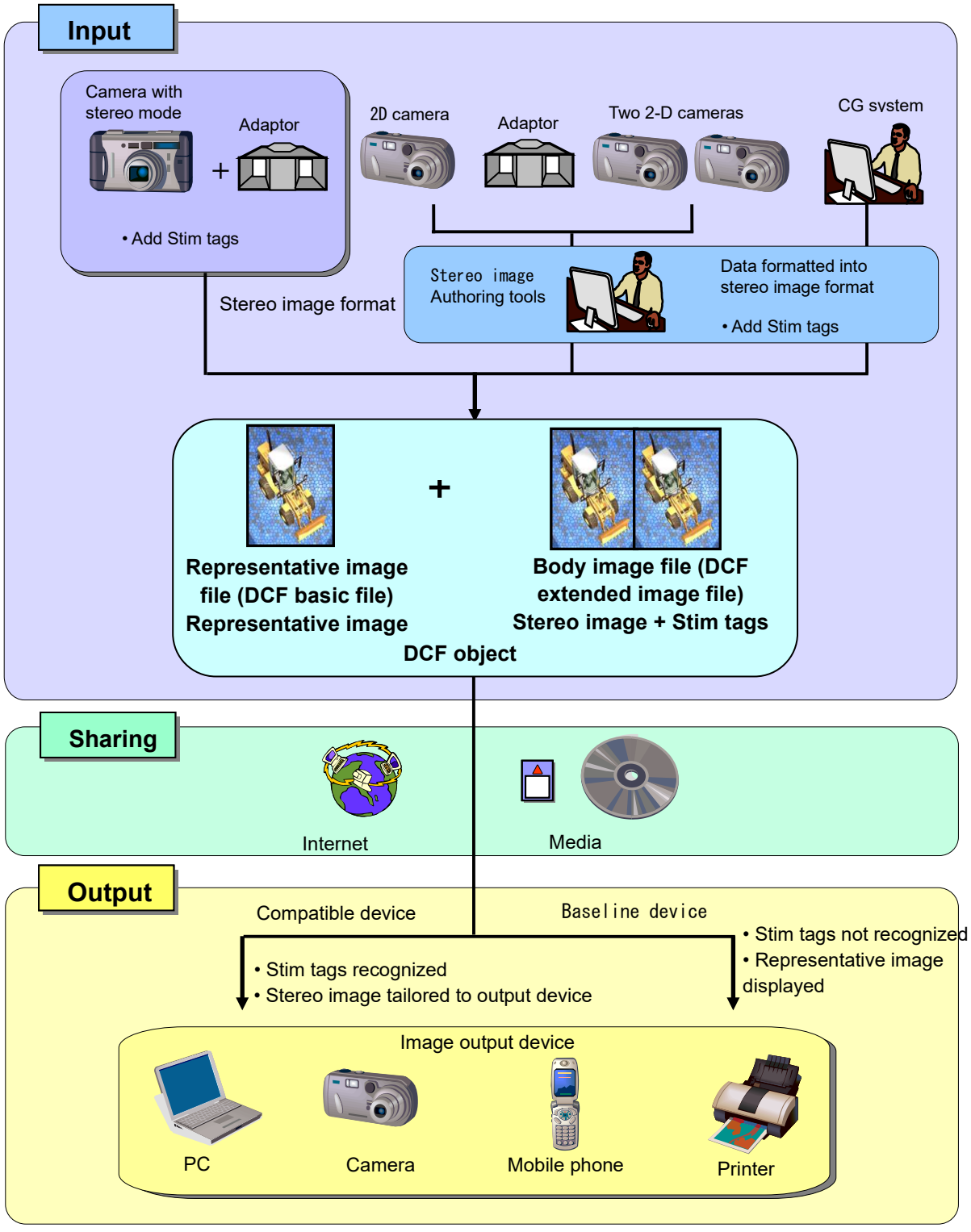
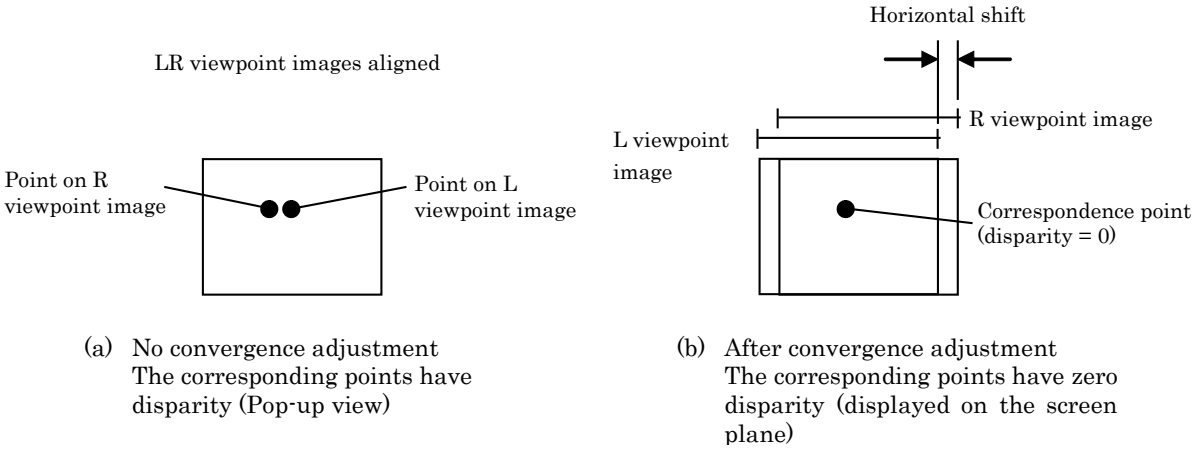


Figure 10 Typical stereo image applications,

3. Explanation for use of Stim tags

3.1. CropOffsetX

It was stated in the body text that a designated area is set aside for cropping use. This parameter can be used for convergence adjustment, the process of shifting images relative to one another in the horizontal (X) direction in order to modify the point at which zero disparity is achieved (i.e. displayed on the screen plane). The amount of adjustment is defined as the amount of shift expressed in pixels (see below), and is added to CropOffsetX for the image.



(a) No convergence adjustment
The corresponding points have disparity (Pop-up view)

(b) After convergence adjustment
The corresponding points have zero disparity (displayed on the screen plane)

Figure 11 Convergence adjustment

3.2. CropOffsetY

It was stated in the body text that a designated area is set aside for cropping use. This parameter can be used for fine adjustment of the vertical display position of the left and right images which in certain situations can enhance the stereoscopic effect. The amount of adjustment is expressed in pixels and added to CropOffsetY.

3.3. ViewType

It was stated in the body text that the view type is determined by shooting conditions such as the distance between the image capture device (i.e., the camera) and the main subject. For basic example, if the main subject is closer to the camera than the convergence point, it can be treated as an image having the pop-up effect.

3.4. ConvergenceBaseImage

As stated in Explanation 3.1, convergence adjustment involves shifting the two viewpoint images relative to one another in the horizontal direction. The choice of viewpoint image (L or R) as the base or fixed image governs the display range of the resulting stereo image. Thus if the base image is not chosen properly, the subject and/or image area may not be viewed stereoscopically as intended by the creator of the stereo image file. This tag is used to specify the base viewpoint image for convergence adjustment as intended by the file creator.

3.4.1. Not using the ConvergenceBaseImage

If designation of the base image with this tag is not used in the case that convergence adjustment is processed by a playback device and the results are displayed, the subject and/or image area in the resulting output may not be viewed stereoscopically as intended by the creator of the stereo image file, as shown in Figure 12 below.

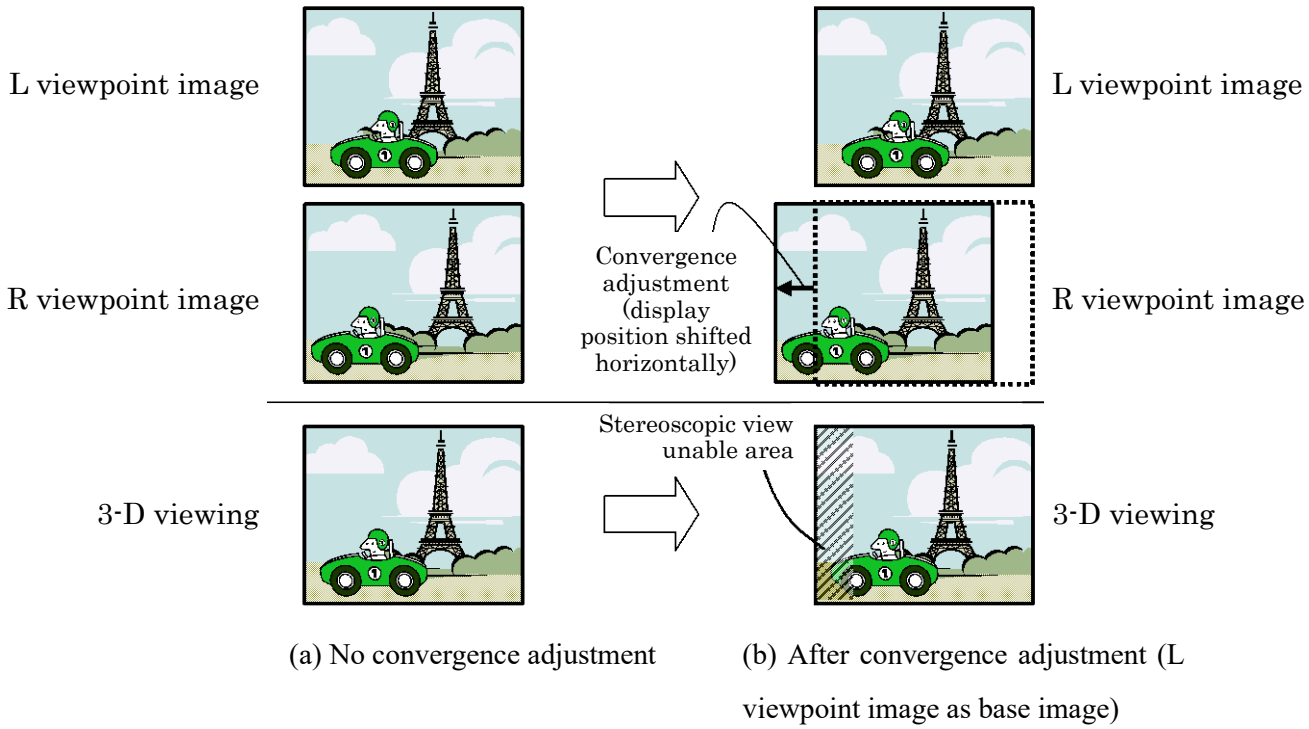


Figure 12 Convergence adjustment — Example 1

3.4.2. Using the ConvergenceBaseImage

Where the ConvergenceBaseImage tag is used to designate the viewpoint image to use as the base (i.e fixed display position), subsequent convergence adjustment will not affect the stereoscopic viewing of the subject and image area as intended by the file creator (see Figure 13).

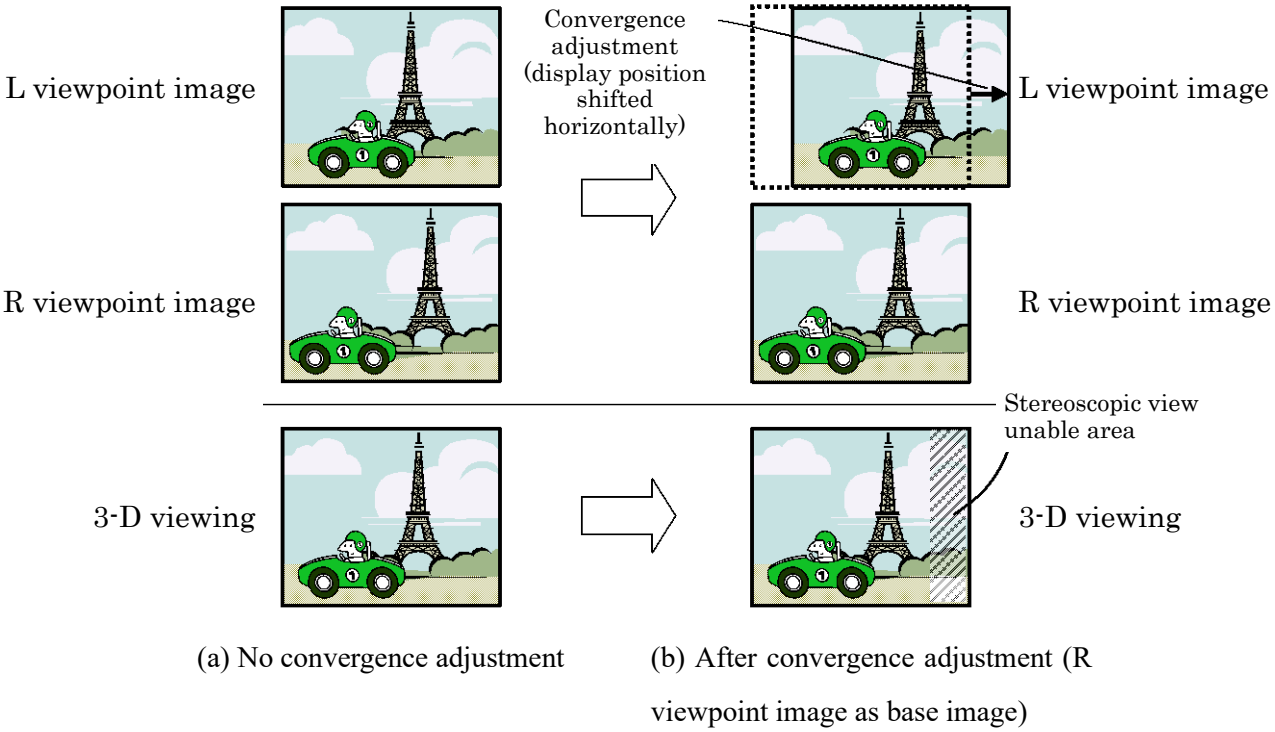


Figure 13 Convergence adjustment — Example 2

3.5. AssumedDisplaySize and AssumedViewDistance

AssumedDisplaySize and AssumedViewDistance indicate the stereo image display size and viewing distance assumed and intended by the file creator, respectively. These are used, for example, to recreate the viewing conditions intended by the file creator, thereby optimizing the stereoscopic effect in image reproduction.

A concrete idea may be that the ratio between display size and viewing distance in image reproduction is matched to the ratio between these two values. This can be achieved by enlarging or reducing the image display size on the display screen for example.

For the purpose of this document, the stereo image display size (W) is not the same as the size of the display area on the playback device, but the size shown in the figure below in the example case of image reduction.

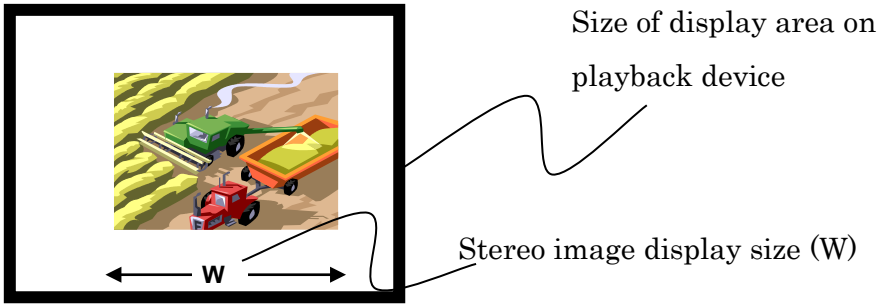


Figure 14 Display size after image reduction

Similarly, image enlargement would virtually cause the stereo image display size (W) to exceed the boundaries of the display area on the playback device as shown in the figure below. The stereo image display size (W) is the total image display size including virtually exceeding portion in this case.

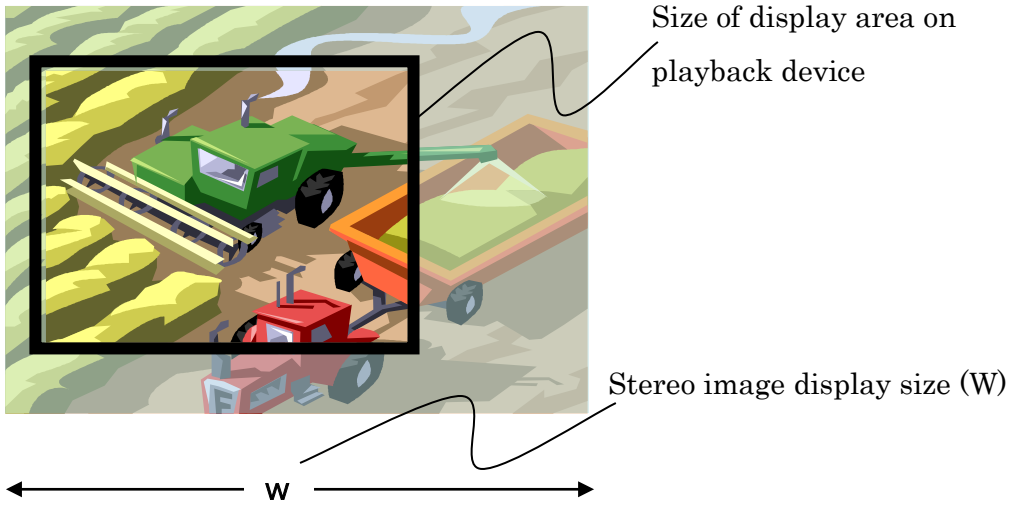


Figure 15 Display size after image enlargement

Although the tag values are expressed in mm, this level of precision is not always required. The required level of physical precision depends on the objectives and conditions.

Table 3 lists example values for AssumedDisplaySize and

AssumedViewDistance.

Table3 Example values for AssumedDisplaySize and AssumedViewDistance under typical viewing conditions

Assumed display size used in creation of stereo image	AssumedDisplaySize [mm]	AssumedViewDistance [mm]
Diagonal 2.5 inch, aspect ratio 3:2	53	350
Diagonal 7 inch, aspect ratio 16:9	155	440
Diagonal 17 inch, aspect ratio 4:3	345	650
Diagonal 32 inch, aspect ratio 16:9	708	1600

* The examples listed above are based on full display mode, with the stereo image display size (W) equal to the horizontal size of the display area on the playback device. The values are intended as an example of this configuration.
* The examples above are only example values and do not in any way constitute recommendations or guarantees.

3.6. RepresentativeDisparityNear and RepresentativeDisparityFar

3.6.1. Meaning of tags

This document has uniquely defined representative disparity as the maximum disparity in the near (close) and far (depth) portions of the stereo image in the body text. In reality, it is difficult to determine the corresponding points and disparities as per the definition with any degree of precision. Even assuming it were possible to obtain a precise value, this may not constitute a valid figure in cases where, for example, the subject (which has the strongest parallax) accidentally appears at the unnoticed edges of the screen. In such a case, the practically valid value should be recorded.

3.6.2. Examples of use

Applications designed for sequential viewing of multiple stereo images on a playback device could be provided with functionality for alerting the viewer and/or restricting the viewing time in response to the intensity of the stereoscopic effect. The examples below describe the case that the intensity of the stereoscopic effect is calculated from the representative disparity value, and the case that the application would alert the viewer and restrict the viewing time based on the

stereoscopic intensity value.

3.6.3. An example of calculation of stereoscopic intensity

For a stereo image viewing application that has been configured to match the horizontal size of a stereo image with the horizontal size of the display screen, the intensity of the stereoscopic effect can be expressed, for example, as a ratio (%) of the representative disparity relative to the horizontal size of the viewpoint image:

$$\gamma\,\mathrm{Near} = (\mathrm{d\,Near} / \mathrm{h}) \cdot 100$$
$$\gamma\,\mathrm{Far} = (\mathrm{d\,Far} / \mathrm{h}) \cdot 100$$

γ Near and γ Far represent the stereoscopic intensity for the near (i.e. pop-up) and far (i.e. depth) portions of the image respectively, dNear and dFar are the RepresentativeDisparityNear and RepresentativeDisparityFar values, and h is the horizontal size of the viewpoint image (see Figure 4). Both may be used as the stereoscopic intensity, or the one with the higher absolute value may be taken as the representative value.

Where the image display involves cropping (see Sections 7.2.6 – 7.2.13), for optimum accuracy the cropping area should be taken into account in the calculation of stereoscopic intensity by replacing the values dNear, dFar and h with following dNear', dFar' and h' respectively in the above expressions.

$$\mathrm{d\,Near'} = \mathrm{d\,Near} + (\mathrm{CropOffsetX_R} - \mathrm{CropOffsetX_L})$$
$$\mathrm{d\,Far'} = \mathrm{d\,Far} + (\mathrm{CropOffsetX_R} - \mathrm{CropOffsetX_L})$$
$$\mathrm{h'} = \mathrm{CropSizeX}$$

*$\mathrm{CropOffsetX_R}$ and $\mathrm{CropOffsetX_L}$ denote the horizontal offset of the cropping area for the R and L viewpoint images respectively.

3.6.4. An example of viewer alert in accordance with stereoscopic intensity

Where the stereoscopic intensity thus calculated is deemed to be too high, the image could be displayed in 2-D together with an alert message in dialog box format as shown in Figure 16. The judgment for the intensity level might be to compare the calculated value with the maximum tolerance level for stereoscopic

intensity of the users, which would be obtained beforehand.

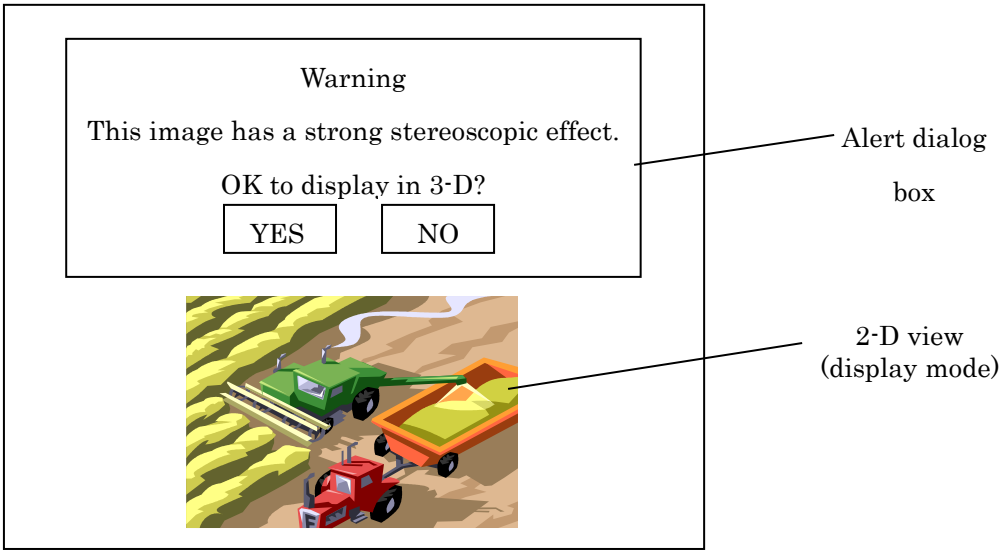


Figure 16 An example of user alert dialog box

3.6.5. An example of restricting viewing time in accordance with stereoscopic intensity

Restriction of viewing time is a kind of processing such as displaying an alert dialog box or automatically switching from 3-D to 2-D display format when the time spent for continuous viewing a stereo image reaches a set limit. It could be tailored to the viewing situation by utilizing the stereoscopic intensity. For instance, the viewing time limit would be triggered when the following condition is satisfied:

$$\int |\gamma| \, dt > 3DViewLimitThreshold$$

*$|\gamma|$ is the larger of γ Near and γ Far with respect to absolute value

3DviewLimitThreshold is the preset threshold by the playback device. The viewing time limit is triggered when the cumulative total of the absolute values of stereoscopic intensity exceeds the threshold value. Thus, the stronger the stereoscopic intensity, the shorter the allowable viewing time because the threshold value is reached more quickly.

3.7. InitialDisplayEffect

In case of playback a stereo image with a dynamic (movie) display device, changing to 3D display (viewing) mode after 2D monocular display mode in the initial period sometimes makes an excellent stereoscopic visual effect, such as pop-upping action effect. This tag is recorded when the file creator requests this effect. The reader can either accept or reject the request. Implementation example of this feature for control could be that the image might be displayed in 2D for an initial period of three seconds and then switched automatically to 3D display mode, or that the image would be displayed in 2D at first and waiting until the viewer manually switches to 3D.

3.8. ConvergenceDistance and CameraArrangementInterval

Although the tag values are expressed in mm, this level of precision is not always required. The required level of physical precision depends on the objectives and conditions.

The tag definitions in this document are based on principal point at the front of the lens on the object side. In practice, the exact location of the principal point of the taking lens is often unknown, especially this location is constantly moving in the case of zoom lenses. In most cases it is acceptable simply to use the location of the camera tripod socket or the most-front part of the lens for example.

However, in order to accommodate close-up shooting and other situations where greater precision is required, the definition has been given explicitly.

3.9. ShootingCount

This tag indicates the level of synchronization and homogeneity among viewpoint images by showing the number of stereo images captured, as stated in Section 7.2.20. For example, either the case that a single camera (or image capture system) is used to capture two images optically split or the case that a single imaging device with two image capture systems for stereo photography is used, where both images are captured simultaneously, is called “1-shot.” On the other hand asynchronous capture of the two images is called “2-shot.”

4. Participating members

The bulk of the deliberations over the formulation of the standards described in this document was performed by the Standard Development Working Group.

The members of the Working Group are listed below.

Standardization Committee

Chair	HATTORI Yuichiro	Canon Inc.
Vice Chair	SATOH Hitoshi	FUJIFILM Corporation
Vice Chair	IMAFUJI Kazuharu	NIKON CORPORATION
Vice Chair	YOSHIDA Hideaki	OM Digital Solutions Corporation
Vice Chair	FUKUSHIMA Tsumoru	Panasonic Corporation
Vice Chair	KATOH Naoya	Sony Corporation

Standard Development Working Group

Leader	MASUDA Hidetoshi	Canon Inc.
Sub Leader	USUI Kazutoshi	NIKON CORPORATION
	Scott Foshee	Adobe Systems Incorporated
	Paul Hubel	Apple, Inc.
	MORI Munehiro	Apple, Inc.
	MURAMATSU Kiyoji	Brother Industries, Ltd.
	TAKAGI Atsushi	Canon Inc.
	KONDO Shigeru	FUJIFILM Corporation
	SATOH Hitoshi	FUJIFILM Corporation
	SHIMIZU Masayoshi	Fujitsu Limited
	IMAI Tsutomu	Morpho, Inc.
	SHIMOYAMADA Yoshitaka	Nidec Copal Corporation
	IMAFUJI Kazuharu	NIKON CORPORATION
	YOSHIDA Hideaki	OM Digital Solutions Corporation
	FUKUSHIMA Tsumoru	Panasonic Corporation
	NUMAKO Norio	RICOH IMAGING COMPANY, LTD.
	KITAJIMA Tatsutoshi	RICOH IMAGING COMPANY, LTD.
	ENDO Masakatsu	Seiko Epson Corporation

SHIOHARA Ryuichi	Seiko Epson Corporation
YABASE Naoto	SIGMA CORPORATION
ISHIZAKA Toshihiro	Sony Corporation
KATOH Naoya	Sony Corporation
KANEHASHI Kenichi	Sony Corporation
MATSUI Akira	Sony Corporation
KUMA Toshitaka	Xacti Corporation.

CIPA DC-006-Translation-2021

Published in December 2021

Published by Camera & Imaging Products Association
MA Shibaura BLDG., 3-8-10, Shibaura, Minato-ku, Tokyo,
108-0023 JAPAN
TEL +81-3-5442-4800 FAX +81-3-5442-4801