INTERNATIONAL STANDARD

ISO/IEC 23000-11

First edition
2009-11-15

Information technology — Multimedia application format (MPEG-A) —

Part 11: Stereoscopic video application format

Technologies de l'information — Format pour application multimédia (MPEG-A) —

Partie 11: Format pour application vidéo stéréoscopique

Reference number
ISO/IEC 23000-11:2009(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

ISO/IEC 23000-11 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

ISO/IEC 23000 consists of the following parts, under the general title *Information technology — Multimedia application format (MPEG-A)*:

— *Part 1: Purpose for multimedia application formats* [Technical Report]

— *Part 2: MPEG music player application format*

— *Part 3: MPEG photo player application format*

— *Part 4: Musical slide show application format*

— *Part 5: Media streaming application format*

— *Part 6: Professional archival application format*

— *Part 7: Open access application format*

— *Part 8: Portable video application format*

— *Part 9: Digital Multimedia Broadcasting application format*

— *Part 10: Video surveillance application format*

— *Part 11: Stereoscopic video application format*

— *Part 12: Interactive music application format*

Introduction

In today's technological arena, there is an abundance of digital content for digital image machinery such as laptops, cell-phones, digital cameras, and mobile devices. Stereoscopic video contents provide users with an experience of natural three-dimensional scenes, which are displayed using acquisition and generation techniques. The market for applying stereoscopic video contents on such devices is taking shape and maturing. Stereoscopic laptops, mobile phones, digital TVs, and multimedia devices are already on the market; however, what seems to be required for an immersive 3D market is a standard file format which is capable of storage, interchange, management, editing, and presentation of stereoscopic video contents.

The Stereoscopic Video application format (AF) defines a file format for stereoscopic video services in mobile environments. It specifies core structures of stereoscopic video AF being organized by the combination of related information for stereoscopic video applications.

Applicable areas of the Stereoscopic Video AF are quite broad, including the internet, telecommunications, and storage devices. The user can download the Stereoscopic Video AF files from the internet or via the telecommunication networks to his/her personal multimedia devices (e.g. Portable Multimedia Player or cell-phone) for local playback.

Information technology — Multimedia application format (MPEG-A) —

Part 11:
Stereoscopic video application format

1 Scope

This part of ISO/IEC 23000 specifies a file format which is capable of storage, interchange, management, editing, and presentation of stereoscopic video contents based on the ISO base media file format. The file format provides the overall structure for storing stereoscopic video contents with the related stereoscopic information in mobile environments.

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

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

ISO/IEC 14496-2, *Information technology — Coding of audio-visual objects — Part 2: Visual*

ISO/IEC 14496-3, *Information technology — Coding of audio-visual objects — Part 3: Audio*

ISO/IEC 14496-10, *Information technology — Coding of audio-visual objects — Part 10: Advanced Video Coding*

ISO/IEC 14496-12, *Information technology — Coding of audio-visual objects — Part 12: ISO base media file format*

ISO/IEC 14496-20, *Information technology — Coding of audio-visual objects — Part 20: Lightweight Application Scene Representation (LASeR) and Simple Aggregation Format (SAF)*

ISO/IEC 15948:2004, *Information technology — Computer graphics and image processing — Portable Network Graphics (PNG): Functional specification*

3GPP TS 26.071, *Mandatory speech CODEC speech processing functions; AMR speech Codec; General description*

TIA/EIA/IS-127, *Enhanced Variable Rate Codec (EVRC)*

3 Terms and definitions

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

3.1
baseline
line between origins of the respective cameras

3.2
convergence distance
distance between a convergence point and a midpoint of baseline

3.3
convergence point
point at which two optical axes of left and right cameras intersect

3.4
disparity
horizontal difference between corresponding points in stereoscopic view

3.5
focal length
distance from a surface of a lens (optical center) or mirror to its focal point (image plane)

3.6
frame
one of the many still images which compose the complete moving picture

NOTE A frame contains an array of luma samples and two corresponding arrays of chroma samples. A frame consists of two fields: a top field and a bottom field.

3.7
lenticular
array of magnifying lenses designed so that, when viewed from slightly different angles, different images are magnified

NOTE A lenticular sheet is placed on a normal display panel to show two or more different views simply by changing the angle of light direction. It can make left and right views display on left and right eyes, respectively, creating a sense of depth.

3.8
max of disparity
maximum disparity value within a stereoscopic fragment

3.9
monoscopic fragment
set of successive samples which represents only monoscopic sequence

3.10
min of disparity
minimum disparity value within the stereoscopic fragment

3.11
parallax barrier
device to allow a liquid crystal display to show a three dimensional image without the need for the viewer to wear glasses

NOTE Placed in front of the normal display panel, a parallax barrier consists of a layer of material with a series of precision slits, allowing each eye to see a different set of pixels, so creating a sense of depth.

3.12
primary view sequence
sequence that has a priority of presentation between sequences of Left/Right view sequence type

3.13
rotation
relative angular variation from the primary-view camera to the secondary-view camera

3.14
secondary view sequence
sequence that has a lower priority of presentation than the primary view sequence between sequences of Left/Right view sequence type

3.15
sequence
series of one or more frames

3.16
stereoscopic camera information
information for stereoscopic camera parameters such as baseline, focal_length, convergence_distance, camera_arrangement, and rotation

3.17
stereoscopic display information
information for the stereoscopic display and visual safety, such as the display size and the viewing distance

3.18
stereoscopic fragment
set of successive samples which represents the stereoscopic sequence satisfying the stereoscopic composition type specified in this part of ISO/IEC 23000

3.19
stereoscopic left fragment
set of successive samples which represents the left view of stereoscopic sequences satisfying the stereoscopic composition type specified in this part of ISO/IEC 23000

3.20
stereoscopic left view sequence
left view sequence of the stereoscopic sequence

3.21
stereoscopic right fragment
set of successive samples which represents the right view of the stereoscopic sequences satisfying the stereoscopic composition type specified in this part of ISO/IEC 23000

3.22
stereoscopic right view sequence
right view sequence of the stereoscopic sequence

4 Abbreviated terms

3D	Three Dimensional
AAC	Advanced Audio Coding
AF	Application Format
AMR	Adaptive Multirate
AVC	Advanced Video Coding
CDMA	Code Division Multiple Access
EVRC	Enhanced Variable Rate Codec
GSM	Global Systems for Mobile communications
HE-AAC	High Efficiency AAC
JPEG	Joint Photographic Experts Group
LASeR	Lightweight Application Scene Representation
PNG	Portable Network Graphics
PMP	Portable Multimedia Player
UMTS	Universal Mobile Telecommunications System

5 Overview

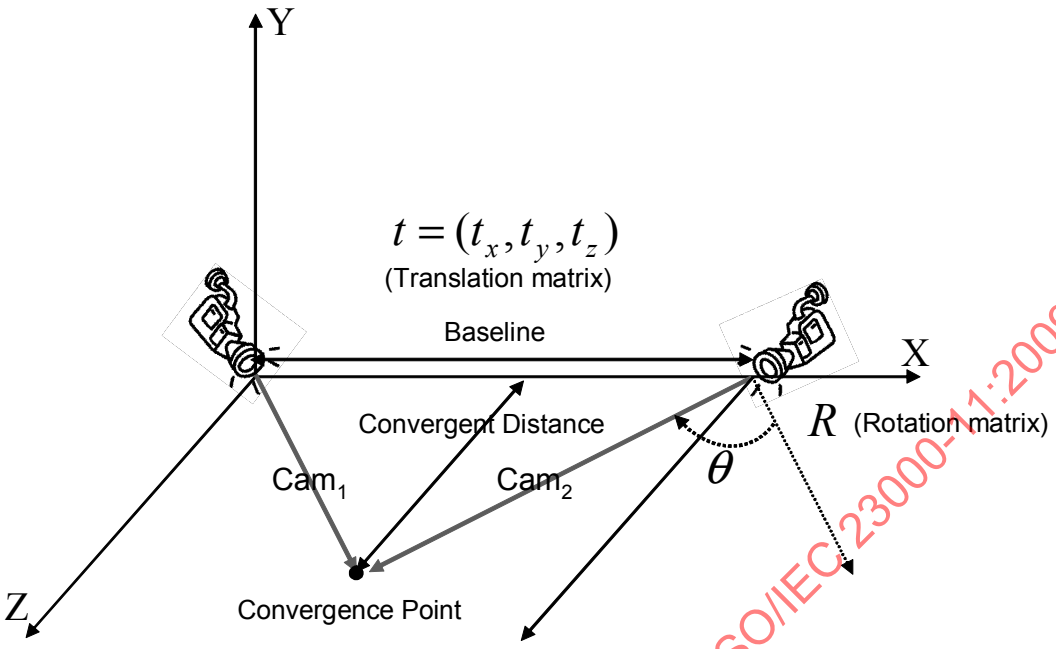
5.1 Overall procedure of stereoscopic contents

The overall procedure for stereoscopic contents can be explained as follows. Both left and right view sequences are acquired from a stereoscopic camera for stereoscopic video sequences, and are composited into a video sequence or two video sequences according to the composition types specified in 5.3. This composited video sequence is encoded and then stored into an AF.

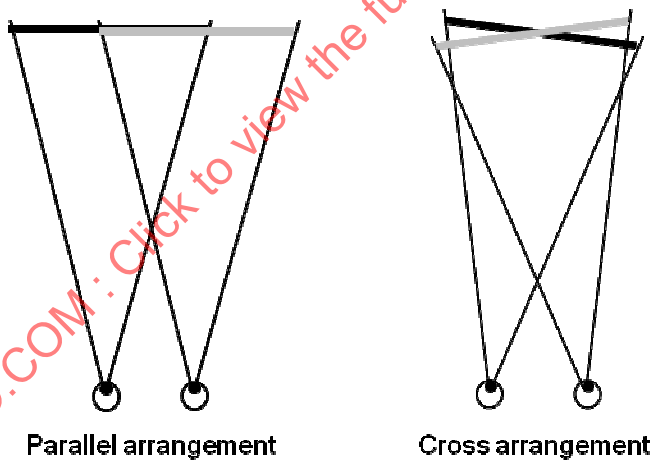
A file generator for Stereoscopic Video AF is to accept the stereoscopic contents with video, audio and LASeR streams. The file satisfying the Stereoscopic Video AF is parsed, decoded and then rendered for a stereoscopic display device.

5.2 Acquisition of the stereoscopic contents

Stereoscopic video sequences are acquired from two cameras, left and right view. As described in Figure 1, camera parameters shall be needed for specifying spatial relationship between two cameras. The stereoscopic contents can be rendered in the display device more precisely by using these camera parameters. The camera parameters shall be described in the `'scdi'` box, which will be specified in 8.5.



(a) Camera coordinates



(b) Camera arrangements

Figure 1 — Example of camera coordinate and camera arrangement used in Stereoscopic Video AF

Figure 1 (a) shows one example of camera coordinates used in the Stereoscopic Video AF. Cam1 and Cam2 indicate right and left views, respectively. This AF simplifies stereoscopic camera coordinates because it considers only stereoscopic contents suitable for binocular display system. In order to decrease the number of camera parameters we assume the coordinates of Cam1 is identical with world coordinates and Cam1 and Cam2 share X axis. Under these assumptions rotation information indicates relative angle value (θ) from Cam1 to Cam2 according to Y axis. Baseline distance means relative translation information of origins from Cam1 to Cam2. In addition, each focal length information is assumed to be identical because stereoscopic contents with different focal length can produce severe eye strain on binocular display. Figure 1 (b) shows camera arrangements– parallel arrangement and cross arrangement.

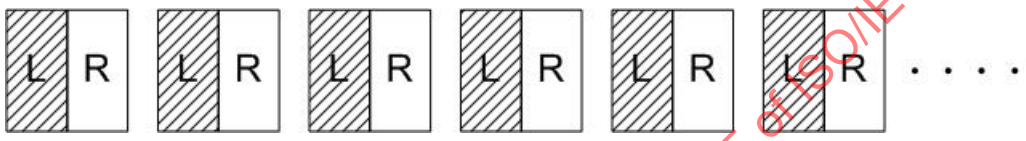
5.3 Stereoscopic contents composition type

In the current market, there are several stereoscopic composition types such as ‘side-by-side type’, ‘top and bottom type’, ‘pixel-by-pixel type’, ‘vertical line interleaved type’, ‘frame sequential type’, ‘Left/Right view sequence type’ and etc.

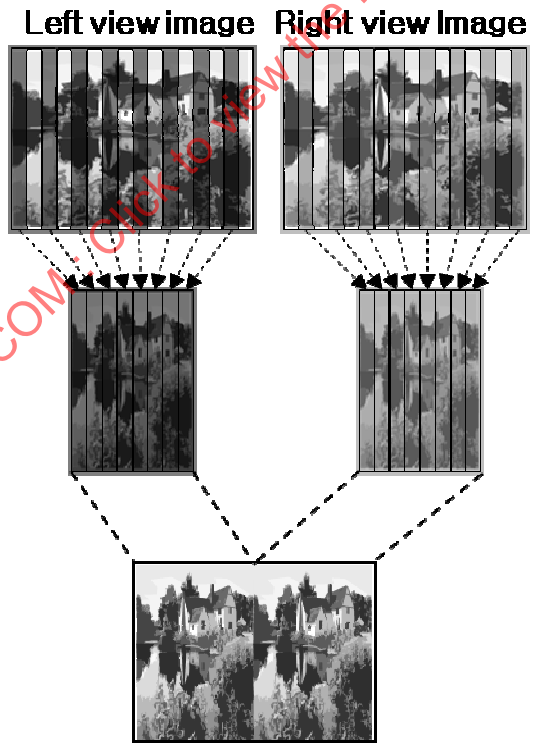
This specification, considering the wide usage and suitability for the mobile display, considers the composition types described in 5.3.1 to 5.3.4.

5.3.1 Side-by-side type

Side-by-side type is one of the most widely used stereoscopic composition types. Two respective left view and right view images are put together into one composition image by making their horizontal resolutions half as being shown in the Figure 2, which shows one example of side-by-side type when the left (right) view part locates in the left (right) side of composition image. It can be compressed in conventional bitrates although there is a quality loss due to the half resolution. In addition, it can be rendered in the legacy player and implemented without modification of the system.



(a) Side-by-side type stereoscopic sequence



(b) Side-by-side type contents for a real image

Figure 2 — Example of the side-by-side type

5.3.2 Vertical line interleaved type

A composition image of this type is made of repeated vertical lines of the left view and the right view images using this type. In the Figure 3, the vertical line of the left view firstly appears and then the vertical line of the right view follows after it. Due to the discontinuity between every vertical line, the compression efficiency is relatively poorer compared with other type. This type is supported by the parallax barrier display, which is most used in the stereoscopic mobile display. The contents can be directly displayed on the parallax barrier without converting them.

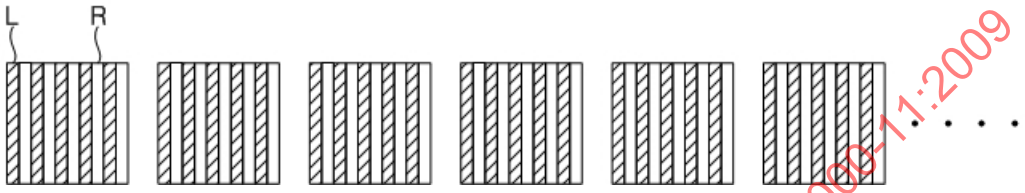


Figure 3 — Example of vertical line interleaved type

5.3.3 Frame sequential type

The frame sequential type is composed of successive left view and right view images as being shown in Figure 4. Some stereoscopic devices display left and right images sequentially while the other devices such as parallax barrier display left and right image in the same time and the same screen. If contents have double frame rate, this type provides full resolution with normal frame rate. In the following example in Figure 4, a left view precedes a right view.

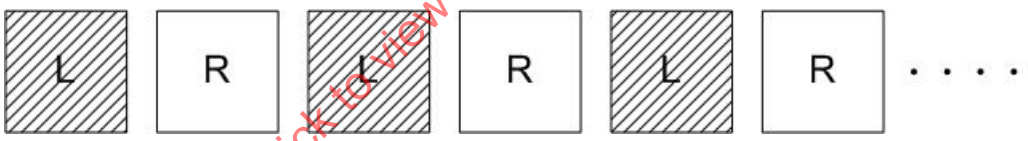


Figure 4 — Example of frame sequential type

5.3.4 Left/Right view sequence type

This type is composed of the independent elementary streams. For example, one stream represents the left view images and the other one does the right view images as shown in Figure 5. In this type, respective two images of left and right view shall be synchronized.

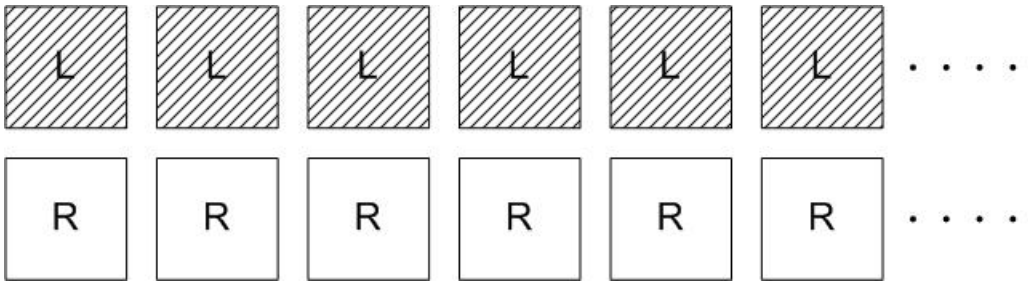


Figure 5 — Example of Left/Right view sequence type

6 Components of Stereoscopic Video AF

6.1 Supported components

Table 1 shows a brief summary of the supported components of the Stereoscopic Video AF which consists of the ISO/IEC Standards and non-ISO/IEC Standards .

The Stereoscopic Video AF includes ISO/IEC 14496-2 Simple Profile at Level 3 and ISO/IEC 14496-10 Baseline Profile at Level 1.3 for visual, ISO/IEC 14496-3 AAC and HE-AAC Profile for audio, 3GPP TS 26.071 AMR and TIA/EIA/IS-127 EVRC for voice, ISO/IEC 14496-20 LASeR for scene description, and various kind of image such as ISO/IEC 10918-1 JPEG and ISO/IEC 15948 PNG. For this specification, ISO/IEC 14496-12 ISO base media file format is used for a base file format structure.

Table 1 — Supported components of Stereoscopic Video AF

Type	Component Name	Specification	Standard
File format	ISO base media file format	ISO/IEC 14496-12	ISO/IEC Standards
Visual	MPEG-4 Video	ISO/IEC 14496-2 Simple Profile Level 3	
	MPEG-4 AVC	ISO/IEC 14496-10 Baseline Profile Level 1.3	
Audio	MPEG-4 Audio AAC	ISO/IEC 14496-3	
	MPEG-4 Audio HE-AAC	ISO/IEC 14496-3	
Data	MPEG-4 LASeR	ISO/IEC 14496-20	
	JPEG Image	ISO/IEC 10918-1	
	PNG Image	ISO/IEC 15948	
Voice	AMR	3GPP TS 26.071	Non-ISO/IEC Standards
	EVRC	TIA/EIA/IS-127	

6.1.1 ISO base media file format

ISO/IEC 14496-12 ISO base media file format is a flexible, extensible format which contains timed media information in order to facilitate interchange, management, editing, and presentation of the media as being shown in Figure 6. The ISO base media file format is a base format for the Stereoscopic Video AF.

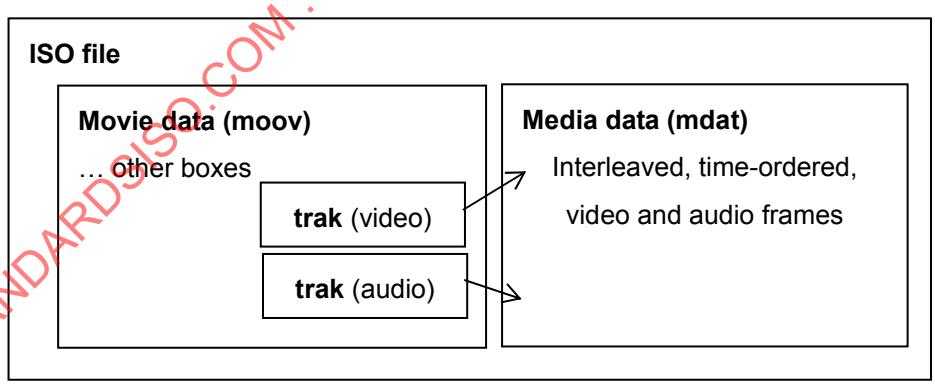


Figure 6 — Example of a simple ISO base media file format

6.1.2 LASeR

ISO/IEC 14496-20 LASeR is a scene description format that specifies various aspects of 2D scene representation and updates of scenes as a part of rich media content. A scene description is composed of graphics, animation, text, and spatial and temporal layout. The LASeR is designed to be suitable for lightweight embedded devices such as mobile phones and PMPs. The LASeR is used for supporting enrich interactive and 2D combined stereoscopic contents services.

6.1.3 AMR

3GPP TS 26.071 AMR is an audio data compression scheme optimized for speech coding. The AMR was adopted as the standard speech codec by 3GPP in October 1998 and is now widely used in GSM and UMTS. It uses link adaptation to select from one of eight different bit rates based on link conditions.

6.1.4 EVRC

TIA/EIA/IS-127 EVRC is a speech codec used in CDMA networks. It was developed in 1995 to replace the QCELP vocoder which used more bandwidth on the carrier's network, thus EVRC's primary goal was to offer the mobile carriers more capacity on their networks while not increasing the amount of bandwidth or wireless spectrum needed.

7 File structures

7.1 Table for boxes

The Stereoscopic Video AF contains various boxes based on the ISO base media file format. It provides new boxes such as the Stereoscopic Video Media Information (`'svmi'`) and the Stereoscopic Camera and Display Information (`'scdi'`).

The normative file structure consists of `'ftyp'`, `'moov'` and `'mdat'` boxes. Mandatory boxes are marked with an asterisk (*).

The `'ftyp'` box indicates the type of the file format which complies to the structure defined for the Stereoscopic Video AF. Thus, an application should be able to play Stereoscopic Video AF files when it supports the brands of `'ftyp'` box field. A detailed description of the brands of Stereoscopic Video AF is provided in 8.1.

The `'moov'` box contains one or more tracks for stereoscopic video sequences, a track for LASeR streams, and also contain tracks for audio, images, text and metadata.

The `'trak'` boxes contain temporal and spatial information of the media data (e.g. stereoscopic video sequences, stereo-monoscopic mixed video sequences, LASeR streams, JPEG images). For Stereoscopic Video AF, each track contains its associated `'mdia'` box, a `'tref'` box and a track level `'meta'` box.

The `'mdia'` box contains a `'svmi'` box for the stereoscopic visual type and fragment information of the stereoscopic contents in the track.

The `'tref'` box provides a `track_ID` of a reference track. In the Stereoscopic Video AF, the stereoscopic contents in the file structures can be separately stored for Left/Right view sequence type as shown in 5.3.4. Thus, the `'tref'` box is used for indicating a pair of stereoscopic left and right view sequences for the Left/Right view sequence type. A detailed description of the `'tref'` box for the Stereoscopic Video AF is provided in 8.2. In case of that the stereoscopic contents are contained in a single track, this box is not present.

The track level `'meta'` box should contain a 'scdi' box and an item location (`'iloc'`) box. The `'scdi'` box provides the information of stereoscopic camera, display and visual safety.

The `'iloc'` box describes the absolute offset in bytes (`extent_offset`) and the size (`extent_length`) of stereoscopic fragments. An `item_ID` is assigned to each fragment of the stereoscopic sequence for resource referencing.

A LASeR track structure is specified in 10.1 of ISO/IEC 14496-20:2008. A richer scene description with stereoscopic contents is specified by the LASeR stream. LASeR refers to stereoscopic tracks by their `track_ID`s.

The 'mdat' box contains the media data which are described in the 'trak' boxes.

The following Table 2 briefly shows the structure of the boxes and their descriptions.

Table 2 — Table for boxes of Stereoscopic Video AF

*	ftyp						*file type and compatibility*
	pdin						*Progressive download Information*
*	moov						*container for all the metadata*
*		mvhd					*movie header, overall declarations*
*		trak					*container for an individual track or stream*
*			tkhd				*track header, overall information about the track*
			tref				*track reference container* *"svdp" for Left/Right view sequence type*
			edts				*edit list container*
				elst			*an edit list*
*			mdia				*container for the media information in a track*
*				mdhd			*media header, overall information about the media*
*				hdlr			*handler, declares the media (handler) type* *"soun" for audio data* *"vide" for visual data* *"sdsm" for LASeR data*
*				minf			*media information container*
					vmhd		*video media header, overall information (video track only)*
					smhd		*sound media header, overall information(sound track only)*
					hmhd		*hint media header, overall information (hint track only)*
					nmhd		*null media header, overall information(some tracks only)*
*					dinf		*data information box, container*
*						dref	*data reference box, declares soure(s) of media data in track*
*					stbl		*sample table box, container for the time/space map*
*						stsd	*sample descriptions (codec types, initialization etc.)* *"lsr1" for LASeR data* *"samr" for AMR NB data* *"sawb" for AMR WB data* *"sevc" for EVRC data*
*						stts	*(decoding) time-to-sample*
						ctts	*(composition) time to sample*
*						stsc	*sample-to-chunk, partial data-offset information*
*						stsz	*sample sizes (framing)*
						stz2	*compact sample sizes (framing)*
*						stco	*chunk offset, partial data-offset information*
						co64	*64-bit chunk offset*
						stss	*sync sample table (random access points)*
*						svmi	*stereoscopic video media infomration*
		lpmc					*IPMP Control Box*
	mdat						*media data container*
	meta						*Metadata*
*		hdlr					*handler, declares the metadata (handler) type*
		iloc					*item location*
		iinf					*item information*
		xml					*XML container*
		bxml					*binary XML container*
		scdi					*stereoscopic camera and display infomration*

7.2 File structures of Stereoscopic Video AF

This subclause describes various structures of the Stereoscopic Video AF specification.

Stereoscopic video sequence can be stored in one video track, and also in two video tracks according to the composition type described in subcluase 5.3. In the latter case, they shall have a dependency between each other. The reference track uses a `reference_type` of `'svdp'` in the `'tref'` to describe the dependency. The detail structure is specified in 7.2.1.

The Stereoscopic Video AF supports either only stereoscopic contents or/and stereo-monoscopic mixed contents. The case of using stereo-monoscopic mixed contents and its related structure are specified in 7.2.2.

7.2.1 File structure for stereoscopic contents

The track for stereoscopic contents contains pure stereoscopic contents or stereo-monoscopic mixed contents. This subclause describes the former file structures.

Figure 7 shows an example of the file structure, containing stereoscopic contents in a single track. The stereoscopic content, which composed by 'side-by-side type', 'vertical line interleaved type' and 'frame sequential type' described in 5.3, shall be contained in the single track.

The tracks are contained in `'moov'` box, one for the stereoscopic contents and the other for the LASeR stream. The track for stereoscopic contents has the `'mdia'` box and the track level `'meta'` box. The stereoscopic video track contains the `'svmi'` box for one set of consecutive samples and so only one `item_ID` is assigned for supporting the `'scdi'` of the stereoscopic video track. The `extent_offset` and `extent_length` indicate the absolute offset and the length of the item in bytes.

As described in ISO/IEC 14496-20, the LASeR track use a handler_type of `'sdsm'` in the handler reference box, a video media header `'vmhd'` and a derivative of the SampleEntry.

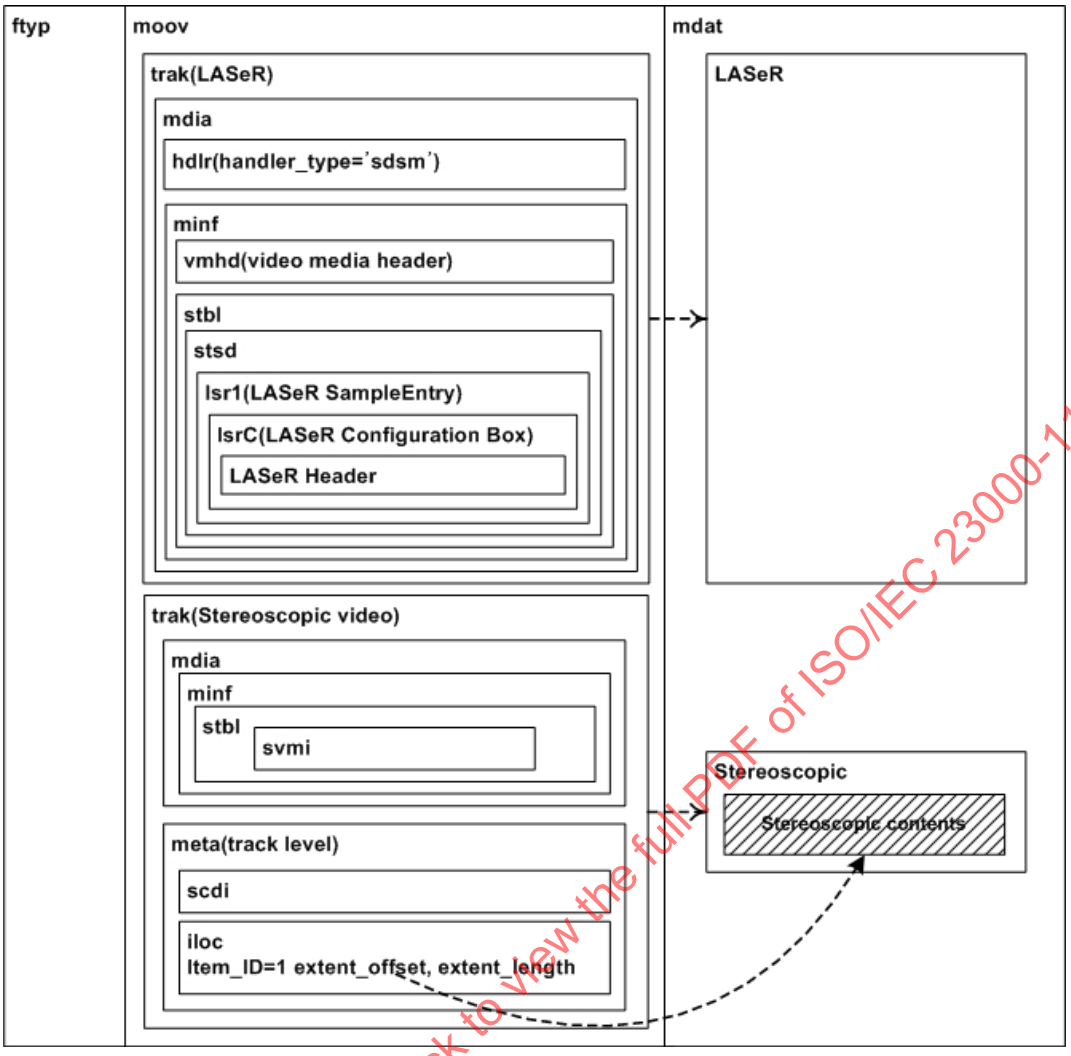


Figure 7 — Example of a file structure for a single stereoscopic track

An example of the file structure for Left/Right view sequence type is illustrated in Figure 8, which describes the file structure of a stereoscopic contents specified in 5.3.4, that is, the composition type for storing the left and right view sequences of the stereoscopic contents in two separate tracks depicted in Figure 5.

The `'moov'` box contains two tracks for Left/Right view sequences and a track for LASeR stream. The tracks for stereoscopic contents include the `'tref'` box, the `'mdia'` box and the track level `'meta'` box, where two view sequences shall be paired with each other view sequence using the `'tref'` box. In Figure 8, the track of `track_ID`=2 contains a `'tref'` box which represents the `reference_type` = `'svdp'` and thus, this track would have `'scdi'` box, and also, the referenced track is the one with `track_ID`=1. In this specification, the track with a `'tref'` box is the secondary view sequence, and the referenced track is the primary view sequence.

The `'mdia'` box contains the `'svmi'` box.

According to the clause 3, when a single camera of the stereoscopic cameras for one view is set to origin, then all the parameters of the `'scdi'` for this camera shall be set to 0, and the `'scdi'` for the other camera shall have the parameter values relative to the origin.

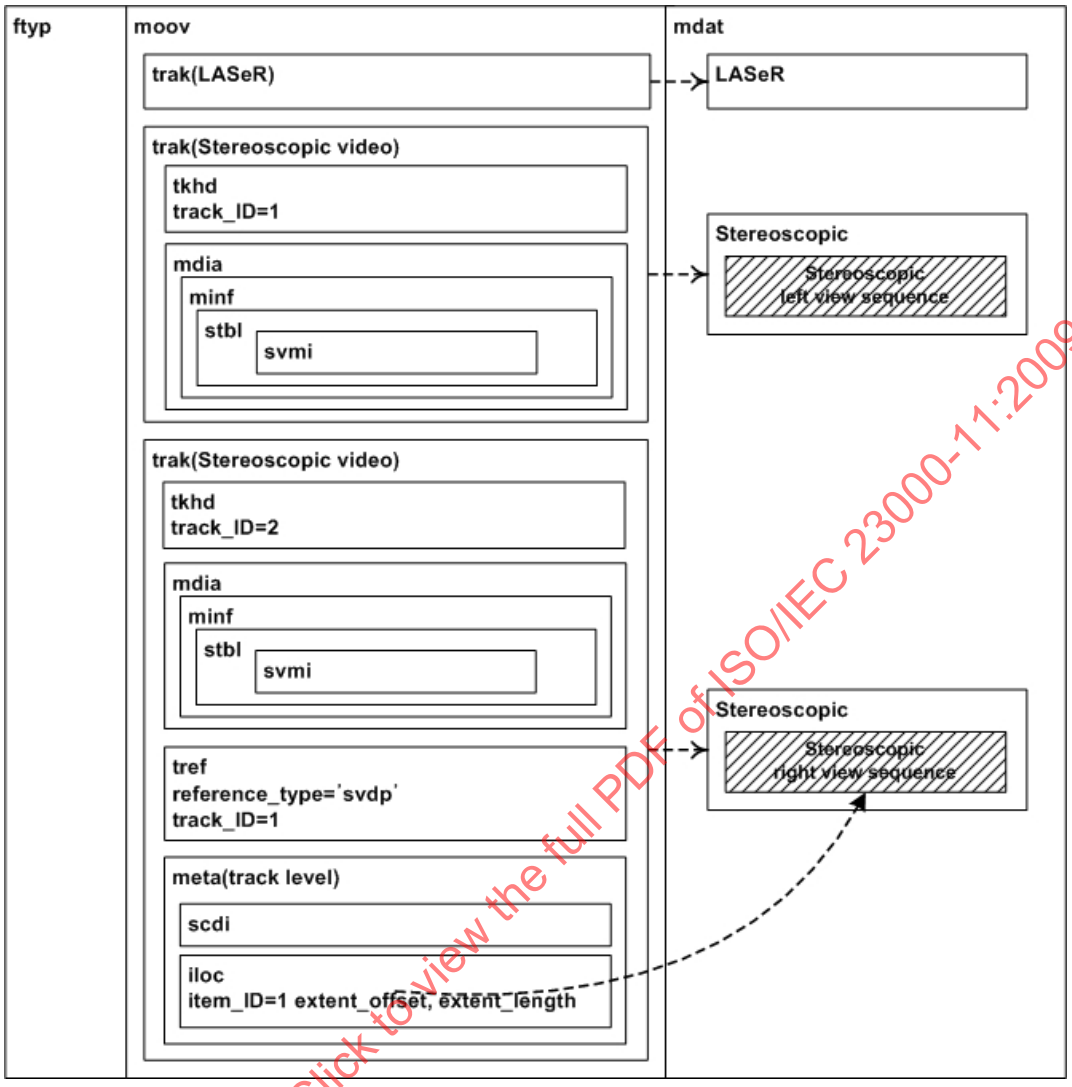


Figure 8 — Example of a file structure for Left/Right view sequence type

7.2.2 File structure for stereo-monoscopic mixed contents

This subclause describes the file structures for a stereo-monoscopic mixed content, which is a video sequence consisting of both stereoscopic and monoscopic fragments in a single track. The stereoscopic and monoscopic fragments should be stored sequentially.

Figure 9 shows an example of the file structure containing a single track for a stereo-monoscopic mixed content on the basis of the file format structure as shown in Figure 7. The `item_ID` under `'iloc'` box is assigned to each stereoscopic fragment sequentially. For example, when a stereoscopic contents is composed as illustrated in the below figure (S-M-S), the `item_ID` of the first fragment in the track, which is the first stereoscopic fragment, is set to 1, and the `item_ID` of the third one (second stereoscopic fragment) in the track is set to 2.

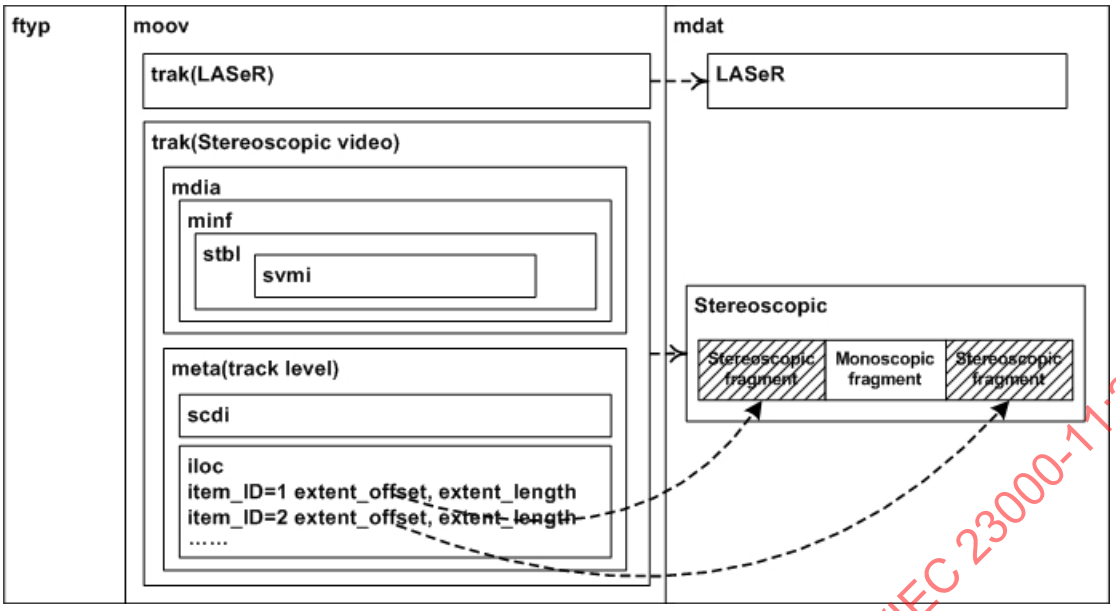


Figure 9 — Example of a file structure for stereoscopic and monoscopic fragments in a single stereoscopic track

Figure 10 describes the file structure of a stereoscopic contents specified in 5.3.4, the composition type for storing the left and the right view sequence of stereoscopic contents in two separate tracks. Stereoscopic fragments of each track have one view sequence on the basis of the file format structure as shown in Figure 8. The `item_ID` is assigned to each stereoscopic fragment of only one track sequentially.

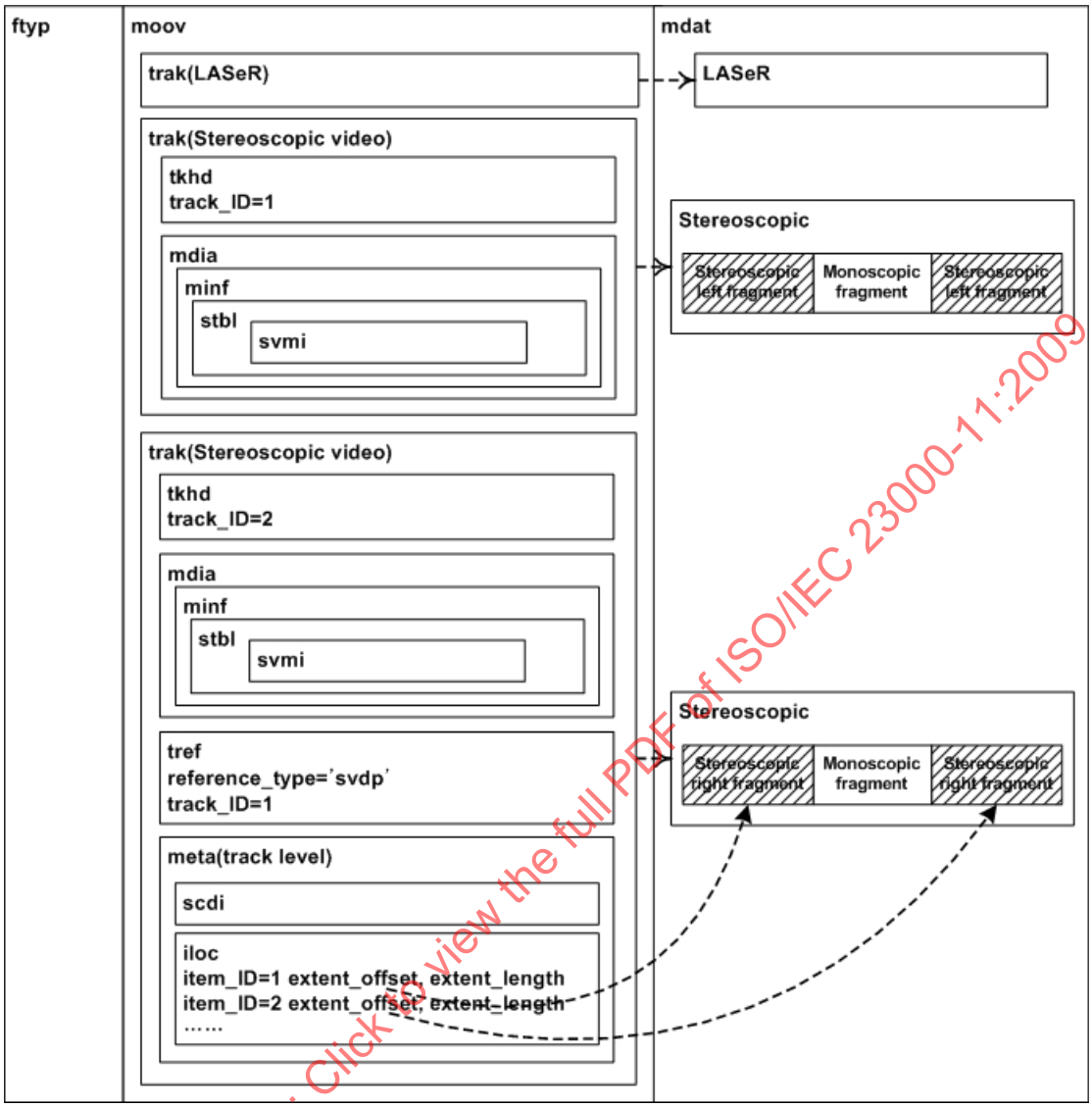


Figure 10 — Example of a file structure for stereoscopic and monoscopic fragments in Left/Right view sequence type

In case of stereo-monoscopic mixed contents being shown in Figure 10, it could cause the same time stamp for monoscopic fragments in the individual tracks. This ambiguity of presentation can be figured out as follows:

1. Check which track is indicating a primary view sequence by the `reference_type` and `track_ID` of the `'tref'` box in the track.
2. Display the each monoscopic fragment of primary view sequence.

8 Syntax and Semantics of the Boxes

8.1 File Type Box

8.1.1 Definition

Box Type : `'ftyp'`
Container: File
Mandatory: Yes
Quantity: Exactly one

The 'ftyp' box is used to identify the type of the Stereoscopic Video AF that this file structure complies to. The brand that identifies files conformant to this specification is 'ss01' and 'ss02' as shown in the Table 3.

Table 3 — The brand of stereoscopic contents

Types	Specifications
ss01	Stereoscopic content
ss02	Stereo-monoscopic mixed content

8.2 Track Reference Box

8.2.1 Definition

Box Type : 'tref'
Container: Track Box ('trak')
Mandatory: No
Quantity: Zero or one

In the case of the Stereoscopic Video AF, the 'tref' box shall be used to identify the track of primary view and secondary view sequences for Left/Right view sequence type

8.2.2 Syntax

```
aligned(8) class TrackReferenceBox extends Box('tref') {}
aligned(8) class TrackReferenceTypeBox (unsigned int(32) reference_type) extends
Box(reference_type) {
   unsigned int(32) track_IDs[];
}
```

8.2.3 Semantics

The Track Reference Box contains track reference type boxes.

track_ID - is an integer that provides a reference from the containing track to another track in the presentation. track_IDs are never re-used and cannot be equal to zero.
reference_type - shall be set to one of the following values:

- 'hint' the referenced track(s) contain the original media for this hint track.
- 'cdsc' this track describes the referenced track.
- 'hind' this track depends on the referenced hint track, i.e., it should only be used if the referenced hint track is used.
- 'svdp' this track describes a reference track, which has a dependency to a referenced track, and also contains the stereoscopic related meta information.

8.3 Sync Sample Box

8.3.1 Definition

Box Type : 'stss'
Container: Sample Table Box ('stbl')
Mandatory: No
Quantity: Zero or one

In case of a stereoscopic contents with Left/Right view sequence type, the 'stss' box which is in the track for the primary view sequence is used for random access.

8.4 Stereoscopic Video Media Information Box

8.4.1 Definition

Box Type : `'svmi'`
Container: Sample Table Box (`'stbl'`)
Mandatory: Yes
Quantity: Exactly one

The `'svmi'` box provides stereoscopic video media information regarding the stereoscopic visual type and also, for the care of some mixed contents, fragments information. The visual type information signals the composition type of the stereoscopic video sequence and the structure of fragments. The fragment information represents the number of fragments, the number of consecutive samples, and whether the current sample is stereoscopic or not.

8.4.2 Syntax

```
aligned(8) class StereoscopicVideoMediaInformationBox extends
   FullBox('svmi', version = 0, 0){
   // stereoscopic visual type information
   unsigned int(8)  stereoscopic_composition_type;
   unsigned int(7)  reserved = 0;
   unsigned int(1)  is_left_first;

   // stereo_mono_change information
   unsigned int(32) stereo_mono_change_count;
   for(i=1; i<=stereo_mono_change_count; i++){
      unsigned int(32)  sample_count;
      unsigned int(7)   reserved = 0;
      unsigned int(1)   stereo_flag;
   }
}
```

8.4.3 Semantics

`stereoscopic_composition_type` - the type of stereoscopic contents that are specified as the following Table 4.

Table 4 — Stereoscopic composition type

Value	Stereoscopic_composition_type
0x00	Side-by-side type
0x01	Vertical line interleaved type
0x02	Frame sequential type
0x03	Left/Right view sequence type
0x04-0xFF	Reserved

`is_left_first` – represents of positions of left and right view sequence for 3D mobile devices as being specified in Table 5. When `is_left_first` is '1' and current stereoscopic video is composed of side-by-side type, left side and right side of the image means left view and right view, respectively. When `is_left_first` is '0', left side and right side means right view and left view, respectively. When `is_left_first` is '1' and current stereoscopic video is composed of vertical line interleaved type, odd line and even line of the image means left view and right view, respectively. When `is_left_first` is '0', odd line and even line means right view and left view, respectively. When `is_left_first` is '1' and current stereoscopic video is composed of frame sequential type, odd frame and even frame of the sequence means left view and right view, respectively. When `is_left_first` is '0', odd frame and even frame means right view and left view, respectively. When `is_left_first` is '1' and current stereoscopic video is composed of Left/Right view sequence type, primary view sequence and secondary view sequence means left view and right view,

respectively. When `is_left_first` is '0', primary view sequence and secondary view sequence means right view and left view, respectively.

Table 5 — The positions of stereoscopic Left/Right view according to the is_left_first value

Type	is_left_first = 1		is_left_first = 0	
	Left view	Right view	Left view	Right view
Side-by-side	Left side	Right side	Right side	Left side
Vertical line interleaved	Odd line	Even line	Even line	Odd line
Frame sequential	Odd frame	Even frame	Even frame	Odd frame
Left/Right view sequence	Primary view sequence	Secondary view sequence	Secondary view sequence	Primary view sequence

`stereo_mono_change_count` - is an integer that gives the number of fragments when stereoscopic to/from monoscopic fragment changes. In case of the `'ftyp'` is `'ss01'`, `stereo_mono_change_count` is set to be 1.
`sample_count` – is an integer that counts the number of consecutive samples.
`stereo_flag` – represents whether the current sample is stereoscopic or not. If this value is 1, then the current sample is stereoscopic, and if this value is 0, then the current sample is monoscopic.

8.5 Stereoscopic Camera and Display Information Box

8.5.1 Definition

Box Type : `'scdi'`
Container: Meta Box (`'meta'`)
Mandatory: No
Quantity: Zero or one

The `'scdi'` box, an optional box, provides primary information of the stereoscopic camera, display and visual safety. Stereoscopic camera and display information specified in this box can be described for stereoscopic fragments. Each fragment including `'scdi'` has a unique `item_ID` which is an identifier to be referenced by other fragments.

8.5.2 Syntax

```
aligned(8) class StereoscopicCameraAndDisplayInformationBox extends
   FullBox('scdi', version = 0, 0){
   unsigned int (16) item_count;
   for( i=0; i<item_count; i++ ){
      unsigned int(16)     item_ID;
      unsigned int(7)      reserved = 0;
      unsigned int(1)      is_item_ID_ref;
      if(is_item_ID_ref){
         unsigned int(16)  ref_item_ID;
      }
      else{
         // stereoscopic display information
         unsigned int(4)   reserved = 0;
         unsigned int(3)   3D_diplay_type;
         unsigned int(1)   is_display_safety_info;
         if(is_display_safety_info) {
            unsigend int(16)  expected_display_width;
            unsigend int(16)  expected_display_height;
            unsigend int(16)  expected_viewing_distance;
            int(16)           min_of_disparity;
            int(16)           max_of_disparity;
         }
```