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

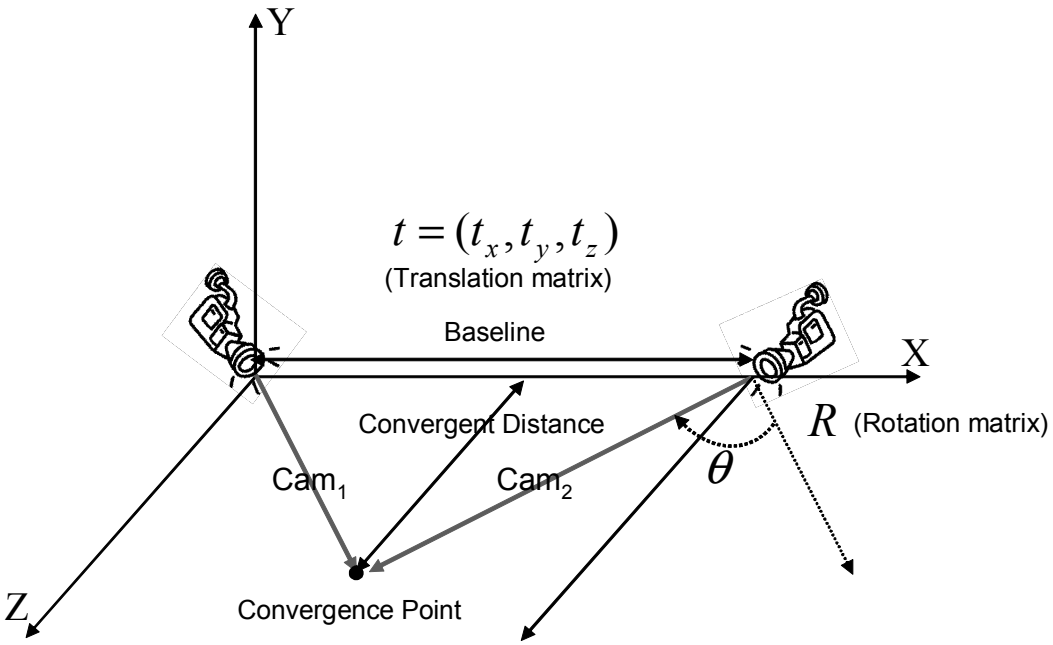
5.1 Overall procedure of stereoscopic contents

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.

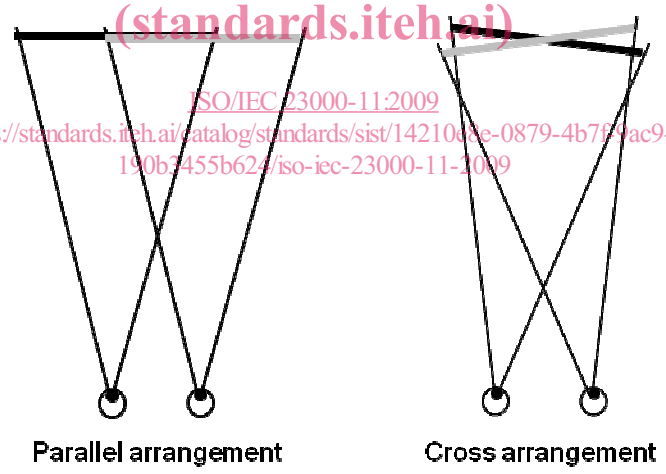
A file generator for Stereoscopic Video AF is to accept the stereoscopic contents with video, audio and LASeR streams. The file satisfying the Stereoscopic Video AF is parsed, decoded and then rendered for a stereoscopic display device.

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.

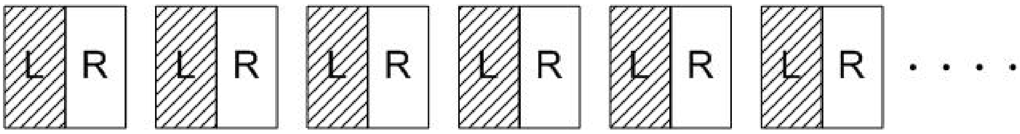
5.3 Stereoscopic contents composition type

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.

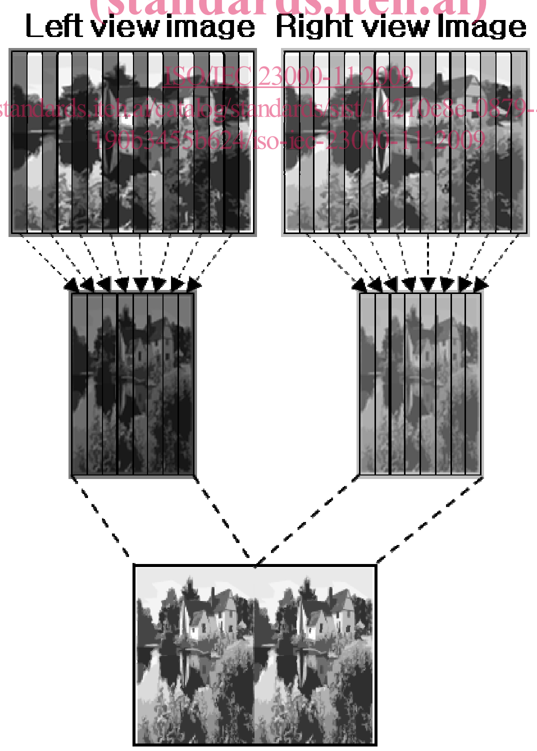
This specification, considering the wide usage and suitability for the mobile display, considers the composition types described in 5.3.1 to 5.3.4.

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