INTERNATIONAL STANDARD

ISO/IEC 23000-10

First edition
2009-05-01

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

Part 10: Video surveillance application format

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

Partie 10: Format pour application à la vidéosurveillance

Reference number
ISO/IEC 23000-10:2009(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 23000-10 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*

Introduction

ISO/IEC 23000 (also known as “MPEG-A”) is an MPEG standard that supports a fast track to standardization by selecting readily tested and verified tools taken from the MPEG body of standards and combining them to form an AF (Application Format). If a needed piece of technology is not provided within MPEG, then additional technologies originating from other organizations can be included by reference in order to facilitate the envisioned AF.

The Video surveillance AF is a file format designed to provide for a first level of interoperability for video-based surveillance systems. It contains MPEG-4 AVC video data and associated MPEG-7 metadata. Usage of other coded video formats will be assisted.

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

Part 10: Video surveillance application format

1 Scope

This part of ISO/IEC 23000 specifies a file format designed to provide for a first level of interoperability for video-based surveillance systems. The file format provides the overall structure for storing video content and associated metadata in a single file.

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 9834-8:2005, *Information technology — Open Systems Interconnection — Procedures for the operation of OSI Registration Authorities: Generation and registration of Universally Unique Identifiers (UUIDs) and their use as ASN.1 Object Identifier components*

ISO/IEC 14496-12:2008, *Information technology — Coding of audio-visual objects — Part 12: ISO base media file format*

ISO/IEC 14496-15:2004, *Information technology — Coding of audio-visual objects — Part 15: Advanced Video Coding (AVC) file format*

ISO/IEC 15938-5:2003, *Information technology — Multimedia content description interface — Part 5: Multimedia description schemes*

ISO/IEC 23001-1:2006, *Information technology — MPEG systems technologies — Part 1: Binary MPEG format for XML*

3 Overview of MPEG Standards Used

3.1 MPEG-4 Advanced Video Coding

ISO/IEC 14496-10 Advanced Video Coding (AVC) is a digital video codec designed to achieve increased compression performance while providing network-friendly data transmission capabilities. The standard was prepared by the Joint Video Team (JVT) which is a collaborative partnership between the ITU-T Video Coding Expert Group (VCEG) and the Moving Picture Experts Group (MPEG). The ITU-T H.264 and the ISO/IEC MPEG-4 Part 10 standard are technically identical. The H.264/AVC project was intended to create a standard that would provide good video quality at substantially lower bit rates than the previous standards (i.e. relative to MPEG-2, H.263, or MPEG-4 Part 2). Application areas covered by the standard are conversational

as well as non-conversational services. The latter comprises broadcast, streaming and surveillance applications.

A conceptually distinction has been made in the specification between a video coding layer (VLC) and a network abstraction layer (NAL). The VLC comprises the signal processing part of the codec e.g. transform, quantization, etc. The output of the VLC is referred to as slices containing an integer number of macroblocks and the information of the slice header. A macroblock being a 16x16 block of luma and corresponding chroma samples.

The NAL provides formatting and encapsulation of the VLC output in a way compliant to the chosen transmission channel or storage media. Packet-oriented as well as bitstream systems are being supported by adding appropriate header information.

Higher layer meta information necessary to appropriately handle the data and to operate the decoder are conveyed in parameter sets. The specification distinguishes between two types of parameter sets: sequence parameter set and picture parameter set. An active sequence parameter set remains unchanged throughout a coded video sequence and an active picture parameter set remains unchanged within a coded picture. Higher layer meta information is supposed to be transmitted reliably and in advance.

A main property of the specification is the decoupling of the decoding process and time (e.g. sampling time, transmission time, presentation time, etc.)

The design requires only 16-bit arithmetic for processing on encoding and decoding side. Furthermore it is the first MPEG video standard achieving exact quality of decoded video because of the definition of an exact-match inverse transform.

3.2 ISO Base Media File Format

The ISO Base Media File Format [see ISO/IEC 14496-12:2008] is designed to contain timed media information for a presentation in a flexible, extensible format that facilitates interchange, management, editing, and presentation of the media. The ISO Base Media File Format is a base format for media file formats. Also the storage format for AVC coded video – the AVC file format [see ISO/IEC 14496-15:2004] – uses the techniques from the ISO Base Media File Format.

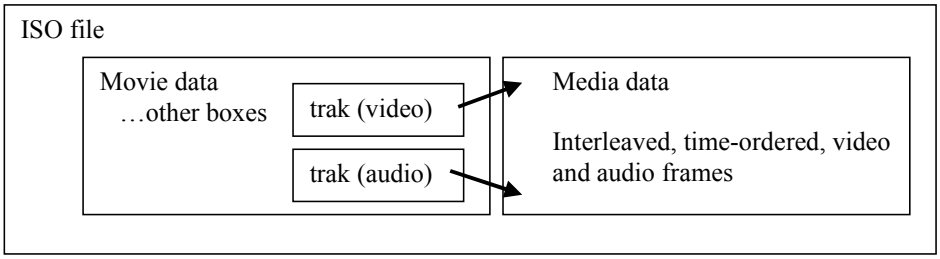

Figure 1 — Example of a simple ISO file used for interchange, containing two streams

The file structure is object-oriented as shown in Figure 1, which means that a file can be decomposed into constituent objects very simply, and the structure of the objects inferred directly from their type. The file format is designed to be independent of any particular network protocol while enabling efficient support for them in general.

It also provides support for metadata in the form of 'meta' boxes at the File, Movie and Track level. This allows support for static (untimed) metadata. Figure 2 schematically illustrates the location of these untimed MPEG-7 Metadata boxes. However, the ISO Base Media File Format also supports storage of timed metadata. These metadata can be synchronized with the video tracks and provide additional information e.g. time code values.

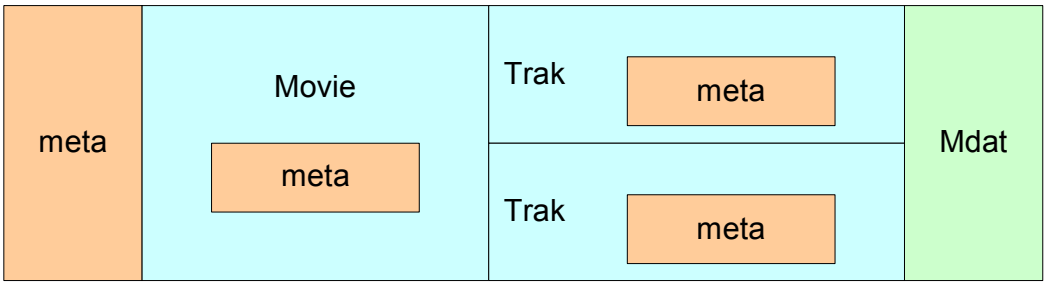

Figure 2 — Support of Static untimed Metadata in ISO/MP4 Files

If it is desired to play parts of a file while the file is still being recorded the media data should be physically stored in a different physical file e.g. on a disc (see ISO/IEC 14496-12:2008 A.3 - physical structure of the media in the ISO Base Media File Format). Movie fragments can be used to enable such feature.

Movie fragments can be used to enable features such as instant replay. In general, all data describing timing, properties and locations of individual video samples are contained in tables within a track. Usually these tables can only be written if all samples of the track are known. To overcome this burden the ISO Base Media File Format specifies the usage of movie fragments to extend a presentation in time. In the surveillance video AF the movie box may contain no or just a limited number of samples (in all the tracks) and the necessary initialization data. Additional samples are described in one or more movie fragments, depending on the use case, e.g. to enable instant replay functionality (the file is played while it is still being recorded).

Each track fragment contains a number of track fragment runs describing the samples individually. If some properties are identical for all samples in a fragment this value can be stored in the track fragment header (e.g. sample duration for constant frame rate video). In a track fragment run all samples are described by a constant number of 32 bit values.

A Video surveillance AF using movie fragments should define track fragment runs with a predictable number of samples in each fragment run and a defined number of track fragment runs in a track fragment.

While writing the file, the video data chunks (and chunks from other tracks) are appended to the end of the media data container which is reasonably physically located at the end of the file or in a separate physical file. The descriptive data about the media data samples is written to the reserved space for movie fragments – it is appended to the track fragment run table. Additionally the number of samples is changed in this track fragment run (see Figure 6). If a new track fragment run is to be created it is appended at the end of the previous track fragment run. Additionally the size of the track fragment box is changed. The same applies for creating new track fragments or new movie fragments. If the space reserved for movie and track fragments is fully used no more samples can be added and a new Video surveillance AF fragment should be created as described in 4.2.

If the file is to be read while it is still being written the reader can access all needed information in the movie and track fragments. The track fragment run table which is currently being written can be accessed up to the sample number given with the sample count value of this track fragment run.

Note that for every video sample a metadata sample must be provided. Therefore the technique described here must be used for all the video tracks and for all corresponding metadata tracks. When using more than one video track it must be ensured that all tracks have the same total duration.

When a Video surveillance AF fragment is being recorded the duration of this Video surveillance AF fragment should be set to zero to indicate that the duration is currently changing. In this case a player application should scan the track/movie fragment boxes to calculate the movie duration.

Special attention must be paid when using edit lists with movie and track fragments to create a compliant presentation.

3.3 MPEG-7 Multimedia Description Scheme

ISO/IEC 15938-5 Multimedia description scheme (MDS) [see ISO/IEC 15938-5:2003] is providing information about content description, management and organization for stored or streamed applications. Furthermore, description schemas are supporting navigation and access as well as user interaction with audiovisual content in real-time or non-real-time environments. Description schemas are the shell or wrapper for other description tools.

3.4 MPEG-7 Visual

ISO/IEC 15938-3 Visual is providing for elementary as well as more sophisticated descriptors for the following categories of features: colour, texture, shape, motion, localization, and face recognition.

3.5 AVC File Format

This AF uses the AVC file format to store the coded video data. ISO/IEC 14496-15 defines the storage of video coded using the ISO/IEC 14496-10 standard.

4 Using the Video surveillance AF

4.1 General

This clause provides necessary information for creating and using Video surveillance AF fragments.

It describes the box types that Video surveillance AF readers will recognize. Other box types may be included but will not be recognized.

4.2 File Structure

A Video surveillance AF contains of a set of self-contained AF fragments which are connected to each other. A Video surveillance AF fragment covers a limited amount of time. Each Video surveillance AF fragment is identified by a UUID (universal unique identifier) [see ISO/IEC 9834-8:2005]. Each Video surveillance AF fragment is linked to a predecessor and successor fragment through their UUIDs (see Figure 3).

All Video surveillance AF data is stored within the Video surveillance AF fragments. If a fragment has no predecessor or successor its value is set the current fragment. Additionally a URI can be given serving as a hint to the location of the predecessor and successor fragments. A Video surveillance AF fragment remains self contained even if unhinged. Note that there is no requirement to use more than one Video surveillance AF fragment. The concept of using fragments e.g. enables ring buffer architectures.

Each fragment shall be a valid AVC file as defined by the AVC file format.

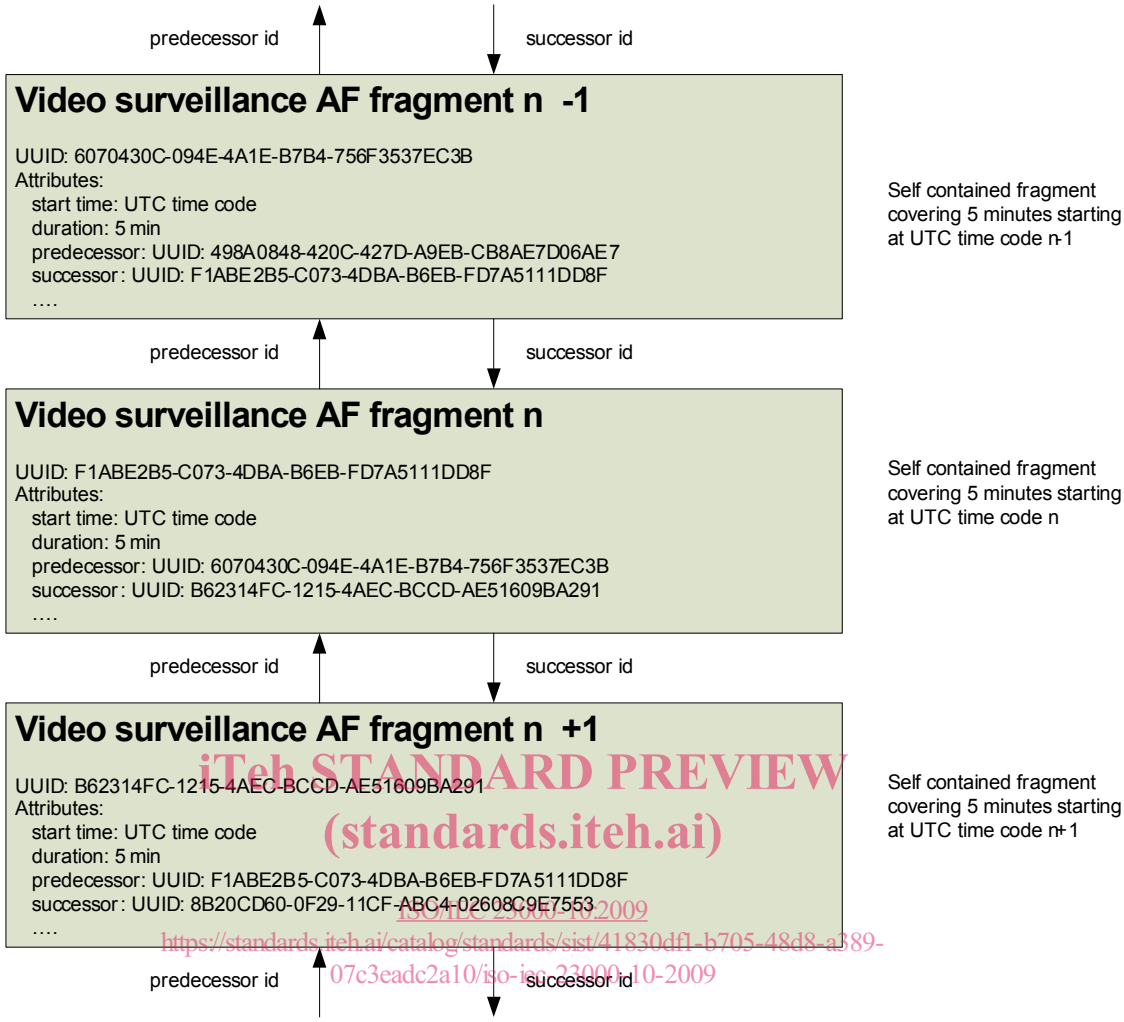

Figure 3 — VSAF fragments linked together by means of predecessor id and successor id

All Video surveillance AF fragments shall use the same number of tracks and the same set of parameters as timing and video coding settings.

The size of a Video surveillance AF fragment can be set as indicated by the application, e.g. providing a constant number of samples in each Video surveillance AF fragment.

Each fragment shall contain the mandatory metadata boxes and may contain additional metadata boxes as specified in Clause 6.

Managing the storage of Video surveillance AF fragments and the connection of fragments to the application is out of the scope of this part of ISO 23000.

4.3 File Contents

The file format for the Video surveillance AF is based on the ISO Base Media File Format. A Video surveillance AF fragment shall contain:

- One or more track boxes of vide type
- One box of meta type at file level and one for each video track at track level
- One or more tracks of timed metadata.

The above Meta Boxes may each additionally contain a further box, containing descriptive metadata as described in Annex B.

4.4 Track Structure

An AF fragment consists of at least one AVC video track (see Clause 5 on restrictions creating the AVC video). If more than one video track from one camera is present these video tracks shall be in the same alternate group (see 4.5.3 on track selection). Additionally, each video track shall link to a metadata track using a track reference (see 6.4 on the metadata tracks and sample structure).

Different video tracks may contain the same video content coded with different parameters or using a different coding technology (at least on video track must be coded as described in Clause 5). Alternatively different video tracks may contain different content, e.g. different views of the area monitored (see Figure 4 and Figure 5).

If there is more than one video track all video tracks shall have the same duration.

NOTE This does not imply that all tracks have the same number of samples. Different video tracks containing the same video content may be coded using different frame rates.

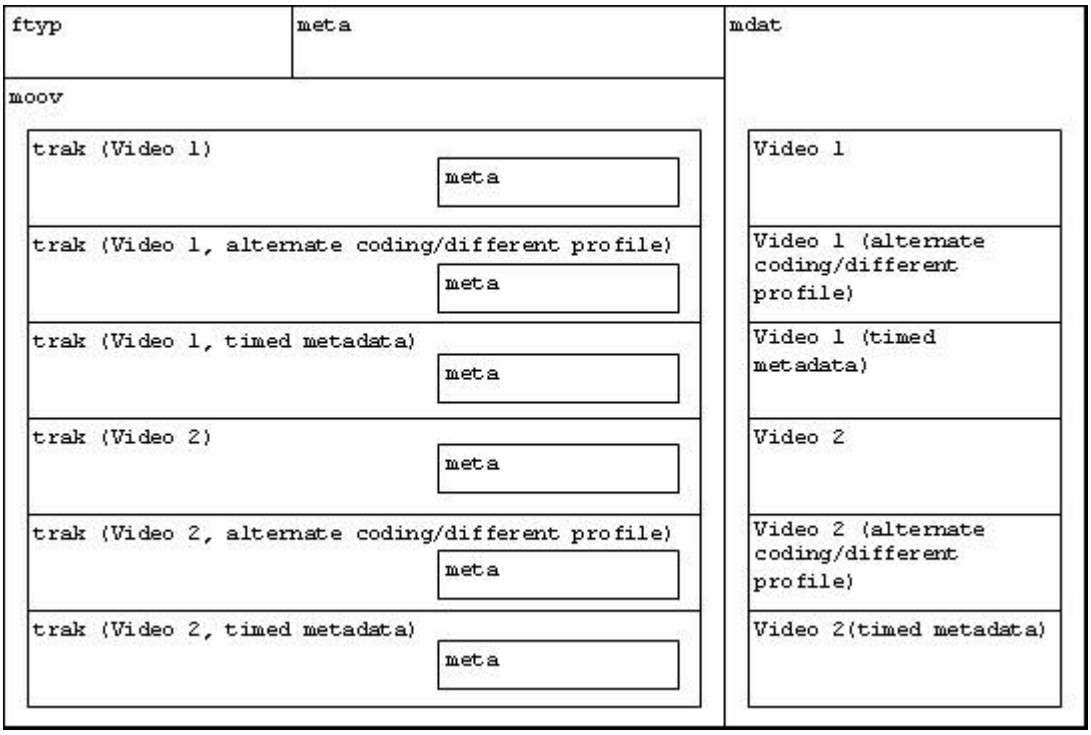

Figure 4 — Example Video surveillance AF fragment illustrating track structure

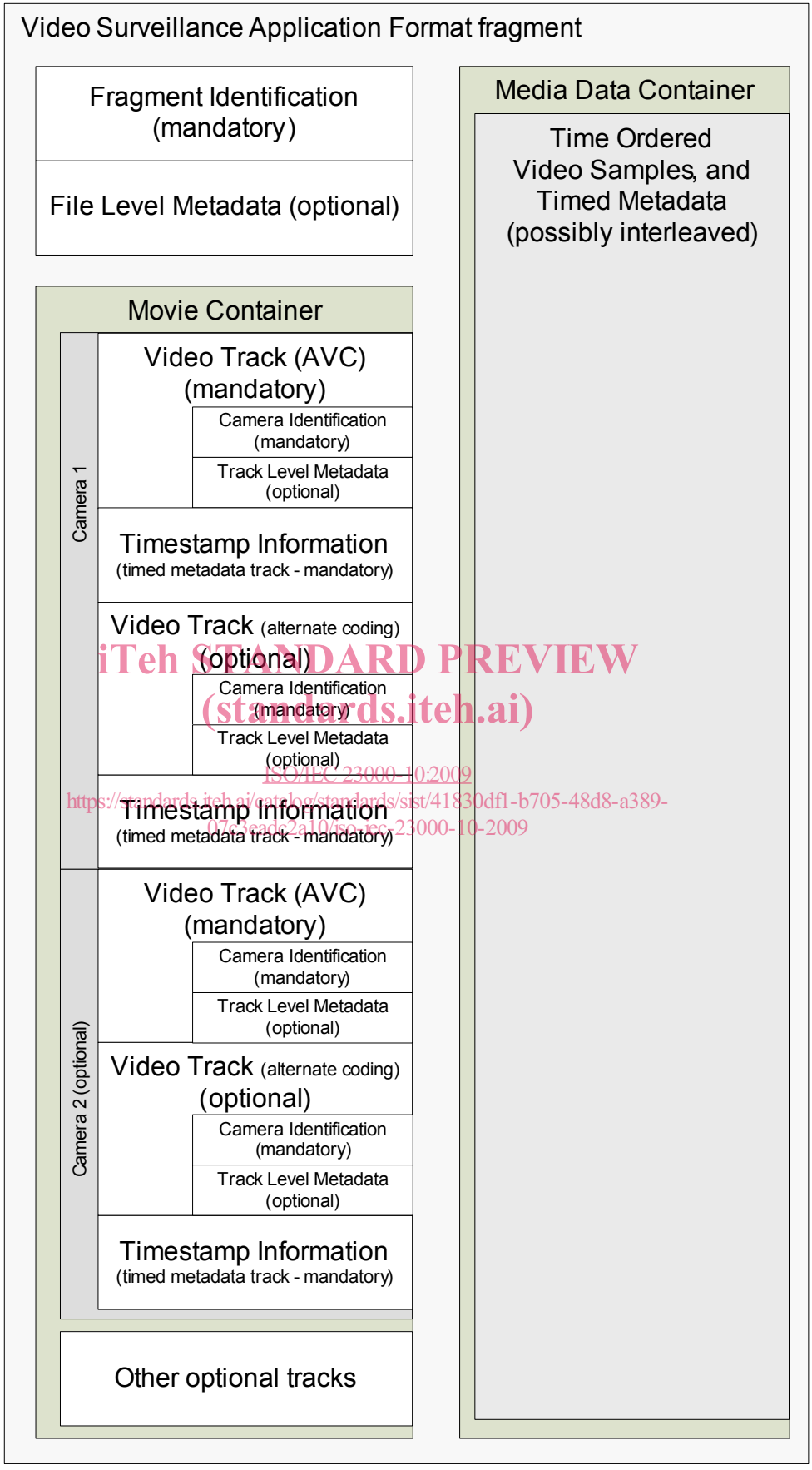

Figure 5 — Example Video surveillance AF fragment illustrating track structure

4.5 Derivation from the ISO Base Media File Format

4.5.1 File Identification

The major_brand identifier for the Video surveillance AF is 'vsf1' (video surveillance format 1). Its meaning is explained herein.

4.5.2 Movie and Track Definition

4.5.2.1 Movie Header Box ('mvhd')

The template fields shall be set to their default values.

The duration shall be set according to the duration of the video tracks in the AF fragment. Note that all video tracks shall have the same duration.

If more than one video track is present e.g. coded with different frame rates, the total duration may differ as indicated by the different frame rates. In this case the track duration shall be set to the greatest value.

4.5.2.2 Track Header Box ('tkhd')

The default value of the track header flags for the video tracks is 7 (track_enabled, track_in_movie, track_in_preview). Width and Height shall correctly document the resolution of a video track. They shall both be set to zero for a metadata track. If an AF fragment contains more than one video track then all video tracks shall be in the same alternate group (see 4.5.3 for detailed description). All other template fields shall be set to their default values.

4.5.2.3 Pixel Aspect Ratio Box ('pasp')

If a pixel aspect ratio different from 1:1 is used for presentation this must be reflected here.

4.5.2.4 Track Reference Box ('tref')

Metadata tracks providing additional timed information (see 6.4 for the metadata track and sample structure) shall be linked to the video tracks they describe by a track reference of type 'cdsc'.

4.5.2.5 Edit Box ('edts')

If edit lists are used for a Video surveillance AF fragment containing more that one video track, a suitable set of edit lists must be provided to ensure synchrony between all video tracks.

4.5.2.6 Media Header Box ('mdhd')

For this AF the timescale shall be set equally to the value used in the movie header box. Creation and modification time shall reflect the time stamps given in the Track Header Box. In particular, if a Track Header Box version 1 is used for a track then a Media Header Box version 1 shall be used. The duration shall be set to the sum of the sample durations (in the scale of the timescale).

4.5.2.7 Handler Reference Box ('hdlr')

This AF specifies the storage of video tracks and additional timed metadata tracks linked to the video tracks therefore handler types 'vide' and 'meta' are required.

The name field of each track should contain a human readable name for the track, e.g. 'camera 1' for the fist camera and 'meta for camera 1' for the metadata track.

4.5.2.8 Media Information Box ('minf')

A Video Media Header shall set all template fields to their default values. The metadata track uses a Null Media Header with flags all set to zero.

4.5.2.9 Data Reference Box ('dref')

Different tracks may use individual physical files or may store interleaved data in the same physical file as indicated by the application.

4.5.2.10 Video Track

The mandatory AVC video track is stored as defined in [see ISO/IEC 14496-15:2004]. The following paragraphs outline the restrictions.

4.5.2.10.1 Elementary Stream Structure

A parameter set elementary stream shall not be used. All parameter sets are stored in the sample description.

4.5.2.10.2 Visual Sample Entry

A visual sample entry of type 'vide' is used to store the video media header which contains an AVC sample entry of type 'avc1'.

MP4 extension descriptors and MP4 bit rate box shall not be used.

Visual width and height must correctly document the size of the video as given with the MPEG-4 AVC parameter sets.

4.5.2.10.3 Sync Samples

All IDR pictures shall be reflected in the sync sample box. A shadow sync sample box shall not be used.

4.5.2.10.4 Layers and Sub-Sequences

Sample groups shall not be used hence layer and sub-sequence definitions shall not be present.

4.5.2.10.5 Sample Dependencies

If a sample dependency box ('sdep') is present (in the sample table or in a track fragment) it shall correctly reflect the dependencies of all samples in the video track. Sample dependency information can be used to enable 'trick modes' such as fast forward/rewind.

4.5.2.11 Sample Groups ('sbgb', 'sgpd')

Sample groups shall not be used for the video and the metadata track.

4.5.2.12 Sample Scale Box ('stsl')

A video track shall reflect the size (width and height) of the visual material i.e. sample scaling information is not needed. Therefore, sample scaling shall not be used.

4.5.2.13 Sub-Sample Information Box ('subs')

Sub-Samples shall not be used hence sub-sample information shall not be present.