INTERNATIONAL STANDARD

ISO/IEC 23000-10

Second edition
2012-12-15

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

Part 10: Surveillance application format

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

Partie 10: Format pour application à la surveillance

Reference number
ISO/IEC 23000-10:2012(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*.

This second edition cancels and replaces the first edition (ISO/IEC 23000-10:2009) and ISO/IEC 23000-10:2009/Amd 1:2010.

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: Surveillance application format*

— *Part 11: Stereoscopic video application format*

— *Part 12: Interactive music application format*

Introduction

This part of ISO/IEC 23000 specifies a file format to enable interoperability for surveillance systems. Features are supported required by wide spread used surveillance systems installations. In the 2nd edition support for audio features has been added, optionally described metadata have been removed and the support for video significantly extended.

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

Part 10:
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 surveillance systems. The file format provides the overall structure for storing media content and associated metadata.

2 Normative references

The following documents, in whole or in part, 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 component

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

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

ISO/IEC 14496-14:2003, *Information technology — Coding of audio-visual objects — Part 14: MP4 file format*

ISO/IEC 14496-15:2010, *Information technology — Coding of audio-visual objects — Part 15: Advanced Video Coding (AVC) file format*

W3C Recommendation, XML Schema Part 2: Datatypes Second Edition, 28 October 2004

3 Overview of Repeatedly Used Standards

3.1 ISO Base Media File Format

The ISO Base Media File Format [see ISO/IEC 14496-12] 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] – uses the techniques from the ISO Base Media File Format. In order to store AVC video together with AAC audio data

[see ISO/IEC 14496-3] the mp4 file format [see ISO/IEC 14496-14] has been derived from ISO Base Media File Format.

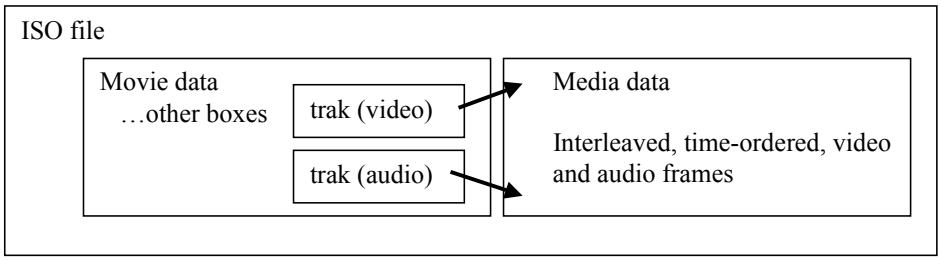

Figure 1 — Example of a simple ISO file used for interchange, containing two streams

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 so called hint tracks.

It also provides support for metadata in the form of 'meta' boxes at the File, Movie and Track level. This allows support for static (un-timed) metadata. Figure 2 schematically illustrates the location of these un-timed 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.

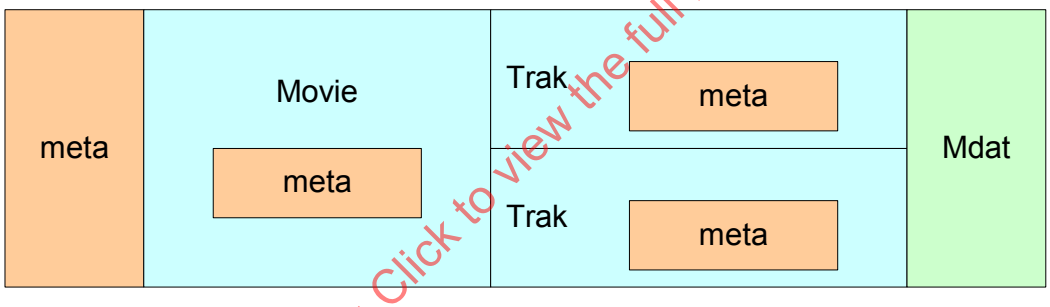

Figure 2 — Support of static un-timed metadata in ISO Base Media Files

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 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 movie fragment consists of zero or one track fragments for each track initialized in the movie header.

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 order to support storage boundary oriented applications a 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 might be 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 Surveillance AF fragment should be created.

If the file is to be read while it is still being written the reader can access all needed information in finalized movie fragments and track fragments. If supported by an application 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 Surveillance AF fragment is being recorded the duration of this 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.

3.2 AVC File Format

In this Surveillance AF support for AVC coded video data is restricted to the main part of ISO/IEC 14496-10, excluding Annex G (Scalable Video Coding) and Annex H (Multiview Video Coding).

4 Using the Surveillance AF

4.1 General

This Clause provides necessary information for creating and using Surveillance AF fragments.

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

4.2 Time reference

Complete UTC timing information shall be provided according to the “dateTime” data type of W3C Recommendation “XML Schema Part2: Datatypes” using the “Z” parameter. The time calculation defined in the ISO base media file format [see ISO/IEC 14496-12] shall be used with 64-bit numbers.

4.3 File Structure

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

All Surveillance AF data is stored within the Surveillance AF fragments. If a fragment has no predecessor or successor its value is set the current fragment. Additionally an URI [see RFC 3986] can be given serving as a hint to the location of the predecessor and successor fragments. A Surveillance AF fragment remains self contained even if unhinged. Note that there is no requirement to use more than one Surveillance AF fragment. The concept of using fragments e.g. enables ring buffer architectures.

Each fragment shall be a valid file as defined by this specification.

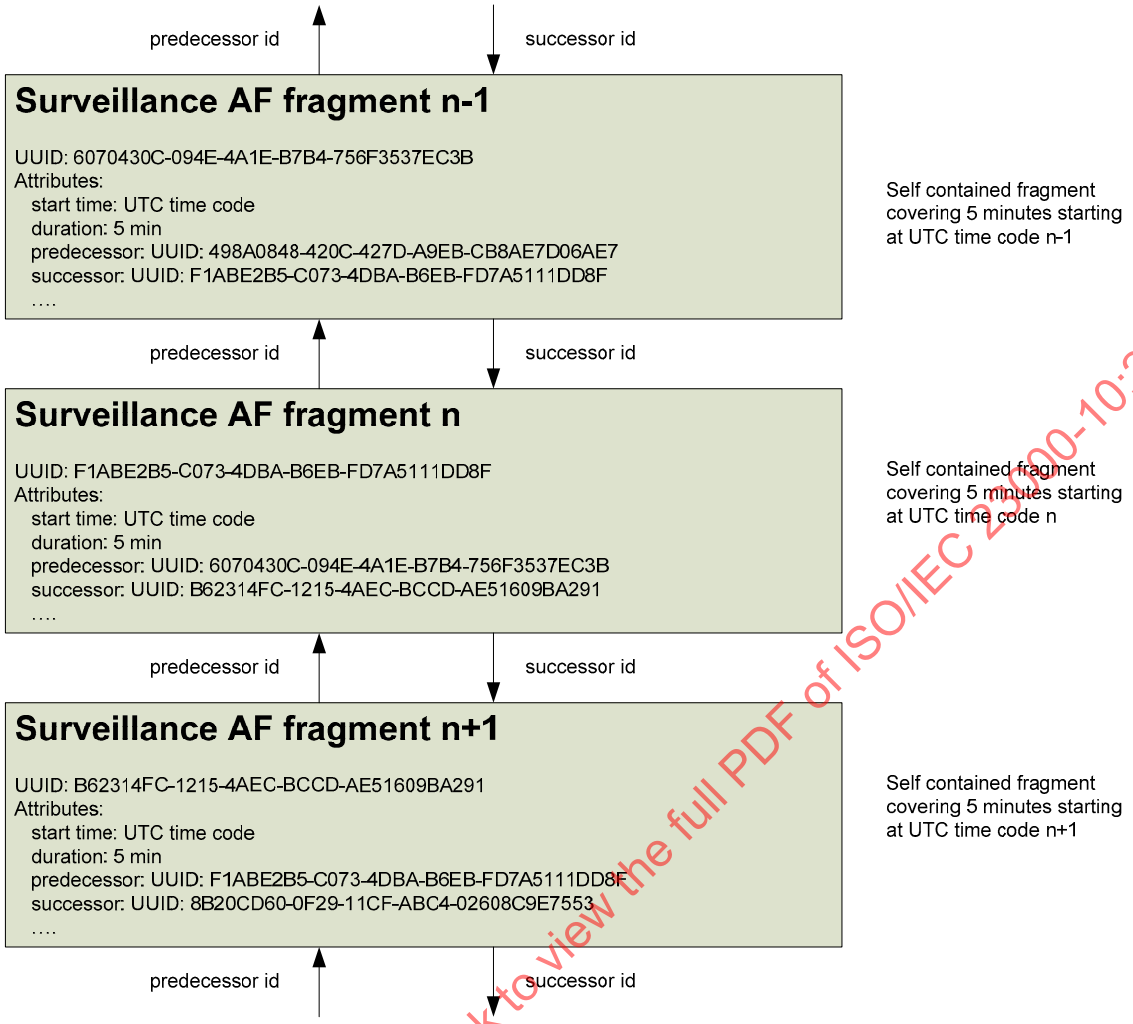

Figure 3 — Surveillance AF fragments linked together by means of predecessor id and successor id

All connected Surveillance AF fragments shall use the same number of tracks and the same set of parameters as timing and media coding settings.

The size of a Surveillance AF fragment can be set as indicated by the application, e.g. providing a constant number of samples in each 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 Surveillance AF fragments and the connection of fragments to the application is out of the scope of this specification.

4.4 File Contents

The file format for the Surveillance AF is based on the ISO Base Media File Format and derived file formats. A Surveillance AF fragment shall contain:

- One box of meta type at file level and one for each media track at track level

If the Surveillance AF fragment contains media data one track of timed metadata shall be provided for each media track, not necessarily contained in the very same Surveillance AF fragment.

4.5 Track Structure

If more than one video track from one camera is present these video tracks shall be in the same alternate group (see 4.6.4 on track selection). Additionally, to each video track a metadata track shall be linked 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. Alternatively different video tracks may contain different content, e.g. different views of the area monitored (see Figure 4 and Figure 5).

The same applies to audio tracks.

If there is more than one media track all media 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.

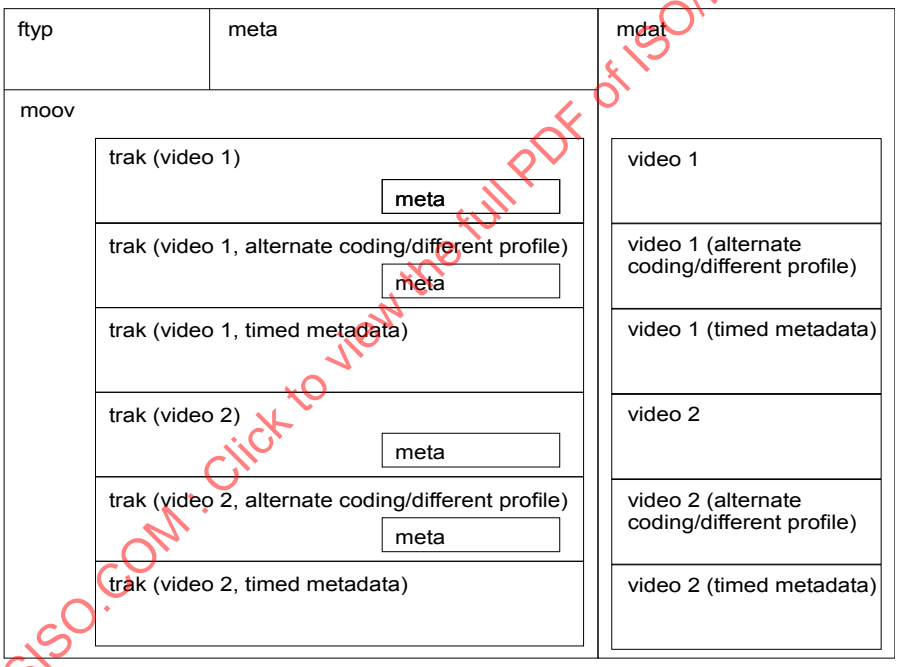

Figure 4 — Example Surveillance AF fragment illustrating track structure

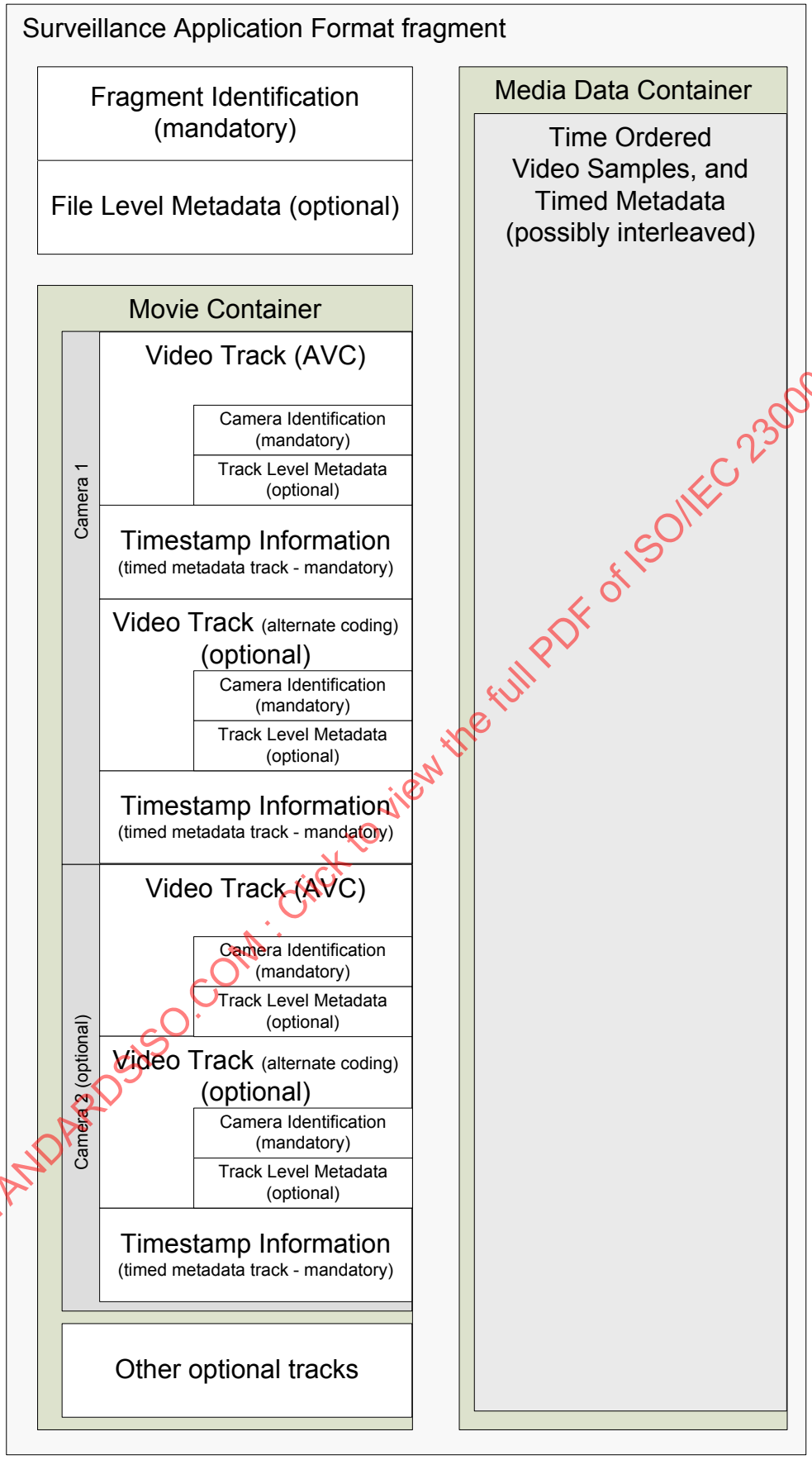

Figure 5 — Example Surveillance AF fragment illustrating track structure

4.6 Derivation from the ISO Base Media File Format

All Surveillance AF fragments shall be self contained. Each fragment shall start with a random access point and contain all necessary information for successfully decoding each sample.

4.6.1 File Identification

The major_brand identifier for the Surveillance AF is 'sff1' (surveillance file format 1).

4.6.2 Movie and Track Definition

4.6.2.1 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 tracks in the AF fragment. Note that all tracks have the same duration.

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

4.6.2.2 Track Header Box ('tkhd')

A Track Header Box version "1" shall be used.

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 audio track then all audio tracks shall be in the same alternate group. If an AF fragment contains more than one video track then all video tracks shall be in the same alternate group (see 4.6.4 for detailed description). All other template fields shall be set to their default values.

4.6.2.3 Pixel Aspect ration Box ('pasp')

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

4.6.2.4 Track Reference Box ('tref')

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

If additional tracks with reference 'hint' are used the entry-format in the sample description shall be 'rrtp' or 'srtp'. Reception hint tracks for RTP and SRTP are thereby described [see ISO/IEC 14496-12].

4.6.2.5 Edit Box ('edts')

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

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

4.6.2.6 Media Header Box ('mdhd')

A Media Header Box version "1" shall be used.

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. The duration shall be set to the sum of the sample durations (in the scale of the timescale).

4.6.2.7 Handler Reference Box ('hdlr')

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

4.6.2.8 Media Information Box ('minf')

A Video Media Header shall set all template fields to their default values. A Sound Media Header shall set all template fields to their default values. Metadata tracks shall use a Null Media Header with flags all set to zero.

4.6.2.9 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.

4.6.2.10 Video Track

The following paragraphs outline the restrictions.

4.6.2.10.1 Layers and Sub-Sequences

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

4.6.2.10.2 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.

4.6.2.10.3 AVC

If AVC video data is present the following restrictions shall be applied.

If a Surveillance AF fragment contains a AVC video track all video data is encapsulated in NAL units and each fragment shall start with an instantaneous decoding refresh access unit as defined in ISO/IEC 14496-10.

4.6.2.10.3.1 SVC and MVC

Annex G (Scalable Video Coding) and Annex H (Multiview Video Coding) of ISO/IEC 14496-10 shall not be used.

4.6.2.10.3.2 Elementary Stream Structure

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

4.6.2.10.3.3 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 AVC parameter sets.

4.6.2.11 Sync Samples

All random access points in a media track shall be reflected in the sync sample box with a sole exception for:

- video tracks: each and every picture represents a random access point
- audio tracks: each and every access unit represents a random access point.

A shadow sync sample box shall not be used.

4.6.2.12 Sample Groups ('sbgp', 'sgpd')

Sample groups shall not be used for media and metadata tracks.

4.6.2.13 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.

4.6.2.14 Sub-Sample Information Box ('subs')

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

4.6.3 Encryption

Sample entry types 'encv', 'enca' and 'enct' are supported. Usage of any of those sample entries requires presence of 'ipro' box. Schema signalling shall be provided in order to identify the protection applied. 'imif' and 'ipmc' boxes shall not be used.

4.6.4 Alternate Track Identification

The Surveillance AF supports the storage of tracks in alternate groups to indicate the dependencies between tracks in a Surveillance AF fragment. Switch groups are not supported by the Surveillance AF.

A Surveillance AF fragment may include a Track Selection Box in the user data box of the track it describes. A Track Selection Box shall be present for each track in an alternate group and contain attributes describing the alternative. Each attribute implicitly points to a descriptive structure in the file format.

The following attributes defined in the ISO Base Media File Format shall be used to describe the alternatives:

Name	*Attribute*	*Pointer*
Codec	'cdec'	Sample Entry (in Sample Description box of media track)
Media type	'mtyp'	Handlertype in Handler box (of media track)
Screen size	'scsz'	Width and height fields of Visual Sample Entries.
Bitrate	'bitr'	Total size of the samples in the track divided by the duration in the track header box
Frame rate	'frar'	Number of samples in the track divided by duration in the track header box

In addition the following attributes specific to the Surveillance AF are used:

Name	*Attribute*	*Pointer*
Camera	‘cami’	Camera/Microphone Identification Box
Meta	‘meta’	Track level metadata (XML or binary XML)

4.6.5 Movie Fragments

Refer to ISO/IEC 14496-12:2008, 8.8.4 and 8.8.6 for a detailed description of movie and track fragments.

In general, movie and track fragments extend a presentation in time. All fragments must be stored in sequence given by an ordinal sequence number.

Each movie fragment contains one or more track fragments for all tracks in the movie.

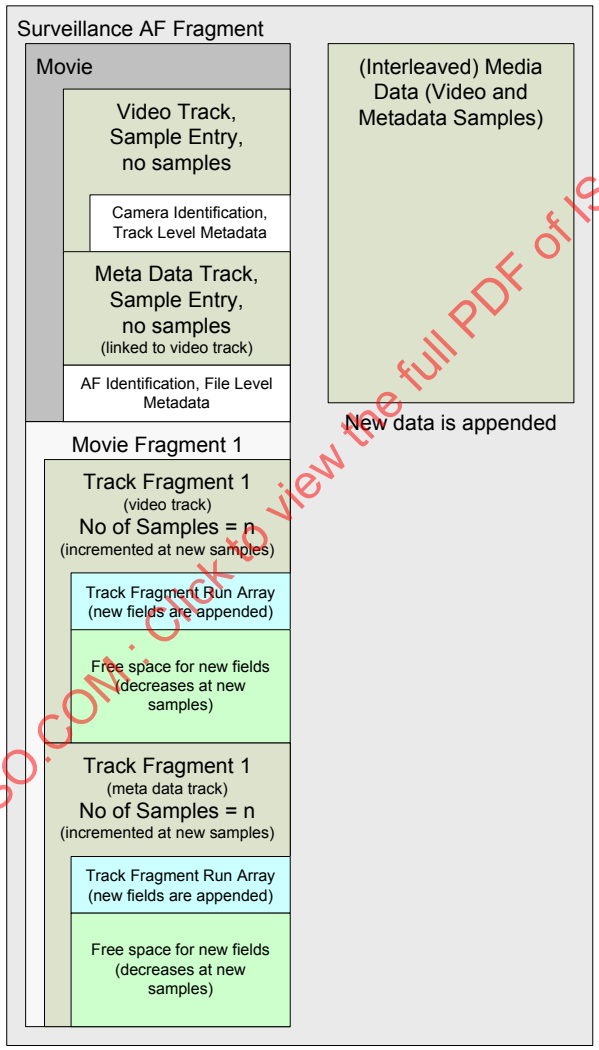

Figure 6 — Example illustrating the usage of video fragments

4.6.6 Metadata Track

Time parallel metadata is used to describe each video sample and audio access unit individually. All samples in the metadata track shall have the same decoding and composition time as the sample they describe.

The metadata track is linked to by the described video track, which contains a Track Reference of type '*vsmd*'. The metadata track may be linked to by more than one video track.

The metadata track is linked to by the described audio track, which contains a Track Reference of type '*asmd*'. The metadata track may be linked to by more than one audio track.

Additional time parallel metadata tracks can be linked to video and audio tracks to provide additional information.

5 Derived file formats

If video and audio data are jointly present in a Surveillance AF the definitions in ISO/IEC 14496-14:2003 shall be observed in addition. BIFS, IOD and OD tracks shall no be used.

If AVC video is present as media data in a Surveillance AF the definitions in ISO/IEC 14496-15:2010 shall be observed in addition.

6 Metadata

6.1 Introduction

In the Surveillance AF, metadata is used to provide additional information about the content. The developers may utilize this information to implement extra functionality in their device applications.

6.2 File Level Metadata

The AF Identification Box is required, and shall be included in every Surveillance AF fragment. An additional Meta Box containing further information may also be included in a Surveillance AF fragment.

6.2.1 AF Identification Box

The AF identifier box covers the following information:

- File identification: An UUID identifying every Surveillance AF fragment
- Successor and predecessor identification: The UUID of the previous/next fragment in composition time shall be included (URIs describing the corresponding location may be included)
- The UTC based time stamp of the first sample in the media tracks and the duration information for the fragment.

6.2.1.1 Definition

Box Type: 'sumi'
Container: Meta Box ('meta'), file level
Mandatory: Yes
Quantity: Exactly one

This box shall provide the Surveillance AF fragment identification UUID and the UUID of the successor and predecessor. It may also provide the URI to the successor and predecessor.