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 iv

Introduction v

1 **Scope** 1

2 **Normative references** 1

3 **Overview of MPEG Standards Used** 1
3.1 **MPEG-4 Advanced Video Coding** 1
3.2 **ISO Base Media File Format** 2
3.3 **MPEG-7 Multimedia Description Scheme** 4
3.4 **MPEG-7 Visual** 4
3.5 **AVC File Format** 4

4 **Using the Video surveillance AF** 4
4.1 **General** 4
4.2 **File Structure** 4
4.3 **File Contents** 6
4.4 **Track Structure** 6
4.5 **Derivation from the ISO Base Media File Format** 8

5 **Video Coding Definition** 11
5.1 **Introduction** 11
5.2 **AVC Profile and Level** 12

6 **Metadata** 12
6.1 **Introduction** 12
6.2 **File Level Metadata** 12
6.3 **Track Level Metadata** 14
6.4 **Timed Metadata** 15

Annex A (informative) **Use cases of Video surveillance AF** 16

Annex B (normative) **Metadata Specification** 17

Bibliography 51

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.

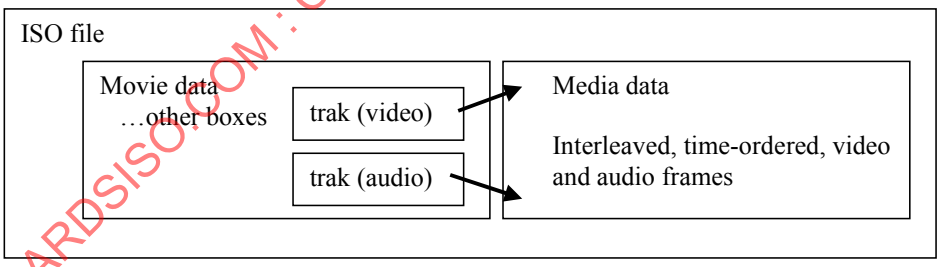

Figure 1 — Example of a simple ISO file used for interchange, containing two streams

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.

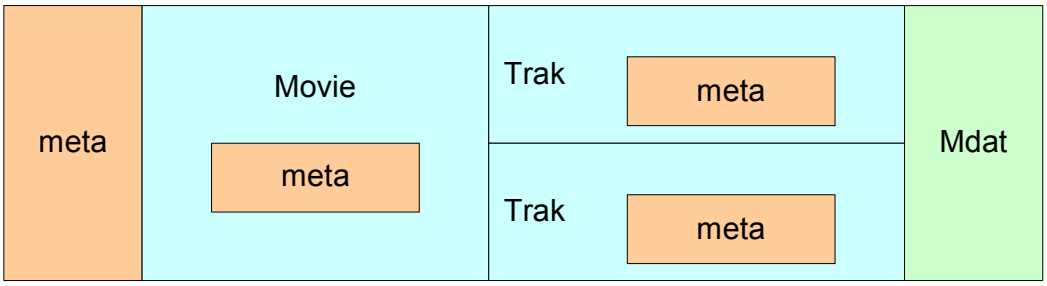

Figure 2 — Support of Static untimed Metadata in ISO/MP4 Files

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.

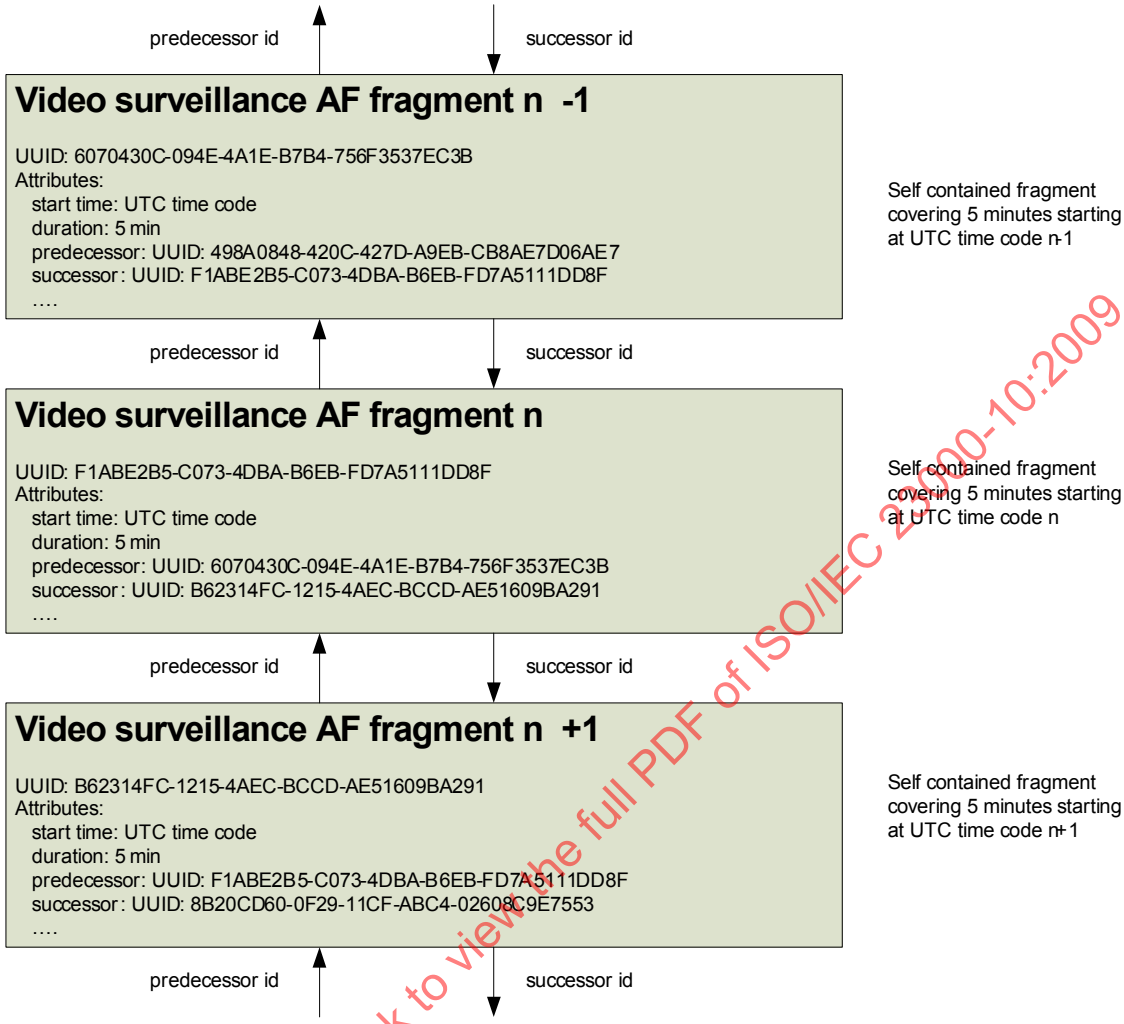

Figure 3 — VSAF fragments linked together by means of predecessor id and successor id

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.

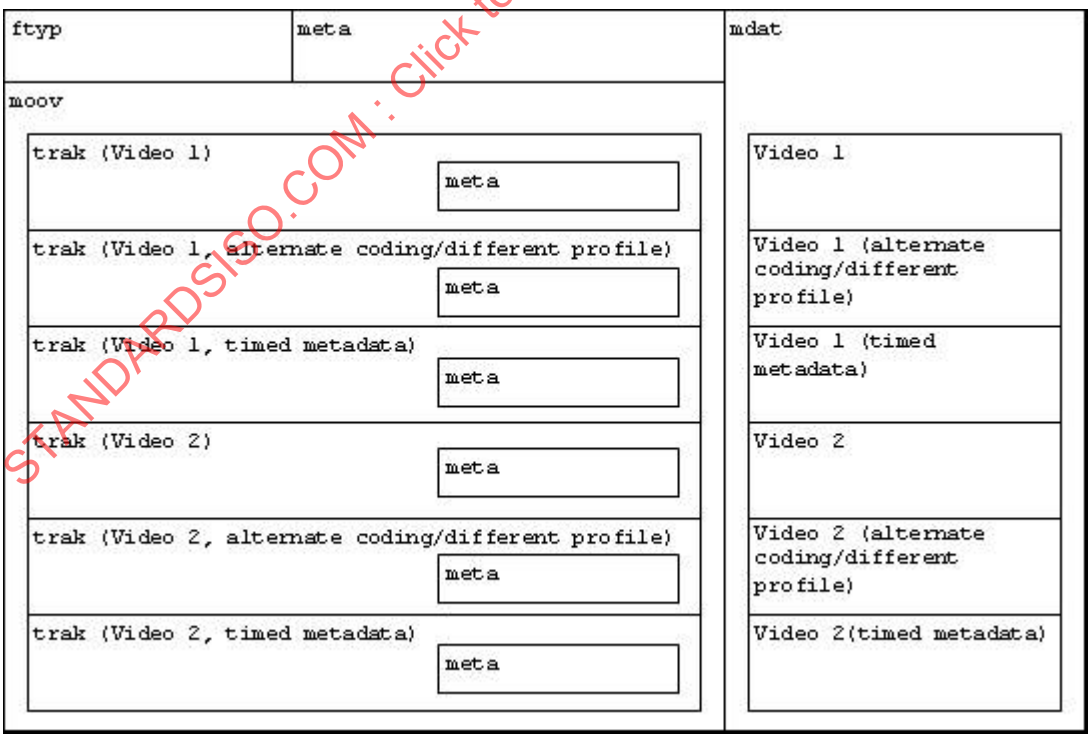

Figure 4 — Example Video surveillance AF fragment illustrating track structure

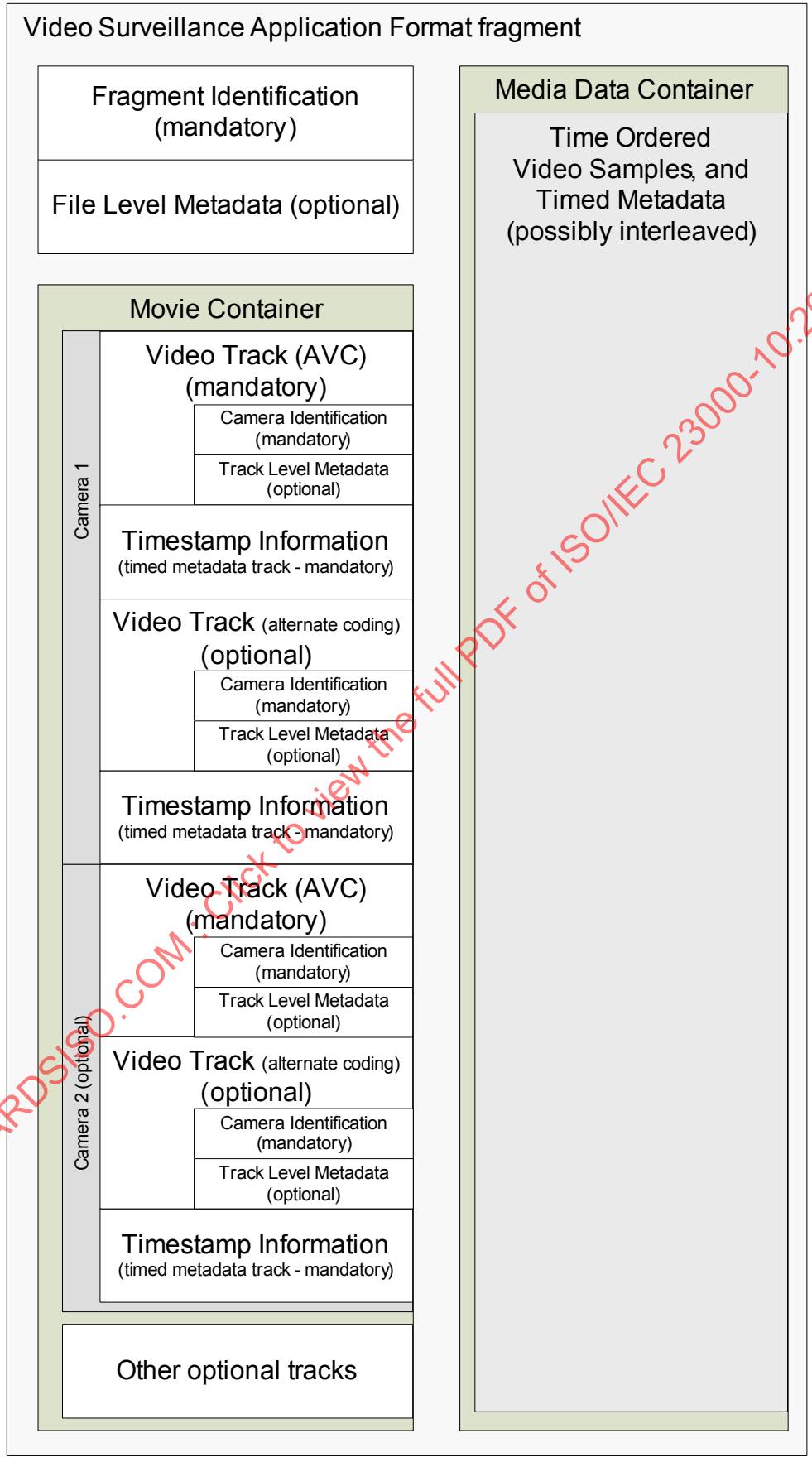

Figure 5 — Example Video surveillance AF fragment illustrating track structure

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.

4.5.2.14 Encryption (Protected Sample Entries, ‘ipro’, ‘sinf’, ’frma’, ’imif’, ‘ipmc’, ‘schi’)

A Video surveillance AF fragment shall not contain protected sample entries (e.g. ‘encv’) or Item Protection Boxes (‘ipro’).

4.5.3 Alternate Track Identification

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

A Video 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 video 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)
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 Video surveillance AF can be used:

Name	*Attribute*	*Pointer*
Camera	‘cami’	Camera Identification Box
Scene	‘scen’	Additional description in Camera Identification Box
Meta	‘meta’	Track level metadata (XML or binary XML)

4.5.4 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.

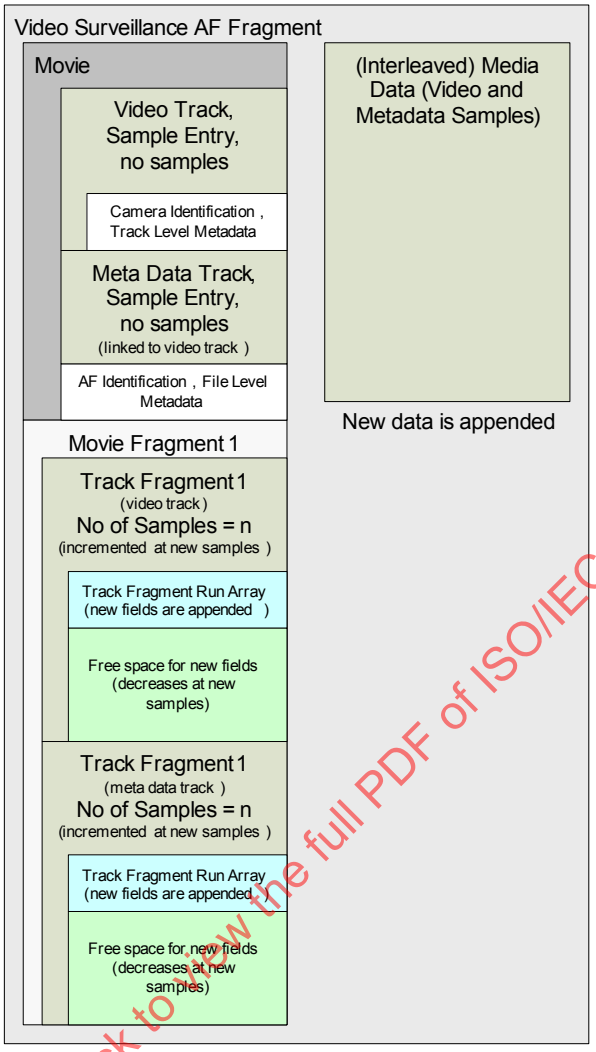

Figure 6 — Example illustrating the usage of video fragments

4.5.5 Metadata Track

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

Additional time parallel metadata tracks can be linked to a video track to provide additional information.

5 Video Coding Definition

5.1 Introduction

An AF fragment consists of at least one MPEG-4 AVC video track conveying header information following the defined parameter set structure. All video data is encapsulated in NAL units and each fragment shall start with an instantaneous decoding refresh access unit.

5.2 AVC Profile and Level

Because of multi-functionality of MPEG-4 AVC, subsets of different tools have been defined in order to allow effective implementations of the standard. These subsets, called "Profiles", limit the tool set which shall be implemented. For each of these Profiles one or more Levels have been set to restrict the computational complexity of implementations.

MPEG-4 AVC accepts various sizes of input picture within the capability specified from the Profile and Level.

In this AF usage of the MPEG-4 AVC video codec is required. The Baseline Profile tool set will be used up to level 3.1 (maximum value of level_idc shall be 31).

Both, constraint_set0_flag and constraint_set1_flag shall be set to 1 simultaneously.

6 Metadata

6.1 Introduction

In the Video 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

On file level two different boxes may be used to store metadata. The AF Identification Box is required, and shall be included in every Video surveillance AF fragment. An additional Meta Box containing further information may also be included in a Video surveillance AF fragment. If present, metadata described in Annex B shall be contained inside this additional Meta Box.

6.2.1 AF Identification Box

The AF identifier box covers the following information:

- File identification: An UUID identifying every Video 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 video tracks and the duration information for the fragment.

6.2.1.1 Definition

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

This box shall provide the Video surveillance AF fragment identification UUID and the UUID of the successor and predecessor. It may also provide the URI to the successor and predecessor. If these URIs are provided there must be the possibility to resolve the URIs.

6.2.1.2 Syntax

```
class AFIdentificationBox extends Box('vsmi')
{
    UInt(128) fragmentUUID;
    UInt(128) predecessorUUID;
    UInt(128) successorUUID;
    UInt(64) startTime;
    UInt(64) duration;
    UInt(16) predecessorURIsize;
    UInt(16) successorURIsize;
    UInt(8)[predecessorURIsize] predecessorURI;
    UInt(8)[predecessorURIsize] successorURI;
}
```

6.2.1.3 Semantics

`fragmentUUID` the UUID (16 Byte) identifying this Video surveillance AF fragment.

`predecessorUUID` the UUID (16 Byte) identifying this Video surveillance AF fragment immediately preceding this Video surveillance AF fragment in composition time.

If this Video surveillance AF fragment has no predecessor, `predecessorUUID` holds the `fragmentUUID` of this fragment.

`successorUUID` the UUID (16 Byte) identifying this Video surveillance AF fragment immediately following this Video surveillance AF fragment in composition time.

If this Video surveillance AF fragment has no successor, `predecessorUUID` holds the `fragmentUUID` of this fragment.

`startTime` the UTC based time represented by the number of 100-nanosecond intervals since January 1, 1601 of the first video sample in this fragment.

`duration` the UTC based time represented by the number of 100-nanosecond intervals since January 1, 1601 of the difference between the last video sample and the first video sample in this fragment.

`predecessorURIsize` the size of `predecessorURI` in bytes including the terminating NULL character. If `predecessorURI` is not applicable, `predecessorURIsize` is set to 0.

`successorURIsize` the size of `successorURI` in bytes including the terminating NULL character. If `successorURI` is not applicable, `successorURIsize` is set to 0.

`predecessorURI` the URI providing the location of the predecessor fragment as NULL terminated string encoded in UTF-8 characters.

`successorURI` the URI providing the location of the successor fragment as NULL terminated string encoded in UTF-8 characters.

6.2.2 Meta Box

This box may be included: if so, it would contain additional time information and may contain annotations that apply to the collection of video tracks, and classification schemas that are used in the track boxes.

The primary data in this Meta Box shall be stored in either XML format or BinaryXML format according to ISO/IEC 23001-1:2006.

A detailed description can be found in Annex B. This provides a list of MPEG-7 elements allowed in the Meta Box, to be stored as primary data. It would be possible to combine these elements with the track level Meta Box contents, to produce a valid MPEG-7 document that conforms to a restriction of the MPEG-7 schema.

6.3 Track Level Metadata

Track level metadata is included in two boxes. A required camera identifier box shall be included and an additional Meta Box may be included. If present, metadata described in Annex B shall be contained inside this additional Meta Box.

6.3.1 Camera Identification Box

This contains the camera identification:

- An UUID identifying the camera
- Additional space for user defined identification

6.3.1.1 Definition

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

Provides the camera identification UUID and user defined identification extensions used to create a particular track in a Video surveillance AF. The camera UUID should be assigned with a physical camera or with a camera location. The camera identification box may be enlarged (indicated by the size of the box) if storage of user defined identification data is required. This extra information might not be understood by all Video surveillance AF readers.

If there are alternative tracks holding different encodings from the same camera then the camera UUID shall be identical for all these alternate tracks.

6.3.1.2 Syntax

```
class CameraIdentificationBox extends Box('cami')
{
    UInt(128) cameraUUID;
}
```

6.3.1.3 Semantics

`cameraUUID` the UUID (first 16 Byte) identifying the camera or the camera location used to create this video track

6.3.2 Meta Box

This box may be included: if so, it should contain the description of cameras and their settings as well as a description of video content.

The primary data in this Meta Box shall be stored in either XML format or BinaryXML format according to ISO/IEC 23001-1:2006.

A detailed description can be found in Annex B. This provides a list of MPEG-7 elements allowed in the Meta Box, to be stored as primary data. It would be possible to combine these elements with the file level Meta Box contents, to produce a valid MPEG-7 document that conforms to a restriction of the MPEG-7 schema.

6.4 Timed Metadata

The Video surveillance AF requires the capture time stamp to be stored for every video frame. The timestamps are stored in timestamp metadata samples in a time parallel metadata track which is linked to the video track by means of a track reference with type '*vsmd*'. For all video samples a timestamp metadata sample shall exist with decoding time equal to the decoding time of the corresponding video sample.

6.4.1 Video surveillance AF Sample Timestamp Metadata Sample Entry and Sample Format

The Video surveillance AF defines the storage of a binary coded timestamp for all video samples of a video track. However future version of the Video surveillance AF might store more information about a video sample. The video surveillance metadata sample entry contains a version number to inform the reader of the sample format used in this metadata track. The Video surveillance AF uses version 1 timestamp metadata sample format. A reader should check the version number to enable future extensions.

6.4.1.1 Definition

The Video surveillance AF Sample Timestamp Metadata Sample Entry extends the Metadata Sample Entry and includes a configuration box.

6.4.1.2 Syntax

```
class VideoSurveillanceMetadataSampleConfigBox extends FullBox('vcmC')
{
   UInt(8) version;
}

class VideoSurveillanceMetadataSampleEntry () extends MetadataSampleEntry('vcmM')
{
   VideoSurveillanceMetadataSampleConfigBox config;
}

class TimestampMetadataSample {
   UInt(64) timestamp;
}
```

6.4.1.3 Semantics

`version` the version of the timestamp metadata track sample format. The Video surveillance AF sets version to 1.

`timestamp` the UTC based time represented by the number of 100-nanosecond intervals since January 1, 1601 when the associated video sample was captured.

Annex A
(informative)

Use cases of Video surveillance AF

In many countries, notably the USA, U.K. and other European and East Asian countries, video surveillance in public places is increasingly used for crime prevention and for the detection of similar incidents. Examples for public places being observed are streets, squares, railway and subway stations.

More and more cameras are being in place forming huge surveillance systems. Within those systems required basic functionalities are identical. The video stream needs to be transmitted from the site to an appropriate place where it will be archived. The video might be looked at by a number of persons and in case of an incident it could be exported to the appropriate authorities.

In order to identify a requested stream it would be necessary to enhance the pure video stream by appropriate metadata. Here, the information about recording time and place as well as camera parameters used for recording would be sufficient to achieve basic interoperability.

For efficient archival the packaging of the video and metadata information into a file format must be supported. That file format can also provide for the inclusion of user data and possibly additional MPEG-7 metadata. This metadata can provide key functionality to support the activities of the CCTV manufacturers, installers and users.

Recording of audio signals in the context of surveillance is not commonly used, and is therefore not included in this application scenario. Companies in this space have indicated their interest to include such functionality in the future.

Annex B
(normative)

Metadata Specification

B.1 Introduction

This Annex contains the metadata specification for the additional Meta Boxes, on file and track level, which may be included in a Video surveillance AF fragment. (This is separate to the specification for the required Meta Boxes at file and Track level, which are to be found in 6.2.1 and 6.3.1 respectively).

B.2 Metadata Definition

The following table summarizes the elements of MPEG-7 schema that are conformance to the requirements of Video surveillance AF. The interpretation is as follows:

- "element/attribute/attributeGroup" – this MPEG-7 metadata shall be instantiated in the metadata of a Video Surveillance compliant file
- xsi:type="[TypeName]" – the element shall have this attribute, when instantiated in the item-level metadata. Therefore, it shall only be instantiated with type [TypeName])
- minOccurs="*n*" – at least *n* occurences of the element shall be instantiated in the item-level metadata
- maxOccurs="*m*" – no more than *m* occurences of the element shall be instantiated in the item-level metadata
- Elements are referenced using MPEG-7 `id` attributes

Textual description of metadata:

1. General text annotation
   a. Textual annotation should be added using either free text or structured annotation. All elements within structured annotation are limited to zero or one.
2. For file level metadata:
   a. ID
      i. The UUID of the VSAF file shall be repeated in the `PublicIdentifier` of the `DescriptionMetadata` element. There shall be one of these descriptors.
      ii. It is assumed the UUID of the camera already contains the information of the cluster the camera belongs to.
   b. Time
      i. The UTC based time stamp of the first sample in the video tracks should be repeated at file level (`CreationTime` of the `DescriptionMetadata` element).

c. Textual annotations should be added using `Comment` of the `DescriptionMetadata` element. Zero or one of the types described in 1.a. shall be used.

d. Classification schemes and `Terms` can be defined and used as described in ISO/IEC 15938-5:2003 section 7.4. The cardinality of the `Definition` element of the `TermDefinitionBaseType` shall be zero or one. The cardinality of the `Name` element of the `TermDefinitionBaseType` shall be zero or one. There shall be zero of the `preferred` attribute of the `Name` element of the `TermDefinitionBaseType`.

e. Maintaining object references

i. Object references are grouped using the Graph DS and referenced using Relation DS elements. The objects can be any DS, as used within the XML document, *e.g.* a still region of video at track level.

ii. The attributes within a `Relation` are restricted to contain unary values. E.g. `source`, `target` and `type` can only contain a single reference to a Classification Scheme `Term`, `id` reference, *etc. i.e.* the values of the `termReferenceType`.

1. General text annotation

The `TextAnnotationType` shall have zero of the `confidence` and `relevance` attributes.

The term `TermUseType` shall be restricted to contain:

- zero of the `TermUseType` elements (no recursion)
- zero or one of the `Name` element
- zero or one of the `Definition` element

Description	Elements of MPEG-7 Schema	Constraints
General text annotation		
1.a Free Text	FreeTextAnnotation	minOccurs = “0”
1.b Structured Annotation	StructuredAnnotation/Who	minOccurs = “0”
	StructuredAnnotation/WhatObject	minOccurs = “0”
	StructuredAnnotation/WhatAction	minOccurs = “0”
	StructuredAnnotation/When	minOccurs = “0”
	StructuredAnnotation/Where	minOccurs = “0”
	StructuredAnnotation/Why	minOccurs = “0”
	StructuredAnnotation/How	minOccurs = “0”

2. File-level table

Description	Elements of MPEG-7 Schema	Constraints
Information associated with captured data		
2.a ID	Mpeg7/DescriptionMetadata/PublicIdentifier	
2.b Time information	Mpeg7/DescriptionMetadata/CreationTime	
Metadata for content		
2.c Annotations that apply to file-level		
i. In a free text format	Mpeg7/DescriptionMetadata/Comment/FreeTextAnnotation	Choice minOccurs = “0” maxOccurs = “unbounded”
ii. In a semantically structured format	Mpeg7/DescriptionMetadata/Comment/StructuredAnnotation	
2.d Classification schemes	Mpeg7/Description[@xsi:type=”ClassificationSchemeDescriptionType”]/ClassificationScheme	minOccurs = “0” maxOccurs = “unbounded”
2.e The identity of objects observed in one or more video sources		
i. Object reference descriptions	Mpeg7/Description[@xsi:type=”ContentEntityType”]/Relationships	minOccurs = “0” maxOccurs = “unbounded”
	Mpeg7/Description[@xsi:type=”ContentEntityType”]/Relationships/Relation	minOccurs = “0” maxOccurs = “unbounded”
	Mpeg7/Description[@xsi:type=”ContentEntityType”]/Relationships/Relation[source,target,type]	minOccurs = “0”
ii. Restricting the cardinality of the Relation attributes	Mpeg7/Description[@xsi:type=”ContentEntityType”]/Relationships/Relation[source=”xxx”],target,type]	(xxx) maxOccurs = “1”
	Mpeg7/Description[@xsi:type=”ContentEntityType”]/Relationships/Relation[target=”xxx”]	
	Mpeg7/Description[@xsi:type=”ContentEntityType”]/Relationships/Relation[type=”xxx”]	

3. For each track:

   a. Metadata for each track are described using the Video Segment DS. *E.g.* `VideoType`. Only one of these types shall be used.

   b. ID.

      i. The camera id shall be repeated from the 'cami' box in `PublicIdentifier` of the `DescriptionMetadata` element. There shall be one of these descriptors.

   c. Equipment

      i. The camera / cluster settings should be given (`Instrument` of the `DescriptionMetadata` element). If present, zero or one of these types shall be used.

      ii. Additional information regarding the cluster to which the camera belongs to should be given by `EntityIdentifier` and its `VideoDomain` element. If present, one of the `EntityIdentifier` types shall be used and zero or more of the `VideoDomain` types should be used. The `VideoDomain` elements reference entries from a Classification Scheme (`ClassificationScheme`).

      iii. The camera stream should be identified with `StreamID`. Zero or more of these types shall be used. It is necessary to include the element `InstanceIdentifier`, although this can be kept empty.

      iv. The camera geographic position should be given using `CreationLocation`. Zero or one of these types shall be used.

      v. Camera 2-dimensional (2D) projection should be provided with the `Spatial2DCoordinateSystemType`. More than one of these descriptors allows a 2D projection function for each preset position for PTZ cameras. Zero or one of these types shall be used

      vi. If the media is outside of the VSAF fragment and referenced using the Data Reference Box (dref) the `MediaURI` shall contain the same reference. If no Data Reference Box (dref) is present the `MediaURI` should contain a valid reference to a media instance. It is necessary to include the element `InstanceIdentifier`, although this can be kept empty.

   d. Time

      i. Video offset has no specific element, so the Description Metadata DS (`DescriptionMetadata`), `Instrument` and its `Tool Setting` elements should be used. The setting name is "`offset`". If present, one of these types shall have the format and precision as given in Section 6.

      ii. The UTC based time stamp of the first sample in the video track (`CreationTime` of the `DescriptionMetadata` element).

      iii. To isolate where the `StillRegion` exists in the video, the `MediaTimePoint` shall be used.

      Note: Duration information for a specific track could be calculated by using the first and the last time stamp of the timed metadata.

   e. Decomposition

      i. Groups of frames should be defined within the video using the `TemporalDecomposition`. If present, one of these types shall be used

ii. Single frames should be decomposed using the StillRegion DS. If a frame (StillRegion DS) is decomposed its time position shall be specified.

iii. To isolate a region within a frame, a choice shall be made between a `Box` or `Polygon`. Zero or one of these can be described per `StillRegion`. If more regions are required for a frame, then another `StillRegion` can be instantiated, referencing the same media time point (`mediaTimePoint`).

f. Visual Descriptions

i. Colour should be described in the StillRegion DS by the VisualDescriptor DS or the GridLayout DS. The VisualDescriptor shall include one colour descriptor. The `GridLayout` can specify an arbitrary number of cells, each should contain one colour descriptor.

ii. `DominantColor` and `ScalableColor` shall be the only descriptors present from the VisualDescriptor DS and GridLayout DS.

g. Semantic descriptions

i. A camera track should define semantic descriptions using `TextAnnotation` – see 1."General text annotation". This is possible with `FreeTextAnnotation` at `DescriptionMetadata`, `Video` and `StillRegion` levels.

ii. A camera track should define semantic descriptions using `TextAnnotation` – see 1."General text annotation". This is possible with a structured annotation using the `StructuredAnnotation` at `DescriptionMetadata`, `Video` and `StillRegion` levels.

iii. In order to provide detailed meaning to semantic descriptions, `Terms` should be referenced from Classification schemes (`ClassificationScheme`) – see 2.d.

h. Maintaining object references

i. Object references are defined as described in 2.e.

Track-level table

Description	Elements of MPEG-7 Schema	Constraints
Information associated with captured data		
3.a Video	Mpeg7/Description[@xsi:type="ContentEntityType"]/Multimedia Content[@xsi:type="VideoType"].	
3.b Identification tag for the camera	Mpeg7/DescriptionMetadata/PublicIdentifier	
3.c Description of equipment used and equipment settings		
i. Camera settings (Aperture, shutter-speed values, peak-to-peak voltage, *etc.*)	Mpeg7/DescriptionMetadata/Instrument	minOccurs = “0”
	Mpeg7/DescriptionMetadata/Instrument/Settings	minOccurs = “0” maxOccurs = “unbounded”
ii. Additional information regarding the cluster	Mpeg7/Description[@xsi:type=”ContentEntityType”]/Multimedia Content[@xsi:type="VideoType"]/MediaInformation/MediaIdentification/EntityIdentifier	
	Mpeg7/Description[@xsi:type=”ContentEntityType”]/Multimedia Content[@xsi:type="VideoType"]/MediaInformation/MediaIdentification/EntityIdentifier/../VideoDomain	minOccurs = “0” maxOccurs = “unbounded”
iii. Identification tags for each of the multiple streams from a single camera.	Mpeg7/Description[@xsi:type=”ContentEntityType”]/Multimedia Content[@xsi:type="VideoType"]/MediaInformation/MediaProfile/MediaInstance/MediaLocator/StreamID	minOccurs = “0”
	Mpeg7/Description[@xsi:type=”ContentEntityType”]/Multimedia Content[@xsi:type="VideoType"]/MediaInformation/MediaProfile/MediaInstance/InstanceIdentifier	
iv. camera geographic position	Mpeg7/DescriptionMetadata/CreationLocation/GeographicPosition	minOccurs = “0”
v. Camera 2D projection	Mpeg7/Description[@xsi:type=”ContentEntityType”]/Multimedia Content[@xsi:type="VideoType"]/Header[@xsi:Spatial2DCoordinatesType]	minOccurs = “0”
vi. Media URI	Mpeg7/Description[@xsi:type=”ContentEntityType”]/Multimedia Content[@xsi:type="VideoType"]/MediaInformation/MediaProfile/MediaInstance/ MediaLocator/MediaURI	minOccurs = “0”
	Mpeg7/Description[@xsi:type=”ContentEntityType”]/Multimedia Content[@xsi:type="VideoType"]/MediaInformation/MediaProfile/MediaInstance/InstanceIdentifier	
3.d Timing Information		
i. Video time offset	Mpeg7/DescriptionMetadata/Instrument/Tool/Setting[Name, Value]	minOccurs = “0”
ii. Time of the video	Mpeg7/DescriptionMetadata/CreationTime	
iii. Time of a frame	Mpeg7/Description[@xsi:type=”ContentEntityType”]/Multimedia Content[@xsi:type="VideoType"]/Video/TemporalDecomposition/StillRegion/MediaTimePoint	

Track-level table (*continued*)

Description	Elements of MPEG-7 Schema	Constraints
3.e Decomposition		
i. Groups of frames	Mpeg7/Description[@xsi:type=”ContentEntityType”]/MultimediaContent[@xsi:type="VideoType"]/Video/TemporalDecomposition	minOccurs = “0”
ii. Single frames	Mpeg7/Description[@xsi:type=”ContentEntityType”]/MultimediaContent[@xsi:type="VideoType"]/Video/TemporalDecomposition/StillRegion	minOccurs = “0” maxOccurs = ”unbounded”
iii. Isolate a region	Mpeg7/Description[@xsi:type=”ContentEntityType”]/MultimediaContent[@xsi:type="VideoType"]/Video/TemporalDecomposition/StillRegion/SpatialLocator/Box	choice minOccurs = “0”
	Mpeg7/Description[@xsi:type=”ContentEntityType”]/MultimediaContent[@xsi:type="VideoType"]/Video/TemporalDecomposition/StillRegion/SpatialLocator/Polygon	choice minOccurs = “0”
3.f. Visual Descriptors		
i. Colour can be described in a frame using the StillRegion or the GridLayout	Mpeg7/Description[@xsi:type=”ContentEntityType”]/MultimediaContent[@xsi:type="VideoType"]/Video/TemporalDecomposition/StillRegion/VisualDescriptor	Choice minOccurs = “0”
	Mpeg7/Description[@xsi:type=”ContentEntityType”]/MultimediaContent[@xsi:type="VideoType"]/Video/TemporalDecomposition/StillRegion/GridLayoutDescriptors/Descriptor	
ii. ScalableColor and DominantColor	Mpeg7/Description[@xsi:type=”ContentEntityType”]/MultimediaContent[@xsi:type="VideoType"]/Video/TemporalDecomposition/StillRegion/VisualDescriptor[@xsi:type=“ScalableColorType”] Mpeg7/Description[@xsi:type=”ContentEntityType”]/MultimediaContent[@xsi:type="VideoType"]/Video/TemporalDecomposition/StillRegion/GridLayoutDescriptors/Descriptor[@xsi:type=“ScalableColorType”]	
	Mpeg7/Description[@xsi:type=”ContentEntityType”]/MultimediaContent[@xsi:type="VideoType"]/Video/TemporalDecomposition/StillRegion/VisualDescriptor[@xsi:type=“DominantColorType”] Mpeg7/Description[@xsi:type=”ContentEntityType”]/MultimediaContent[@xsi:type="VideoType"]/Video/TemporalDecomposition/StillRegion/GridLayoutDescriptors/Descriptor[@xsi:type=“DominantColorType”]	
3.g. Semantic descriptions		
i. In a free text format	Mpeg7/DescriptionMetadata/Comment/FreeTextAnnotation	Choice minOccurs = “0” maxOccurs = “unbounded”
	Mpeg7/Description[@xsi:type=”ContentEntityType”]/MultimediaContent[@xsi:type="VideoType"]/TextAnnotation/FreeTextAnnotation	
	Mpeg7/Description[@xsi:type=”ContentEntityType”]/MultimediaContent[@xsi:type="VideoType"]/Video/TemporalDecomposition/StillRegion/TextAnnotation/FreeTextAnnotation	

Track-level table (*continued*)

Description	Elements of MPEG-7 Schema	Constraints
ii. In a semantically structured format	Mpeg7/Description[@xsi:type=”ContentEntityType”]/MultimediaContent[@xsi:type="VideoType"]/TextAnnotation/StructuredAnnotation	Choice minOccurs = “0” maxOccurs = “unbounded”
	Mpeg7/DescriptionMetadata/Comment/StructuredAnnotation	
	Mpeg7/Description[@xsi:type=”ContentEntityType”]/MultimediaContent[@xsi:type="VideoType"]/Video/TemporalDecomposition/StillRegion/TextAnnotation/StructuredAnnotation	
iii. Referencing Classification Schemes	Mpeg7/Description[@xsi:type=”ContentEntityType”]/MultimediaContent[@xsi:type="VideoType"]/Video/TemporalDecomposition/StillRegion/TextAnnotation/StructuredAnnotation/Who/Term[termID] Mpeg7/Description[@xsi:type=”ContentEntityType”]/Relationships/Relation[type,target,source]	minOccurs = “0”
3.h. Maintaining object references		
	Mpeg7/Description[@xsi:type=”ContentEntityType”]/Relationships	minOccurs = “0” maxOccurs = “unbounded”
	Mpeg7/Description[@xsi:type=”ContentEntityType”]/Relationships/Relation	minOccurs = “0” maxOccurs = “unbounded”

File level

```
<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
<Mpeg7 xmlns="urn:mpeg:mpeg7:schema:2001"
xmlns:mpeg7="urn:mpeg:mpeg7:schema:2001"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="urn:mpeg:mpeg7:schema:2001 Mpeg7-2001_vsaf-07-2008.xsd
">
    <DescriptionMetadata>
        <Comment>
            <FreeTextAnnotation>free text annotation</FreeTextAnnotation>
            <StructuredAnnotation>
                <Who>
                    <Name>HOW</Name>
                </Who>
                <WhatObject>
                    <Name>WHATOBJECT</Name>
                </WhatObject>
                <WhatAction>
                    <Name>WHATACTION</Name>
                </WhatAction>
                <Where>
                    <Name>WHERE</Name>
                </Where>
                <When>
                    <Name>WHEN</Name>
                </When>
                <Why>
                    <Name>WHY</Name>
                </Why>
                <How>
                    <Name>HOW</Name>
                </How>
            </StructuredAnnotation>
        </Comment>
        <PublicIdentifier>6E3564F8-8CCB-47af-9F7E-
FD4693AA9D4B</PublicIdentifier>
        <CreationLocation>
            <GeographicPosition>
                <Point latitude="0.100000" longitude="0.100000" />
            </GeographicPosition>
        </CreationLocation>
        <CreationTime>2008-01-01T01:01:01:0F30</CreationTime>
        <Instrument>
            <Tool>
                <Name>Automatically generated by VSAF Meta-data API</Name>
            </Tool>
        </Instrument>
    </DescriptionMetadata>
    <Description xsi:type="ClassificationSchemeDescriptionType">
        <!-- The relationship between Joe Doe (on patrol) and Fred Bloggs
(working on the computer) and -->
        <!-- a moving object detected by the surveillance camera 8A4EB5D3-B2D8-
4b99-A430-11032F0AAAF4 -->
        <!-- These values will be based upon the output of all cameras and time
etc. -->
        <mpeg7:Relationships id="rship1">
            <mpeg7:Relation id="r1"
                source="
urn:mpeg:maf:cs:vsaf:example2:2008:vs:infrastructure:camera:ptz:camera1"
                type=" urn:mpeg:maf:cs:vsaf:example2:2008:process:object:unknown1"
```

```
                target="
urn:mpeg:maf:cs:vsaf:example1:2008:acme:people:security:guard:joe_doe"
                strength="0.8"/>
            <mpeg7:Relation id="r2"
                source="
urn:mpeg:maf:cs:vsaf:example2:2008:vs:infrastructure:camera:ptz:camera1"
                type=" urn:mpeg:maf:cs:vsaf:example2:2008:process:object:unknown1"
                target="
urn:mpeg:maf:cs:vsaf:example1:2008:acme:people:staff:fred_bloggs"
                strength="0.2"/>
        </mpeg7:Relationships>
        <ClassificationScheme uri="urn:mpeg:maf:cs:vsaf:example1:2008">
            <Term id="t1" termID="acme">
                <Name>ACME</Name>
                <Definition>The ACME company classification scheme</Definition>
                <Term id="t2" termID="people">
                    <Name>People</Name>
                    <Definition>People recorded by the ACME
surveillance</Definition>
                    <Term id="t3" termID="security">
                        <Name>Security</Name>
                        <Definition>Security are people who are work for the ACME
surveillance system</Definition>
                        <Term id="t4" termID="guard">
                            <Name>Guard</Name>
                            <Definition>Security guards to protect the premise and
patrol</Definition>
                            <Term id="t5" termID="joe_doe">
                                <Name>Joe Doe</Name>
                                <Definition>Joe Doe, the security guard</Definition>
                            </Term>
                        </Term>
                    </Term>
                    <Term id="t6" termID="staff">
                        <Name>Staff</Name>
                        <Definition>Staff are the employees who work for the ACME
company</Definition>
                        <Term id="t7" termID="fred_bloggs">
                            <Name>Fred Bloggs</Name>
                            <Definition>Fred Bloggs, the company
employee</Definition>
                        </Term>
                    </Term>
                </Term>
            </Term>
        </ClassificationScheme>
        <ClassificationScheme uri="urn:mpeg:maf:cs:vsaf:example2:2008">
            <Term id="t8" termID="vs">
                <Name>vs</Name>
                <Definition>The surveillance system</Definition>
                <Term id="t9" termID="process">
                    <Name>Process</Name>
                    <Definition>The definition of terms used in the computer
processing components</Definition>
                    <Term id="t10" termID="object">
                        <Name>Object</Name>
                        <Definition>An detected object</Definition>
                        <Term id="t11" termID="blob">
                            <Name>Blob</Name>
                            <Definition>The moving region within a
video</Definition>
```

```
                    </Term>
                    <Term id="t12" termID="known">
                        <Name>Known</Name>
                        <Definition>The moving region within a video with a
known identity</Definition>
                    </Term>
                    <Term id="t13" termID="unknown">
                        <Name>Unknown</Name>
                        <Definition>The moving region within a video with a
unknown identity</Definition>
                        <Term id="t14" termID="known1">
                            <Name>Unknown1</Name>
                            <Definition>The identity of an object of unknown
identity</Definition>
                        </Term>
                    </Term>
                </Term>
            </Term>
            <Term id="t15" termID="infrastructure">
                <Name>Infrastructure</Name>
                <Definition>The surveillance system's
infrastructure</Definition>
                <Term id="t16" termID="camera">
                    <Name>Camera</Name>
                    <Definition>Cameras watching and recording</Definition>
                    <Term id="t17" termID="PTZ">
                        <Name>PTZ</Name>
                        <Definition>A camera with pan-tilt-zoom
mobility</Definition>
                        <Term id="t18" termID="Carmera1">
                            <Name>8A4EB5D3-B2D8-4b99-A430-11032F0AAAF4</Name>
                            <Definition>The camera overlooking
entrance</Definition>
                        </Term>
                    </Term>
                </Term>
                <Term id="t19" termID="cluster">
                    <Name>Cluster</Name>
                    <Definition>Camera cluster</Definition>
                    <Term id="t20" termID="cluster1">
                        <Name>73D1C6F8-5405-4a82-B2AF-295FDD3B5D30</Name>
                        <Definition>The cluster of cameras surveying all
entrance points</Definition>
                    </Term>
                </Term>
            </Term>
        </Term>
    </ClassificationScheme>
  </Description>
</Mpeg7>
```

Track level

```
<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
<Mpeg7 xmlns="urn:mpeg:mpeg7:schema:2001"
xmlns:mpeg7="urn:mpeg:mpeg7:schema:2001"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="urn:mpeg:mpeg7:schema:2001 Mpeg7-2001_vsaf-07-2008.xsd">
    <DescriptionMetadata>
        <Comment>
            <FreeTextAnnotation>Sensor details</FreeTextAnnotation>
            <StructuredAnnotation>
                <Who href="
urn:mpeg:maf:cs:vsaf:example1:2008:acme:people:security:guard:joe_doe"/>
                <WhatObject>
                    <Name>WHATOBJECT</Name>
                </WhatObject>
                <WhatAction>
                    <Name>WHATACTION</Name>
                </WhatAction>
                <Where>
                    <Name>WHERE</Name>
                </Where>
                <When>
                    <Name>WHEN</Name>
                </When>
                <Why>
                    <Name>WHY</Name>
                </Why>
                <How>
                    <Name>HOW</Name>
                </How>
            </StructuredAnnotation>
            <FreeTextAnnotation>More free Text</FreeTextAnnotation>
        </Comment>
        <PublicIdentifier>8A4EB5D3-B2D8-4b99-A430-
11032F0AAAF4</PublicIdentifier>
        <CreationLocation>
            <GeographicPosition>
                <Point latitude="0.100000" longitude="0.100000" />
            </GeographicPosition>
        </CreationLocation>
        <CreationTime>2008-01-01T01:01:01:0F25</CreationTime>
        <Instrument>
            <Tool>
                <Name>Automatically generated by VSAF Meta-data API</Name>
            </Tool>
            <Setting name="Focus" value="1" />
            <Setting name="Offset" value="25" />
        </Instrument>
    </DescriptionMetadata>
    <Description xsi:type="ContentEntityType">
        <!-- The relationship between the movement detected by the surveillance
camera 8A4EB5D3-B2D8-4b99-A430-11032F0AAAF4 -->
        <!-- and the known people -->
        <!-- this will be based upon appearance -->
        <Relationships id="rship1">
            <Relation id="r1" source="blob1"
                target="
urn:mpeg:maf:cs:vsaf:example2:2008:process:object:unknown1"
                type="
urn:mpeg:maf:cs:vsaf:example2:2008:vs:process:object:unknown"
```

```
                strength="0.8"/>
        </Relationships>
        <Relationships id="rship2">
            <Relation id="r2" source="blob2"
            target=" urn:mpeg:maf:cs:vsaf:example2:2008:process:object:unknown1"
                type="
urn:mpeg:maf:cs:vsaf:example2:2008:vs:process:object:unknown"
                strength="0.9"/>
        </Relationships>
        <Relationships id="rship3">
            <Relation id="r3" source="blob3"
            target=" urn:mpeg:maf:cs:vsaf:example2:2008:process:object:unknown1"
                type="
urn:mpeg:maf:cs:vsaf:example2:2008:vs:process:object:unknown"
                strength="0.9"/>
        </Relationships>
        <MultimediaContent xsi:type="VideoType">
            <Header id="coord1" xRepr="0"
xsi:type="Spatial2DCoordinateSystemType" yRepr="0">
                <LocalCoordinateSystem name="preset1">
                    <Pixel>0 0</Pixel>
                    <CoordPoint>0.100000 0.100000</CoordPoint>
                    <Pixel>0 0</Pixel>
                    <CoordPoint>0.100000 0.100000</CoordPoint>
                    <Pixel>0 0</Pixel>
                    <CoordPoint>0.100000 0.100000</CoordPoint>
                    <MappingFunct>ProjectionMatrixMappingFunction</MappingFunct>
                </LocalCoordinateSystem>
            </Header>
            <Header id="coord2" xRepr="0"
xsi:type="Spatial2DCoordinateSystemType" yRepr="0">
                <LocalCoordinateSystem name="preset2">
                    <Pixel>0 0</Pixel>
                    <CoordPoint>0.100000 0.100000</CoordPoint>
                    <Pixel>0 0</Pixel>
                    <CoordPoint>0.100000 0.100000</CoordPoint>
                    <Pixel>0 0</Pixel>
                    <CoordPoint>0.100000 0.100000</CoordPoint>
                    <MappingFunct>ProjectionMatrixMappingFunction</MappingFunct>
                </LocalCoordinateSystem>
            </Header>
            <Video>
                <MediaInformation>
                    <!-- Camera cluster information -->
                    <MediaIdentification>
                        <EntityIdentifier>73D1C6F8-5405-4a82-B2AF-
295FDD3B5D30</EntityIdentifier>
                        <VideoDomain href="
urn:mpeg:maf:cs:vsaf:example2:2008:vs:infrastructure:camera:ptz:camera1" />
                        <VideoDomain href="
urn:mpeg:maf:cs:vsaf:example2:2008:vs:infrastructure:cluster:cluster1" />
                    </MediaIdentification>
                    <MediaProfile>
                        <MediaInstance>
                            <InstanceIdentifier/>
                            <MediaLocator>
                                <MediaUri>file:///data/video.avi</MediaUri>
                                <StreamID>1</StreamID>
                            </MediaLocator>
                        </MediaInstance>
                    </MediaProfile>
```

```
            </MediaInformation>
            <TextAnnotation>
                <FreeTextAnnotation>Free text annotation</FreeTextAnnotation>
                <StructuredAnnotation>
                    <Who>
                        <Name>HOW</Name>
                    </Who>
                    <WhatObject>
                        <Name>WHATOBJECT</Name>
                    </WhatObject>
                    <WhatAction>
                        <Name>WHATACTION</Name>
                    </WhatAction>
                    <Where>
                        <Name>WHERE</Name>
                    </Where>
                    <When>
                        <Name>WHEN</Name>
                    </When>
                    <Why>
                        <Name>WHY</Name>
                    </Why>
                    <How>
                        <Name>HOW</Name>
                    </How>
                </StructuredAnnotation>
                <FreeTextAnnotation>Another Free text
annotation</FreeTextAnnotation>
            </TextAnnotation>
            <TemporalDecomposition>
                <StillRegion id="blob1">
                    <TextAnnotation>
                        <FreeTextAnnotation>Detected blob -- probably
human</FreeTextAnnotation>
                        <StructuredAnnotation>
                            <Who>
                                <Name>HOW</Name>
                            </Who>
                            <WhatObject>
                                <Name>WHATOBJECT</Name>
                            </WhatObject>
                            <WhatAction>
                                <Name>WHATACTION</Name>
                            </WhatAction>
                            <Where>
                                <Name>WHERE</Name>
                            </Where>
                            <When>
                                <Name>WHEN</Name>
                            </When>
                            <Why>
                                <Name>WHY</Name>
                            </Why>
                            <How>
                                <Name>HOW</Name>
                            </How>
                        </StructuredAnnotation>
                        <FreeTextAnnotation>Another Free text
annotation</FreeTextAnnotation>
                    </TextAnnotation>
                    <SpatialLocator>
```

```
                    <Box xmlns:mpeg7="urn:mpeg:mpeg7:schema:2001"
mpeg7:dim="2 2">1 2 3 4</Box>
                </SpatialLocator>
                <MediaTimePoint>2008-01-01T01:01:01:0F25</MediaTimePoint>
                <VisualDescriptor xsi:type="DominantColorType">
                    <SpatialCoherency>0</SpatialCoherency>
                    <Value>
                        <Percentage>1</Percentage>
                        <Index>1 1 1</Index>
                    </Value>
                </VisualDescriptor>
            </StillRegion>
            <StillRegion id="blob2">
                <TextAnnotation>
                    <FreeTextAnnotation>Detected blob -- probably
human</FreeTextAnnotation>
                    <StructuredAnnotation>
                        <Who>
                            <Name>HOW</Name>
                        </Who>
                        <WhatObject>
                            <Name>WHATOBJECT</Name>
                        </WhatObject>
                        <WhatAction>
                            <Name>WHATACTION</Name>
                        </WhatAction>
                        <Where>
                            <Name>WHERE</Name>
                        </Where>
                        <When>
                            <Name>WHEN</Name>
                        </When>
                        <Why>
                            <Name>WHY</Name>
                        </Why>
                        <How>
                            <Name>HOW</Name>
                        </How>
                    </StructuredAnnotation>
                    <FreeTextAnnotation>Another Free text
annotation</FreeTextAnnotation>
                </TextAnnotation>
                <SpatialLocator>
                    <Polygon>
                        <Coords xmlns:mpeg7="urn:mpeg:mpeg7:schema:2001"
mpeg7:dim="2 2">5 6 7 8</Coords>
                    </Polygon>
                </SpatialLocator>
                <MediaTimePoint>2008-01-01T01:01:01:0F25</MediaTimePoint>
                <VisualDescriptor numOfBitplanesDiscarded="0"
numOfCoeff="256" xsi:type="ScalableColorType">
                    <Coeff>
                        1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
                        1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
                        1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
                        1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
                        1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

```
                        </Coeff>
                    </VisualDescriptor>
                </StillRegion>
                <StillRegion id="blob3">
                    <TextAnnotation>
                        <StructuredAnnotation>
                            <Who>
                                <Name>HOW</Name>
                            </Who>
                            <WhatObject>
                                <Name>WHATOBJECT</Name>
                            </WhatObject>
                            <WhatAction>
                                <Name>WHATACTION</Name>
                            </WhatAction>
                            <Where>
                                <Name>WHERE</Name>
                            </Where>
                            <When>
                                <Name>WHEN</Name>
                            </When>
                            <Why>
                                <Name>WHY</Name>
                            </Why>
                            <How>
                                <Name>HOW</Name>
                            </How>
                        </StructuredAnnotation>
                        <StructuredAnnotation>
                            <Who>
                                <Name>HOW</Name>
                            </Who>
                            <WhatObject>
                                <Name>WHATOBJECT</Name>
                            </WhatObject>
                            <WhatAction>
                                <Name>WHATACTION</Name>
                            </WhatAction>
                            <Where>
                                <Name>WHERE</Name>
                            </Where>
                            <When>
                                <Name>WHEN</Name>
                            </When>
                            <Why>
                                <Name>WHY</Name>
                            </Why>
                            <How>
                                <Name>HOW</Name>
                            </How>
                        </StructuredAnnotation>
                    </TextAnnotation>
                    <MediaTimePoint>2008-01-01T01:01:01:0F25</MediaTimePoint>
                    <GridLayoutDescriptors
descriptorMask="11000000000000000000000000" numOfPartX="5" numOfPartY="5">
                        <Descriptor xsi:type="DominantColorType">
                            <SpatialCoherency>0</SpatialCoherency>
                            <Value>
                                <Percentage>1</Percentage>
                                <Index>1 1 1</Index>
                            </Value>
```

```
                       </Descriptor>
                       <Descriptor numOfBitplanesDiscarded="0"
numOfCoeff="256" xsi:type="ScalableColorType">
                           <Coeff>
                              1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
                              1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
                              1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
                              1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
                              1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1
                           </Coeff>
                       </Descriptor>
                    </GridLayoutDescriptors>
                 </StillRegion>
             </TemporalDecomposition>
          </Video>
       </MultimediaContent>
   </Description>
</Mpeg7>
```

Schema

```
<?xml version="1.0" encoding="UTF-8"?>
<!--                                          ****************************************
****************************************

   This XML document was originally developed in the course of development of
the
   ISO/IEC 15938 standard (MPEG-7). This XML document contains either a part of
   the MPEG-7 schema implementation for one or more MPEG-7 tools as specified
by
   the MPEG-7 Requirements or MPEG-7 description examples conformant to the
   MPEG-7 schema.

   ISO/IEC gives users of MPEG-7 free license to this XML document or
modifications
   Thereof for use in hardware or software products claiming conformance to
MPEG-7.

   Those intending to use this XML document in hardware or software products
are
   advised that its use may infringe existing patents. The original developers
of
   this XML document and his/her company, the subsequent editors and their
companies,
   and ISO/IEC have no liability for use of this XML document or modifications
   thereof in an implementation.

   Copyright is not released for non MPEG-7 conforming products. The
organizations
   Who contributed to this XML document retain the full right to use the code
for
   their own purpose, assign or donate their contribution to a third party and
   inhibit third parties from using their contribution for non MPEG-7
conforming
   products.

   Copyright (c) 1999-2003 ISO/IEC.

   This XML document is provided for informative purposes only. If any parts of
this
   XML document contradict the normative part of the corresponding standard
document
   then the normative part should be used as the definitive specification.

   This notice must be included in all copies or derivative works.
   ************************************
**************************************** -->
<!--                                          ****************************************
************************************
   Modified as part of the development of ISO/IEC 23000-10 Video Surveillance
Application Format
   Authors:
   James Annesley, Kingston University, London, United Kingdom
   Houari Sabirin, Information and Communications University, Daejeon, Korea
   last update: James Annesley 22.07.08
   ************************************
**************************************** -->
<!-- mpeg7-2001_vsaf-07-2008.xsd -->

<schema                                   xmlns="http://www.w3.org/2001/XMLSchema"
```

```
xmlns:mpeg7="urn:mpeg:mpeg7:schema:2001"
targetNamespace="urn:mpeg:mpeg7:schema:2001"
    elementFormDefault="qualified" attributeFormDefault="unqualified">
    <annotation>
        <documentation>
            The following schema definition corresponds to a subset of the
description tools of:
            ISO/IEC 15938-2 (DDL)
            ISO/IEC 15938-3 (Visual)
            ISO/IEC 15938-5 (MDS)
        </documentation>
        <documentation>
            The following schema definition corresponds to the description tools
of: ISO/IEC 23000-10 (MPEG-A: Part 10 - Video Surveillance
            Application Format)
        </documentation>
    </annotation>
    <!-- ############################################################ -->
    <!--  import xml components                                       -->
    <!-- ############################################################ -->
    <import namespace="http://www.w3.org/XML/1998/namespace"
schemaLocation="http://www.w3.org/2001/03/xml.xsd" />
    <annotation>
        <documentation>This document contains tools defined as MPEG-7 specific
extension of XML Schema in ISO/IEC 15938-2</documentation>
    </annotation>
    <!--
*****************************************************************************-->
    <!--    ****************************    ISO/IEC    15938-2    DDL
**************************-->
    <!--
*****************************************************************************-->
    <annotation>
        <documentation>The following section contains tools defined in ISO/IEC
15938-2</documentation>
    </annotation>
    <!-- listOfPositiveIntegerForDim -->
    <simpleType name="listOfPositiveIntegerForDim">
        <list itemType="positiveInteger" />
    </simpleType>
    <!-- dim -->
    <attribute name="dim">
        <simpleType>
            <restriction base="mpeg7:listOfPositiveIntegerForDim">
                <minLength value="1" />
            </restriction>
        </simpleType>
    </attribute>
    <!-- basicTimePointType -->
    <simpleType name="basicTimePointType">
        <restriction base="string">
            <pattern value="\-?(\d+(\-\d{2}(\-
\d{2})?)?)?(T\d{2}(:\d{2}(:\d{2}(:\d+(\.\d{2})?)?)?)?)?(F\d+)?((\-
|\+)\d{2}:\d{2})?" />
        </restriction>
    </simpleType>
    <!--
***************************************************************************-->
    <!--    ********************    ISO/IEC    15938-3    Visual
******************************-->
    <!--
```

```
**************************************************************************-->
    <annotation>
       <documentation>The following section contains visual tools defined in:
ISO/IEC 15938-3 (Visual)</documentation>
    </annotation>
    <!-- ########################################################### -->
    <!-- Definition of GridLayout Datatype                           -->
    <!-- ########################################################### -->
    <complexType name="GridLayoutType" final="#all">
       <sequence>
          <element          name="Descriptor"          type="mpeg7:VisualDType"
maxOccurs="unbounded" />
       </sequence>
       <attribute name="numOfPartX" type="mpeg7:unsigned8" use="required" />
       <attribute name="numOfPartY" type="mpeg7:unsigned8" use="required" />
       <attribute name="descriptorMask" use="optional">
          <simpleType>
             <restriction base="string">
                <pattern value="(0|1)*" />
             </restriction>
          </simpleType>
       </attribute>
    </complexType>
    <!-- ########################################################### -->
    <!-- Definition of Spatial2DCoordinateSystem Header              -->
    <!-- ########################################################### -->
    <complexType name="Spatial2DCoordinateSystemType" final="#all">
       <complexContent>
          <extension base="mpeg7:HeaderType">
             <sequence>
                <element name="LocalCoordinateSystem" minOccurs="0">
                   <complexType>
                      <sequence>
                         <sequence maxOccurs="3">
                            <element name="Pixel">
                               <simpleType>
                                  <restriction base="mpeg7:integerVector">
                                     <length value="2" />
                                  </restriction>
                               </simpleType>
                            </element>
                            <element name="CoordPoint">
                               <simpleType>
                                  <restriction base="mpeg7:floatVector">
                                     <length value="2" />
                                  </restriction>
                               </simpleType>
                            </element>
                         </sequence>
                         <element name="MappingFunct" type="string" />
                      </sequence>
                      <attribute name="name" type="string" use="required" />
                   </complexType>
                </element>
             </sequence>
             <attribute name="xRepr" type="mpeg7:unsigned8" use="required" />
             <attribute name="yRepr" type="mpeg7:unsigned8" use="required" />
          </extension>
       </complexContent>
    </complexType>
    <!-- ########################################################### -->
```

```
    <!-- Definition of DominantColor D                                    -->
    <!-- ################################################################ -->
    <complexType name="DominantColorType" final="#all">
        <complexContent>
            <extension base="mpeg7:VisualDType">
                <sequence>
                    <element name="SpatialCoherency" type="mpeg7:unsigned5" />
                    <element name="Value" maxOccurs="8">
                        <complexType>
                            <sequence>
                                <element name="Percentage" type="mpeg7:unsigned5" />
                                <element name="Index">
                                    <simpleType>
                                        <restriction>
                                            <simpleType>
                                                <list itemType="mpeg7:unsigned12" />
                                            </simpleType>
                                            <length value="3" />
                                        </restriction>
                                    </simpleType>
                                </element>
                                <element name="ColorVariance" minOccurs="0">
                                    <simpleType>
                                        <restriction>
                                            <simpleType>
                                                <list itemType="mpeg7:unsigned1" />
                                            </simpleType>
                                            <length value="3" />
                                        </restriction>
                                    </simpleType>
                                </element>
                            </sequence>
                        </complexType>
                    </element>
                </sequence>
            </extension>
        </complexContent>
    </complexType>
    <!-- ################################################################ -->
    <!-- Definition of ScalableColor D                                    -->
    <!-- ################################################################ -->
    <complexType name="ScalableColorType" final="#all">
        <complexContent>
            <extension base="mpeg7:VisualDType">
                <sequence>
                    <element name="Coeff" type="mpeg7:integerVector" />
                </sequence>
                <attribute name="numOfCoeff" use="required">
                    <simpleType>
                        <restriction base="integer">
                            <enumeration value="16" />
                            <enumeration value="32" />
                            <enumeration value="64" />
                            <enumeration value="128" />
                            <enumeration value="256" />
                        </restriction>
                    </simpleType>
                </attribute>
                <attribute name="numOfBitplanesDiscarded" use="required">
                    <simpleType>
                        <restriction base="integer">
```

```
                        <enumeration value="0" />
                        <enumeration value="1" />
                        <enumeration value="2" />
                        <enumeration value="3" />
                        <enumeration value="4" />
                        <enumeration value="6" />
                        <enumeration value="8" />
                    </restriction>
                </simpleType>
            </attribute>
        </extension>
    </complexContent>
</complexType>
<!-- Definition of RegionLocator Datatype -->
<complexType name="RegionLocatorType" final="#all">
    <sequence>
        <element name="Box" minOccurs="0">
            <complexType>
                <simpleContent>
                    <extension base="mpeg7:BoxListType" />
                </simpleContent>
            </complexType>
        </element>
        <element name="Polygon" minOccurs="0">
            <complexType>
                <sequence>
                    <element name="Coords" type="mpeg7:IntegerMatrixType" />
                </sequence>
            </complexType>
        </element>
    </sequence>
</complexType>
<!-- Definition of BoxList Datatype -->
<complexType name="BoxListType">
    <simpleContent>
        <restriction base="mpeg7:IntegerMatrixType">
            <minLength value="4" />
            <maxLength value="6" />
        </restriction>
    </simpleContent>
</complexType>
<!--
****************************************************************************-->
<!--        ************************        ISO/IEC        15938-5        MDS
*****************************-->
<!--
****************************************************************************-->
<annotation>
    <documentation>This section contains MDS tools defined in ISO/IEC 15938-
5</documentation>
</annotation>
<!-- ######################################## -->
<!--  Definition of MPEG-7 Base types (4.2)  -->
<!-- ######################################## -->
<!-- Definition of MPEG-7 base type -->
<complexType name="Mpeg7BaseType" abstract="true">
    <complexContent>
        <restriction base="anyType" />
    </complexContent>
</complexType>
<!-- Definition of generic D -->
```

```
    <complexType name="DType" abstract="true">
        <complexContent>
            <extension base="mpeg7:Mpeg7BaseType" />
        </complexContent>
    </complexType>
    <!-- Definition of generic DS -->
    <complexType name="DSType" abstract="true">
        <complexContent>
            <extension base="mpeg7:Mpeg7BaseType">
                <sequence>
                    <element  name="Header"  type="mpeg7:HeaderType"  minOccurs="0"
maxOccurs="unbounded" />
                </sequence>
                <attribute name="id" type="ID" use="optional" />
            </extension>
        </complexContent>
    </complexType>
    <!-- Definition of Visual D -->
    <complexType name="VisualDType" abstract="true">
        <complexContent>
            <extension base="mpeg7:DType" />
        </complexContent>
    </complexType>
    <!--  kept for DescriptionMetadata -->
    <!-- HeaderType -->
    <complexType name="HeaderType" abstract="true">
        <complexContent>
            <extension base="mpeg7:Mpeg7BaseType">
                <attribute name="id" type="ID" use="optional" />
            </extension>
        </complexContent>
    </complexType>
    <!-- Mpeg7Type -->
    <complexType name="Mpeg7Type" abstract="true">
        <sequence>
            <element                                          name="DescriptionMetadata"
type="mpeg7:DescriptionMetadataType" minOccurs="1" />
        </sequence>
    </complexType>
    <!-- Definition of Mpeg7 Element -->
    <element name="Mpeg7">
        <complexType>
            <complexContent>
                <extension base="mpeg7:Mpeg7Type">
                    <sequence>
                        <element                                          name="Description"
type="mpeg7:CompleteDescriptionType" minOccurs="0" />
                    </sequence>
                </extension>
            </complexContent>
        </complexType>
    </element>
    <!-- ################################################################ -->
    <!--  Definition of Complete description top-level types (4.4.2)  -->
    <!-- ################################################################ -->
    <!-- Definition of CompleteDescription Top-level Type -->
    <complexType name="CompleteDescriptionType" abstract="true">
        <sequence>
            <element  name="Relationships"  type="mpeg7:GraphType"  minOccurs="0"
maxOccurs="unbounded" />
        </sequence>
```

```
    </complexType>
    <!-- ################################################################ -->
    <!--  Definition of Content description top-level types (4.4.3)  -->
    <!-- ################################################################ -->
    <!-- Definition of ContentDescription Top-level Type -->
    <complexType name="ContentDescriptionType" abstract="true">
       <complexContent>
          <extension base="mpeg7:CompleteDescriptionType" />
       </complexContent>
    </complexType>
    <!-- Definition of ContentEntity Top-level Type -->
    <complexType name="ContentEntityType">
       <complexContent>
          <extension base="mpeg7:ContentDescriptionType">
             <sequence>
                <element                                    name="MultimediaContent"
type="mpeg7:MultimediaContentType" />
             </sequence>
          </extension>
       </complexContent>
    </complexType>
    <!-- ################################################################ -->
    <!--  Definition of Content management top-level types (4.4.4)  -->
    <!-- ################################################################ -->
    <!-- Definition of ContentManagement Top-level Type -->
    <complexType name="ContentManagementType" abstract="true">
       <complexContent>
          <extension base="mpeg7:CompleteDescriptionType" />
       </complexContent>
    </complexType>
    <!-- Definition of ClassificationSchemeDescription Top-level Type -->
    <complexType name="ClassificationSchemeDescriptionType">
       <complexContent>
          <extension base="mpeg7:ContentManagementType">
             <sequence>
                <element                                 name="ClassificationScheme"
type="mpeg7:ClassificationSchemeType" maxOccurs="unbounded" />
             </sequence>
          </extension>
       </complexContent>
    </complexType>
    <!-- ################################################ -->
    <!--  Definition of Multimedia content entity tools (4.4.5)   -->
    <!-- ################################################ -->
    <!-- Definition of MultimediaContent Content Entity -->
    <complexType name="MultimediaContentType" abstract="true">
       <complexContent>
          <extension base="mpeg7:DSType" />
       </complexContent>
    </complexType>
    <!-- Definition of Video Content Entity -->
    <complexType name="VideoType">
       <complexContent>
          <extension base="mpeg7:MultimediaContentType">
             <sequence>
                <element name="Video" type="mpeg7:VideoSegmentType" />
             </sequence>
          </extension>
       </complexContent>
    </complexType>
    <!-- ################################################ -->
```