

Part V

Content Description

MPEG-7

motivation

- ❑ exponentially increasing amount of audiovisual information is becoming available in digital form (digital archives, WWW, broadcast data streams, personal and professional DBs, etc.)
- ❑ new ways to produce, offer, filter, search, and manage digitized multimedia information
- ❑ broadband is being offered with increasing audio and video quality and speed of access.
- ❑ information value often depends on how easy information can be found, retrieved, accessed, filtered and managed.

MPEG-7

motivation (continued)

- ❑ users will be confronted with such a large number of contents that efficient and accurate access to this content will seem to be unimaginable.
- ❑ identifying and managing content efficiently is becoming more difficult, because of the sheer volume.
- ❑ problem not restricted to database retrieval applications such as digital libraries, but extends to areas like broadcast channel selection, multimedia editing, and multimedia directory services

MPEG-7

solution to this problem: MPEG-7

- ❑ ISO/IEC standard; formally called "Multimedia Content Description Interface"
- ❑ Both human users and automatic systems that process audiovisual information are within the scope of MPEG-7.
- ❑ aims at offering a comprehensive set of audiovisual description tools that form the basis for applications enabling the needed quality access to content
- ❑ active people: broadcasters, electronics manufacturers, content creators and managers, publishers and intellectual property rights managers, telecommunication service providers and academia

MPEG-7 Main Goals

- ❑ description of multimedia content,
- ❑ flexibility in data management, and
- ❑ globalization and interoperability of data resources.

MPEG-7 aims to standardize

- ❑ a set of Description Schemes and Descriptors to describe data,
- ❑ a language to specify Description Schemes, such as the Description Definition Language(DDL), and
- ❑ a scheme for coding the Description.

Terminology

Definition: A **Feature** is a distinctive characteristic of the data that signifies something to somebody.

- ❑ Features require a meaningful Feature representation (Descriptor) and its instantiation (Descriptor Value) for a given data set.
- ❑ examples: the color of an image, fundamental frequency of a speech segment, rhythm of an audio segment, camera motion in a video, genre of a piece of music, title of a movie, and actors in a movie.

Terminology

Definition: A **Descriptor** is a representation of a Feature. A Descriptor defines the syntax and semantics of the Feature representation.

- ❑ A descriptor must precisely define the semantics of the feature, the associated data type, legal values, and an interpretation of the Descriptor Values.
- ❑ An example might be `Color: string`. The data type may be composite, meaning that it can be formed by concatenating multiple instances of a data type. An example of this would be `RGB-Color: [integer, integer, integer]`.

Descriptor and Descriptor Value

- ❑ several descriptors may represent a single feature—that is, address different relevant requirements.
- ❑ examples of multiple descriptors for a single feature include enumerated lists, color moments, and histograms that represent color.
- ❑ the syntax of a Descriptor might look like this:
`<name, typekind, spec> value </name>`

Definition: A **Descriptor Value** is an instantiation of a Descriptor for a given data set (or subset thereof). Descriptor Values are combined via a Description Scheme to form a Description.

Multimedia

Part V Content Description

MPEG-7

Descriptor Items

Item	Description
Name	Descriptor ID
Typekind	Data type of Descriptor Value: free text (plus language identifier), structured text (plus structure identification), integer, real, date, time/time index, version, enumeration (such as RGB values, vector), relation (Descriptor-Descriptor, Descriptor-Description Scheme, Description-Description), complex (a bitmap, histogram)*, function*, and possibly trigger (executable code)*
Spec	Contains a specification for the data type. For example, "free text" might specify the language the text is written in (Dutch, English, Japanese, and so on).
Value	An instantiation of the Descriptor is assigned to the Feature as pertaining to the data. Descriptor Values are combined via the mechanism of a Description Scheme to form a Description
*Value could contain the actual data or a link	

654

Christian Breiteneder

Multimedia

Part V Content Description

MPEG-7

Terminology

Definition: A *Description Scheme* specifies the structure and semantics of the relationships between its components, which may be both Descriptors and Description Schemes.

The diagram shows a hierarchical structure. At the top is a box labeled 'DDL'. Below it, there are two boxes labeled 'DS'. The left 'DS' box is enclosed in a dashed rectangle labeled 'defined in standard'. This 'DS' box has three children: two ovals labeled 'D' and another 'DS' box. This nested 'DS' box has two children, both ovals labeled 'D'. The right 'DS' box is not enclosed in the dashed rectangle and has one child, an oval labeled 'D'. A dashed arrow points from the 'DDL' box to the right 'DS' box. A dashed arrow points from the right 'DS' box to its child 'D' oval. A small blue triangle points to the 'D' oval under the right 'DS' box, with the text 'not in standard; defined using DDL' next to it.

655

Christian Breiteneder

Multimedia	Part V Content Description	MPEG-7
<h2>Terminology</h2>		
<p>Definition: A <i>Description Scheme</i> specifies the structure and semantics of the relationships between its components, which may be both Descriptors and Description Schemes.</p>		
655	Christian Breiteneder	

Multimedia	Part V Content Description	MPEG-7
<h2>Description Scheme</h2>		
<ul style="list-style-type: none"> ❑ The last figure presents an abstract representation of possible relations between Descriptors and Description Schemes. The arrows from DDL to Description Scheme signify that the Description Schemes are generated using the DDL. ❑ The distinction between a Description Scheme and a Descriptor is that a Descriptor is concerned with the representation of a Feature whereas the Description Scheme deals with the structure of a Description. 		
656	Christian Breiteneder	

Multimedia	Part V Content Description	MPEG-7
<h2>Description Scheme</h2>		
<ul style="list-style-type: none"> ❑ A simple Description Scheme for describing technical aspects of a shot (elements written in bold represent other Description Schemes): 		
<div>Shot_Technical_Aspects</div> <div> Lens providing information about type (wide-angle), movement (zooms), state (deep focus), masking, and so on. </div> <div> Camera providing information about distance (close-up), angle (overhead), move-ment (pan_left), position (viewpoint of the frame), and so on. </div> <div> Speed </div> <div> Color </div> <div> Granularity </div> <div> Contrast </div>		
657	Christian Breiteneder	

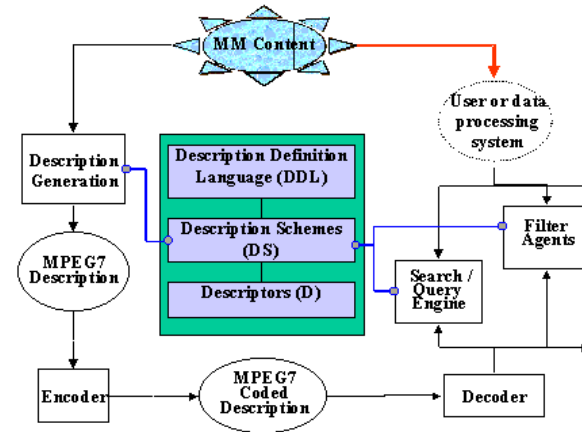
Terminology

Definition: A *Description* consists of a Description Scheme (structure) and the set of Descriptor Values (instantiations) that describe the data. A Description contains or refers to a fully or partially instantiated Description Scheme.

Definition: A *coded description* is a Description that has been encoded to fulfill relevant requirements such as compression efficiency, random access, and so on.

Definition: The *Description Definition Language (DDL)* allows the creation of new Description Schemes and, possibly, Descriptors. It also allows the extension and modification of existing Description Schemes.

MPEG-7 in Practice



The left side portrays how data is annotated, whereas the right side demonstrates how described data can be retrieved.

Square boxes describe processing tools (e.g. encoding or decoding). Circular boxes describe static elements, such as a Description.

For example, the "description generation" box is a description generation engine that produces an "MPEG-7 description" as output, shown here as a rectangular box.

Requirements

The DDL's design forms a core part of the work within MPEG-7. The DDL provides a solid descriptive (for example, SGML-based) underpinning for users to create their own Description Schemes and Descriptors. For this purpose, MPEG-7 has identified a number of requirements the DDL should cover:

- ❑ *Compositional capabilities*. The DDL supplies the ability to compose new Description Schemes and Descriptors, where a Description Scheme might be composed from multiple Description Schemes. A newly created Description Scheme must allow the creation of MPEG-7 compliant Descriptions.

Requirements

- ❑ *Transformational capabilities*. The DDL allows the reuse, extension, and inheritance of existing Descriptors and Description Schemes.
- ❑ *Unique identification*. The DDL provides mechanisms to uniquely identify Description Schemes and Descriptors so that they can be referred to unambiguously.
- ❑ *Data types*. The DDL provides a set of primitive data types, such as e.g. text, integer, real, date, time/time index, version to succinctly describe composite data types that might arise from processing digital signals (such as histograms, graphs, RGB values). Also, the DDL must provide a mechanism to relate Descriptors to data of multiple media types of inherent structure (e.g. audio, video, audio-visual presentations, etc.)

Requirements

- ❑ *Relationships within a Description Scheme and between Description Schemes.* The DDL provides the capability to express relationships between Description Schemes and among elements of a description Scheme. The DDL expresses the semantics of these relations, such as spatial, temporal, structural, and conceptual relations.
- ❑ *Relationship between Description and data.* The DDL shall supply a rich model for links and/or references between one or several Descriptions and the described data.

Parts of the Standard

The MPEG-7 Standard consists of the following parts:

1. MPEG-7 Systems - the tools that are needed to prepare MPEG-7 Descriptions for efficient transport and storage, and to allow synchronization between content and descriptions. Tools related to managing and protecting intellectual property
2. MPEG-7 Description Definition Language - the language for defining new Description Schemes and perhaps eventually also for new Descriptors.
3. MPEG-7 Audio - the Descriptors and Description Schemes dealing with (only) Audio descriptions

Parts of the Standard

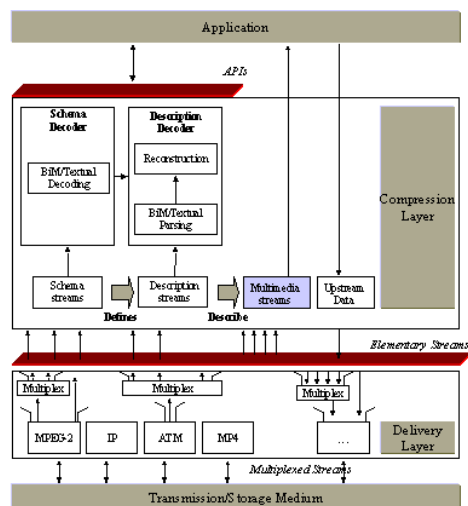
4. MPEG-7 Visual - the Descriptors and Description Schemes dealing with (only) Visual descriptions
5. MPEG-7 Multimedia Description Schemes - the Descriptors and Description Schemes dealing with generic features and multimedia descriptions
6. MPEG-7 Reference Software - a software implementation of relevant parts of the MPEG-7 Standard
7. MPEG-7 Conformance - guidelines and procedures for testing conformance of MPEG-7 implementations.

MPEG-7 Systems

MPEG-7 Systems defines the terminal architecture and the normative interfaces:

- ❑ *Terminal*—the entity that makes use of coded representation of the multimedia content; may correspond to a stand-alone application or be part of an application system:
 - ❑ application
 - ❑ compression layer
 - ❑ delivery layer
 - ❑ transmission/storage medium

Terminal Architecture



Terminal Architecture

- ❑ transmission/storage medium—refers to the lower layers of the delivery infrastructure (network layer and below as well as storage); these layers deliver multiplexed streams to the DL.
- ❑ delivery layer (DL)—encompasses mechanisms allowing synchronization, framing and multiplexing of MPEG-7 content.
 - ❑ MPEG-7 content may be delivered independently or together with the content they describe. Not all MPEG-7 streams have to be downstream (server to the client).
 - ❑ provides to the Compression layer MPEG-7 elementary streams. MPEG-7 elementary streams consist in consecutive individually accessible portion of data named *Access Units*.

Terminal Architecture

delivery layer (continued)

- ❑ An *access unit* is the smallest data entity to which timing information can be attributed
- ❑ MPEG-7 elementary streams contain information of different nature:
 - ❑ Schema information: this information defines the structure of the MPEG-7 description;
 - ❑ Descriptions information: this information is either the complete description of the multimedia content or fragments of the description.

Terminal Architecture

- ❑ compression layer—the flow of Access Units (either textual or binary encoded) is parsed, and the content description is reconstructed.
 - ❑ MPEG-7 does not mandate the reconstruction of a textual representation as an intermediate step of the decoding process.
 - ❑ The MPEG-7 binary stream can be either parsed by the BiM parser, transformed in textual format and then transmitted in textual format to further reconstruction processing, or the binary stream can be parsed by the BiM parser and then transmitted in proprietary format to further processing.

MPEG-7 Visual

MPEG-7 visual description tools consist of *basic structures* and descriptors that cover the following basic visual features:

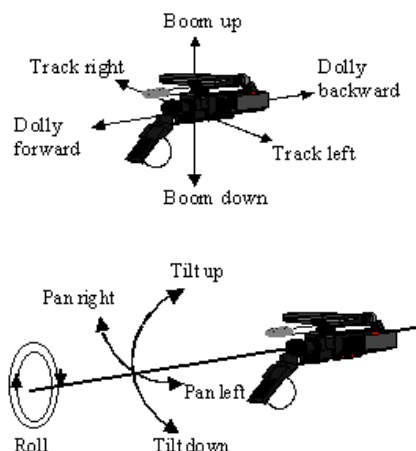
- ☐ Color
- ☐ Texture
- ☐ Shape
- ☐ Motion
- ☐ Localization
- ☐ Others

Each category consists of *elementary* and *sophisticated descriptors*

Descriptors

- ☐ 5 Basic structures: the Grid Layout, the Time Series, MultiView, the Spatial 2D Coordinates, and Temporal Interpolation.
- ☐ 8 Color Descriptors: Color space, Dominant Colors, Color Quantization, GoF/GoP Color, Color-Structure, Color Layout and Scalable Color Histogram
- ☐ 3 Texture Descriptors: Homogeneous Texture, Texture Browsing and Edge Histogram
 - ☐ Texture Browsing Descriptor—useful for representing homogeneous texture for browsing type applications, and requires only 12 bits (maximum). It provides a perceptual characterization of texture (similar to a human) in terms of regularity, coarseness and directionality

Camera Motion Descriptors



Motion Descriptors

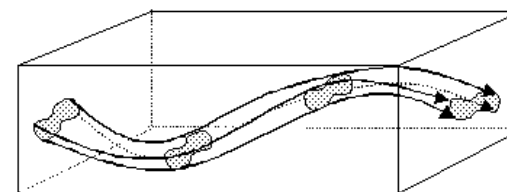
- ☐ motion trajectory of an object—a simple, high level feature, defined as the localization in time and space, of one representative point of this object.
- ☐ the descriptor is essentially a list of keypoints (x, y, z, t) along with a set of optional interpolating functions that describe the path of the object between keypoints, in terms of acceleration. The speed is implicitly known by the keypoints specification.
- ☐ The keypoints are specified by their time instant and either their 2-D or 3-D Cartesian coordinates, depending on the intended application. The interpolating functions are defined for each component $x(t)$, $y(t)$, and $z(t)$ independently.

Motion Descriptors

- ❑ Parametric motion—has been extensively used within various related image processing and analysis areas, including motion-based segmentation and estimation, global motion estimation, mosaicing and object tracking. Parametric motion models have been already used in MPEG-4, for global motion estimation and compensation and sprite generation
- ❑ activity descriptor—captures the intuitive notion of 'intensity of action' or 'pace of action' in a video segment. Examples of high 'activity' include scenes such as 'goal scoring in a soccer match', 'scoring in a basketball game', 'a high speed car chase' etc..

Localization

- ❑ Region Locator—enables localization of regions within images or frames by specifying them with a brief and scalable representation of a Box or a Polygon.
- ❑ Spatio Temporal Locator—describes spatio-temporal regions in a video sequence, such as moving object regions, and provides localization functionality (for hypermedia and CBR)



Others

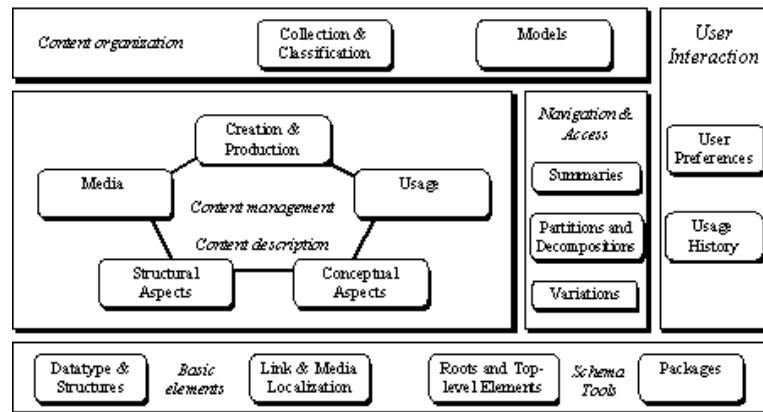
- ❑ Face Recognition descriptor—can be used to retrieve face images which match a query face image.
 - ❑ The descriptor represents the projection of a face vector onto a set of basis vectors which span the space of possible face vectors. The Face Recognition feature set is extracted from a normalized face image.
 - ❑ This normalized face image contains 56 lines with 46 intensity values in each line. The centers of the two eyes in each face image are located on the 24th row and the 16th and 31st column for the right and left eye respectively.

MPEG-7 Multimedia Description Schemes

Organization of MDS tools

- ❑ Basic Elements,
- ❑ Schema Tools,
- ❑ Content Description,
- ❑ Content Management,
- ❑ Content Organization,
- ❑ Navigation and Access, and
- ❑ User Interaction

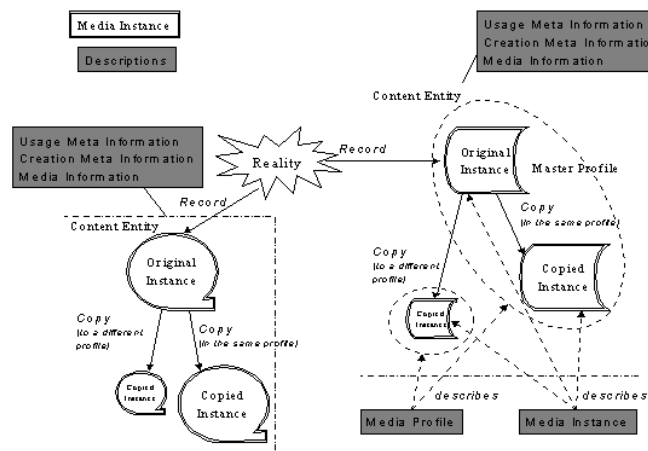
Organization of MDS Tools



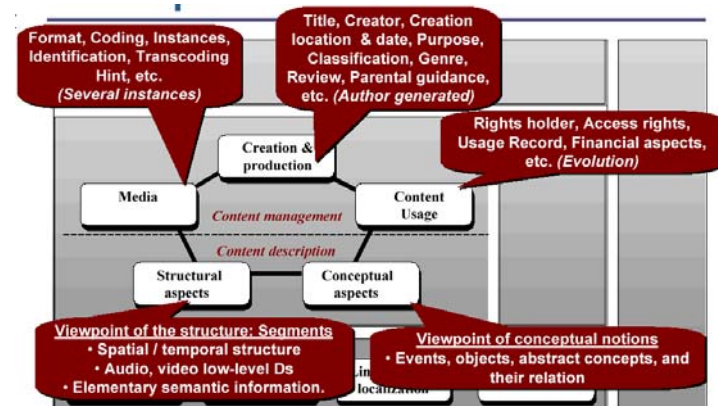
Content Management (CM)

- ❑ CM description tools allow the description of the life cycle of the content, from content creation to consumption.
- ❑ The content described by MPEG-7 descriptions can be available in different modalities, formats, coding schemes, and there can be several instances.
 - ❑ For example, a concert can be recorded in two different modalities: audio and audio-visual. Each of these modalities can be encoded by different encoding schemes. This creates several media profiles.

Content Management



Content Management and Description



Content Description

Description of the content structural aspects

- ❑ core element of this part is the *Segment DS* (description of the physical and logical aspects of audio-visual content)
- ❑ Segment DSs may be used to form segment trees.
- ❑ MPEG-7 also specifies a Graph DS that allows the representation of complex relations between segments. It is used to describe spatio-temporal relationships, between segments that are not described by the tree structures
- ❑ A segment represents a section of an audio-visual content item.

Content Description—Segment DS

- ❑ *Segment DS* is an abstract class (in the sense of oop). It has five subclasses: AudioVisual Segment DS, Audio Segment DS, Still Region DS, Moving Region DS and Video Segment DS. Therefore, it may have both spatial and temporal properties.
- ❑ examples: A *temporal segment* may be a set of samples in an audio sequence, represented by an *Audio Segment DS*, a set of frames in a video sequence, represented by a *Video Segment DS* or a combination of both audio and visual information described by an *Audio Visual Segment DS*. A *spatial segment* may be a region in an image or a frame in a video sequence, represented by a *Still Region DS*.
- ❑ a segment can be decomposed into sub-segments through the Segment Decomposition DS.

Segment DS

- ❑ A segment is not necessarily connected, but may be composed of several non-connected components.
- ❑ Connectivity refers to both spatial and temporal domains. A temporal segment (Video Segment, Audio Segment and AudioVisual Segment) is said to be temporally connected if it is a sequence of continuous video frames or audio samples.
- ❑ A spatial segment (Still Region) is said spatially connected if it is a group of connected pixels.
- ❑ A spatio-temporal segment (Moving Region) is said spatially and temporally connected if the temporal segment where it is instantiated is temporally connected and if each one of its temporal instantiations in frames is spatially connected (Note: different from the classical connectivity in a 3D space).

Segment Discription

Any segment may be described by

- ❑ Creation information
- ❑ Usage information
- ❑ Media information
- ❑ Textual annotation
- ❑ Specific features

Specific Features for Segment Discription

Features	Video Segment	Still Region	Moving Region	Audio Segment
Time	X	-	X	X
Shape	-	X	X	-
Color	X	X	X	-
Texture	-	X	-	-
Motion	X	-	X	-
Camera motion	X	-	-	-
Mosaic	X	-	-	-
Audio features	-	-	X	X

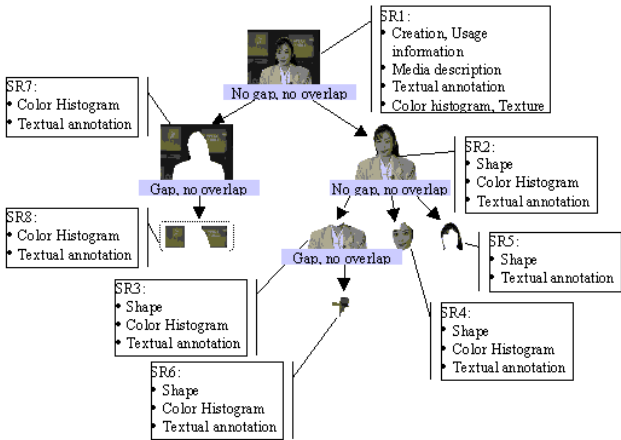
Example of Segment Description

Example — Still image, discribed as a tree

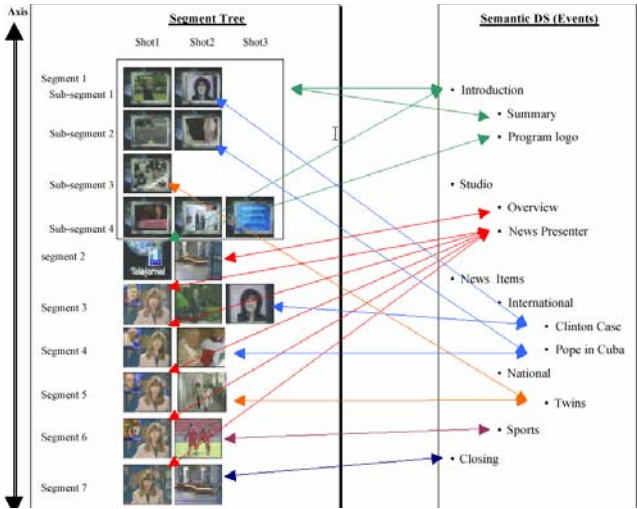
- ❑ Original image — discribed as still region SR1: creation (title, creator), usage information (copyright), media information (file format), textual annotation (summarizing the image content), color histogram, texture descriptor
- ❑ Segment Tree — composed of 8 still regions, each of these are assigned types of features.

It is not necessary to repeat the creation, usage information and media information in the tree hierarchy, since the children segments are assumed to inherit their parent value

Example of Image Description



Segment Tree



Segment Graph

Hierarchical structures as trees are adequate for efficient access, retrieval and scalable description. But for certain applications they imply constraints that may make them inappropriate. In such cases we use —

- ❑ Segment Graph DS; defined by
 - ❑ Set of Nodes - representing segments
 - ❑ Set of edges - specifying the relationship between the nodes

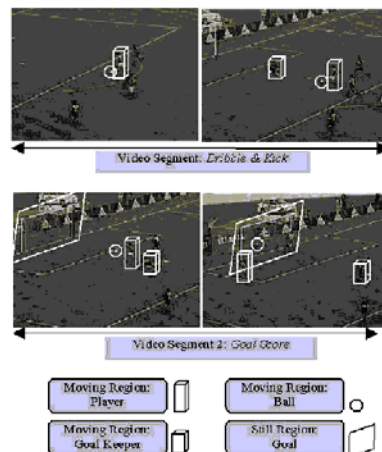
Example of Segment Relationship Graph

Example — Excerpt of a football match

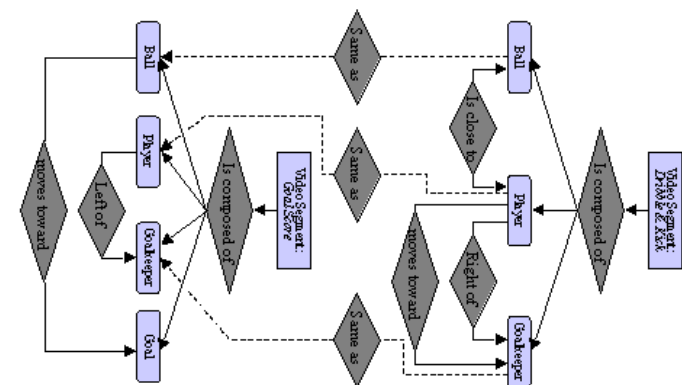
Two Video segments, one still region, three moving regions have been defined

- ❑ Video Segment *Dribble&Kick* — involves the moving regions *Player*, *Goalkeeper*, *Ball*. The *Ball* remains close to the *Player* who is moving towards the *Goalkeeper*. The *Player* appears on the *Right of the Goalkeeper*.
- ❑ Video Segment *Goal Score* — involves the moving regions *Player*, *Goalkeeper*, *Ball* and the still region *Goal*. *Player Left of the Goalkeeper*, the *Ball* moves towards the *Goal*.

Example of Segment Relationship Graph



Example of Segment Relationship Graph

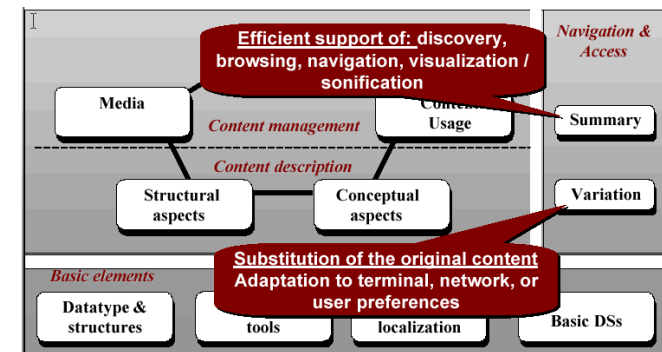


Navigation and Access

MPEG-7 provides DSs that facilitate navigation and access of audio-visual content by specifying

- ❑ Summaries — enable efficient browsing, navigation, discovery, visualization, sonification
- ❑ Views and Partitions — provide views of the av-data in the space or frequency domain, which allows multi-resolution and progressive access
- ❑ Variations — specify the relation between different of av-material, which allows adaptive selection of the different variations of the content under different delivery conditions

Navigation and Access



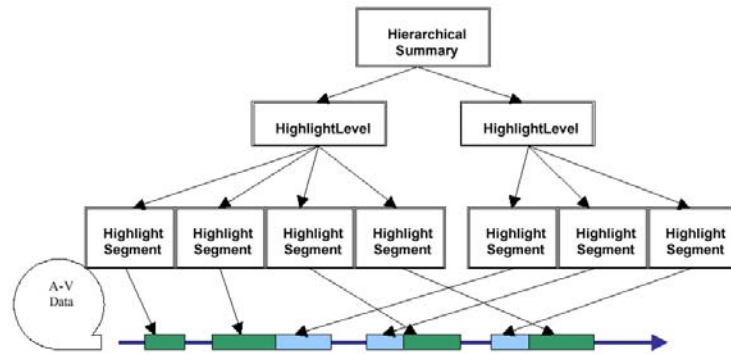
Summary

Summary DSs allow the av-content to be navigated in either a hierarchical or sequential fashion —

- ❑ Hierarchical Summary
- ❑ Sequential Summary

Hierarchical Summary

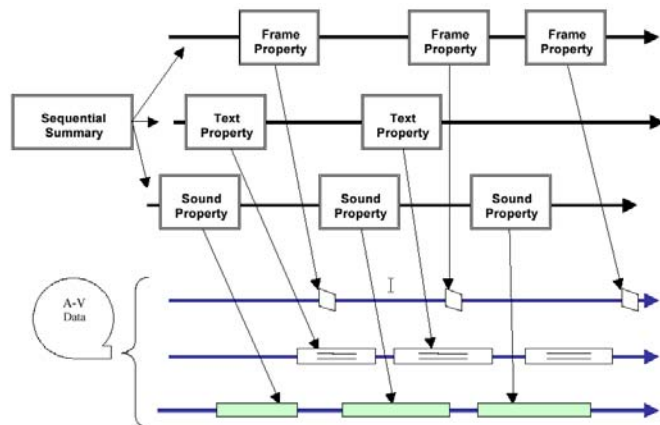
- ❑ Hierarchical Summary DS — organizes the content into successive levels that describe the av-content at different levels of detail from coarse to fine. The hierarchy forms a tree.
- ❑ Highlightlevel DS — specifies the elements of the Hierarchical Summary DS by providing a summary at a particular level of detail. It refers to a sequence of av-segments (image, key-frame, video or audio clip)
Example: news program
 - ❑ level1 - local news, sports, weather, society...
 - ❑ more detailed level - detailed highlights from sport segment corresponding to the events of a football game



Sequential Summary

- ❑ Sequential Summary DS — It specifies a summary consisting of a sequence of audio clips, a sequence of images or video frames, possibly synchronized with audio or text, that compose a slide-show or a audio-visual skim
- ❑ Sequential Summary may be stored separately from the original av-content - allows fast navigation and access.
- ❑ Alternatively, the Sequential Summaries may link directly to the original av-content - reduce storage.

Sequential Summary

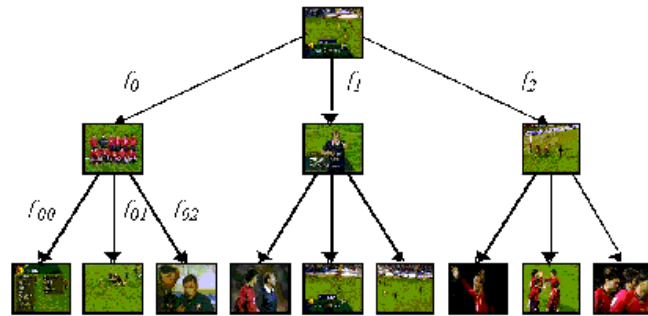


Example: Hierarchical Summary

Example — Football game

- ❑ Root — the whole game summarized into a single frame
- ❑ Next level — three frames that summarize different segments of the video (team, referee, goal score)
- ❑ Bottom level — additional frames, depicting in more detail the scenes depicted in the segments.

Example: Hierarchical Summary



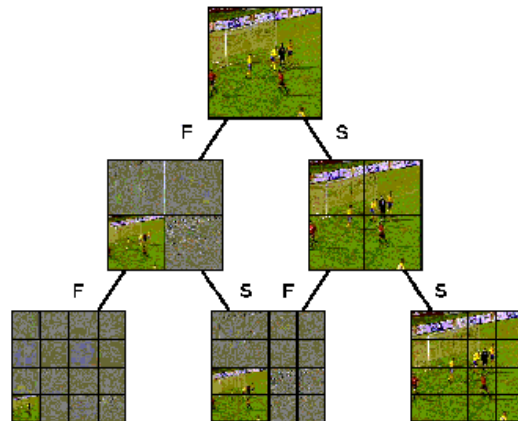
Partitions and Decomposition

Example —The Space and Frequency Graph of an image

Nodes correspond to the different space and frequency views of the image —

- ☐ views in space - spatial segments (transitions S)
- ☐ views in frequency - wavelet subbands (transitions F)
- ☐ views in space and frequency - wavelet subbands of spatial segments (transition S+F)

This kind of decomposition enables efficient multi-resolution access and progressive retrieval of the image data



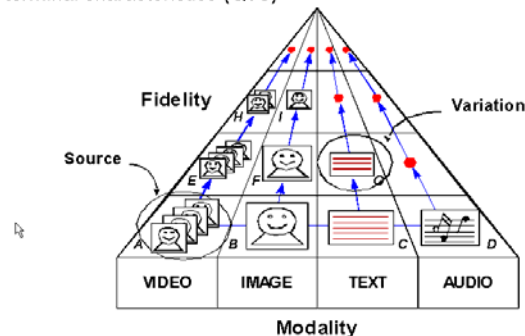
Variation of the Content

a server, proxy or terminal is to select the most suitable variation of av-content (capabilities of terminal devices, network conditions, user preferences.....)

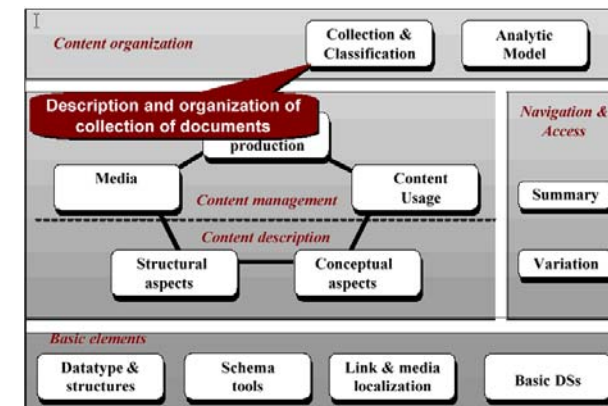
- ☐ Variation DS — specifies the different variations of an av-content, such as summaries, compressed or low resolution version, different languages and modalities (audio, video, image, text)
 - ☐ Variation fidelity value - quality of variation compared to the original
 - ☐ Variation type attribute - summary, abstract, compression, modality translation, color reduction, language translation..

Variation

Universal Multimedia Access: Adapt delivery to network and terminal characteristics (QoS)



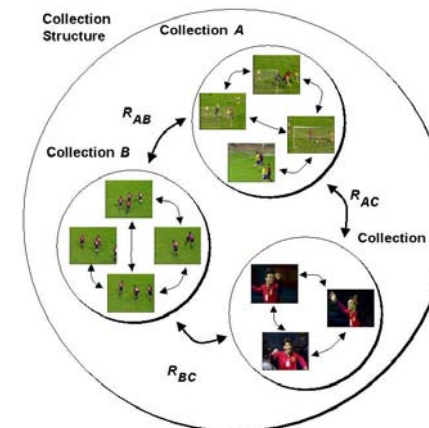
Content Organization



Collection

- ❑ Collection Structure DS — groups av-content, segments, events, or objects into collection clusters and specifies properties that are common to the elements; describes relationships among collection clusters
- ❑ Example: Football game —
 - each collection consists of a set of images with common properties (similar events in the game)
 - ❑ relationships within each collection - degree of similarity of images
 - ❑ relationships across the collections - degree of similarity of collections

Collection



User Interaction

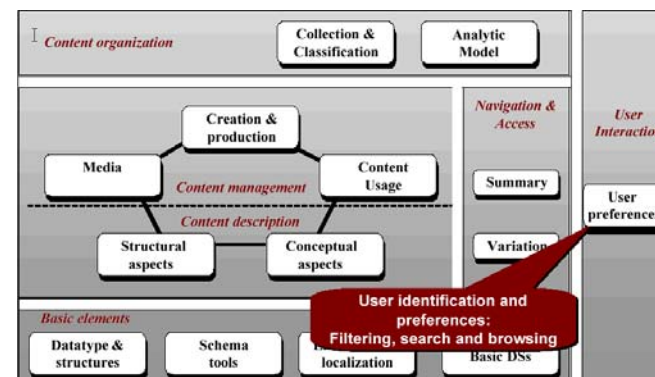
User Preference DSs

MPEG-7 av-content descriptions can be matched to the User-Preference DS descriptions in order to select and personalize av-content for more efficient access, presentation and consumption

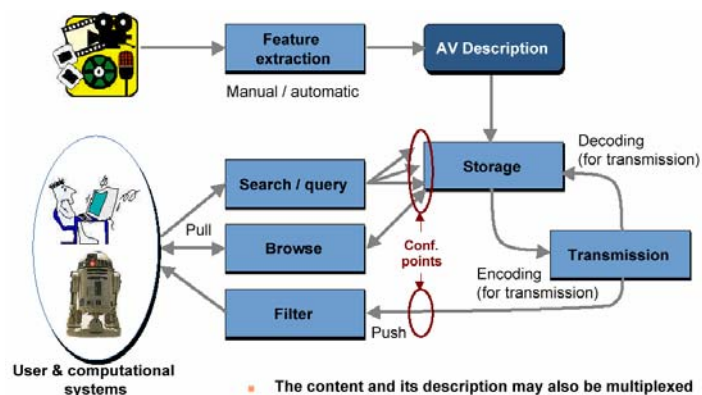
UserPreference DS allows —

- ☐ specification of preferences for different types of content and modes of browsing, including context dependency in terms of time and place
- ☐ weighting of the relative importance of different preferences
- ☐ specification of the privacy characteristics of preferences

User Interaction



Summary—Information Flow



Summary—Further Information

- ☐ major MPEG-7 documents are public:
www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm