EE665000 視訊處理

MPEG-7 Standard Overview

MPEG-7: Content and Objective

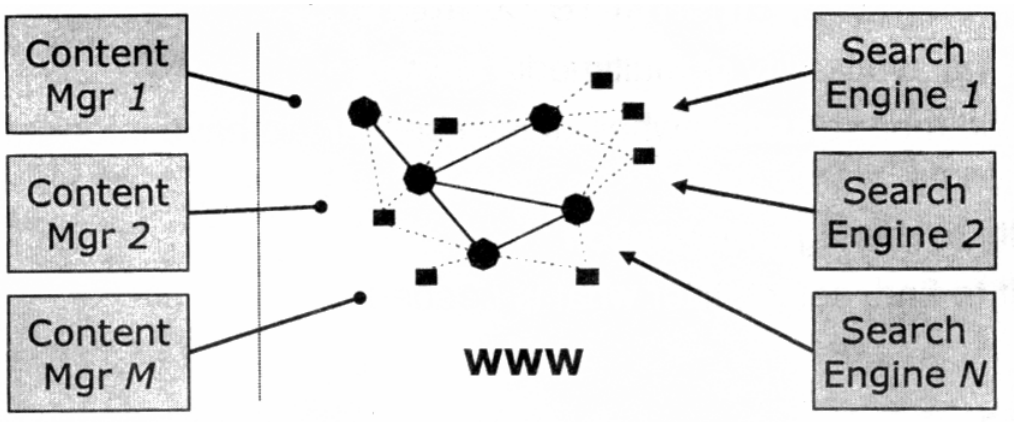
- Content, content, and more content …
  - Increasing availability of multimedia content
  - More and more situation where it is necessary to have information about the content
  - Difficulty to manage information
  - Difficult to find, select, filter what is needed

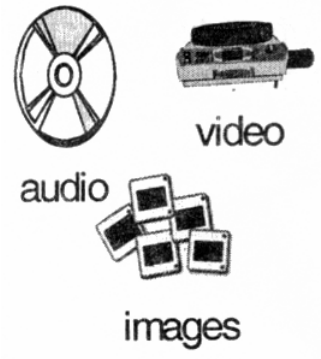
- Objective
  - Standardize content-based descriptions for various types of audio-visual information, allowing quick and efficient content identification, and addressing a large range of applications
  - MPEG-1, -2, and -4 represent the content itself (“ the bits”)
  - MPEG-7 should represent information about the content (“the bits *about* the bits”)

Data Targets & Types of Descriptions

- Most types of audio-visual info are considered (targets):
  - Audio and speech
  - Moving video, still pictures, graphics, 3D models
  - Object relations in a scene, etc.
  - Descriptions are independent of data format
- Descriptions can be classified into two broad categories:
  - Information that is present in the content
    - Low-level features that are automatically extracted e.g., (video) color, texture, motion (audio) pitch, tempo, volume
    - High-level (semantic) feature related to the human interpretation of the content, e.g., “car driving fast” or “tiger attacks deer”
  - Information that cannot be deduced from the content
    - E.g., data and time, author, copyright data, genre, parental rating, links to other related material

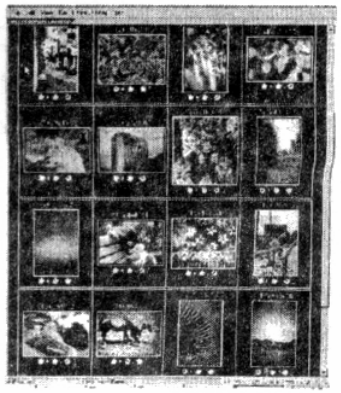
Why do we need a standard?





Interoperable Search & Retrieval:

- **Search**: color, texture, shape, motion, spatial
- **Browse**: video parsing, shot detection, key-frames
- **Filter**: object detection, classification

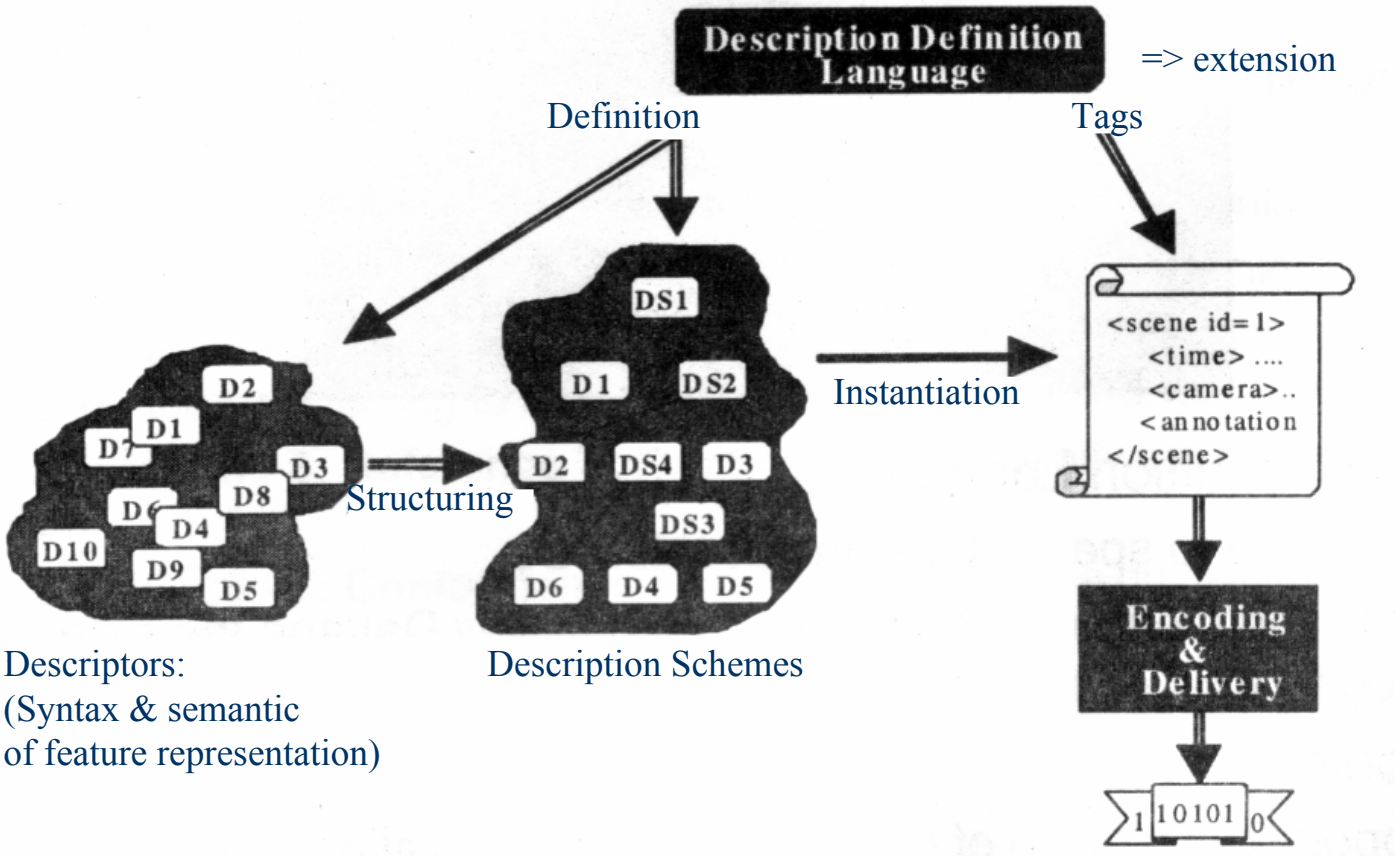


Scope of MPEG-7

- The description generation (feature extraction, indexing) is not a normative part of the standard
- The description consumption (search engine, retrieval process) is also not a normative part of the standard
- Just the description is normative
  - Standardizes the minimum needed to ensure interoperability
  - Specifies four types of normative elements
    - Descriptors
    - Description schemes (DSs)
    - Description Definition Language (DDL)
    - Coding schemes

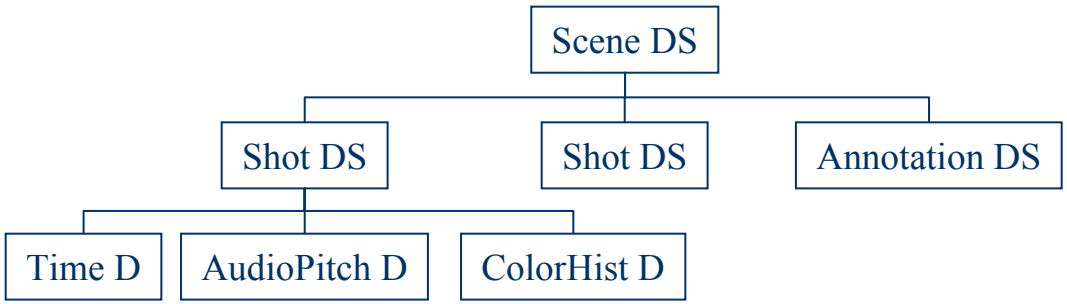
MPEG-7 Elements (I)

- MPEG-7 main elements

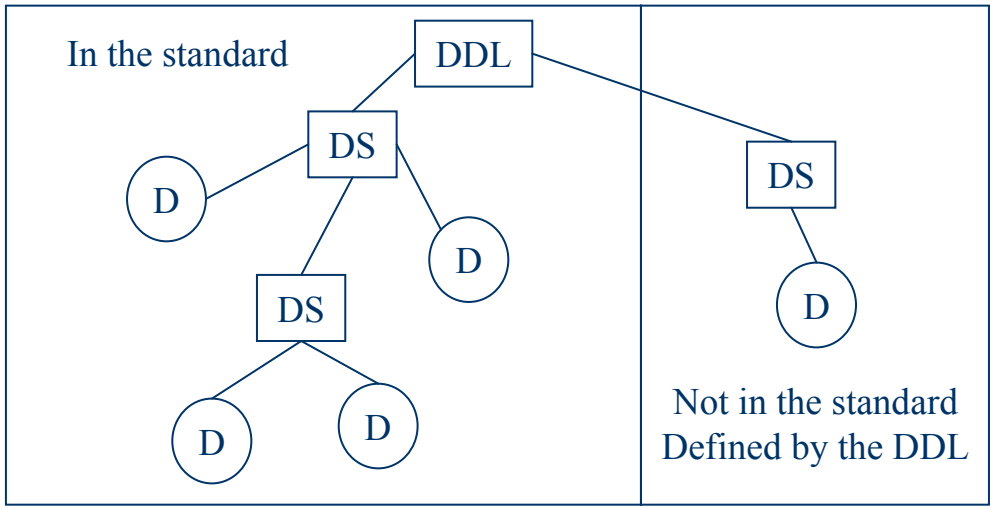


MPEG-7 Elements (II)

- Description Tools
  - Descriptors (Ds): representations of features
    - To describe various types of features of multimedia information
    - To define the syntax/semantics of each feature representation
    - Color example – specify histogram in RGB color space
  - Description Schemes (DSs)
    - Specifies the structure and semantics of the relationships between its components, which may be either Ds or other DSs



MPEG-7 Elements (III)



- Description Definition Language (DDL)
  - The language to specify DSs and possibly Ds
  - Will allow the creation of new DSs (and possibly Ds) and the extension and modification of existing DSs
- System tools
  - To support multiplexing of description, synchronization issues, transmission mechanisms, file format, etc

MPEG-7 Workplan

- Start to work … Oct. 1996
- Call for proposals … Oct. 1998
- Submission deadline … Feb. 1999
- Working draft … Mar. 2000
- Committee draft … Oct. 2000
- Final committee draft … Jan. 2001
- Draft international standard … July 2001
- International standard … Oct. 2001
- Extensions … Feb. 2002
- Amd … Sept. 2002

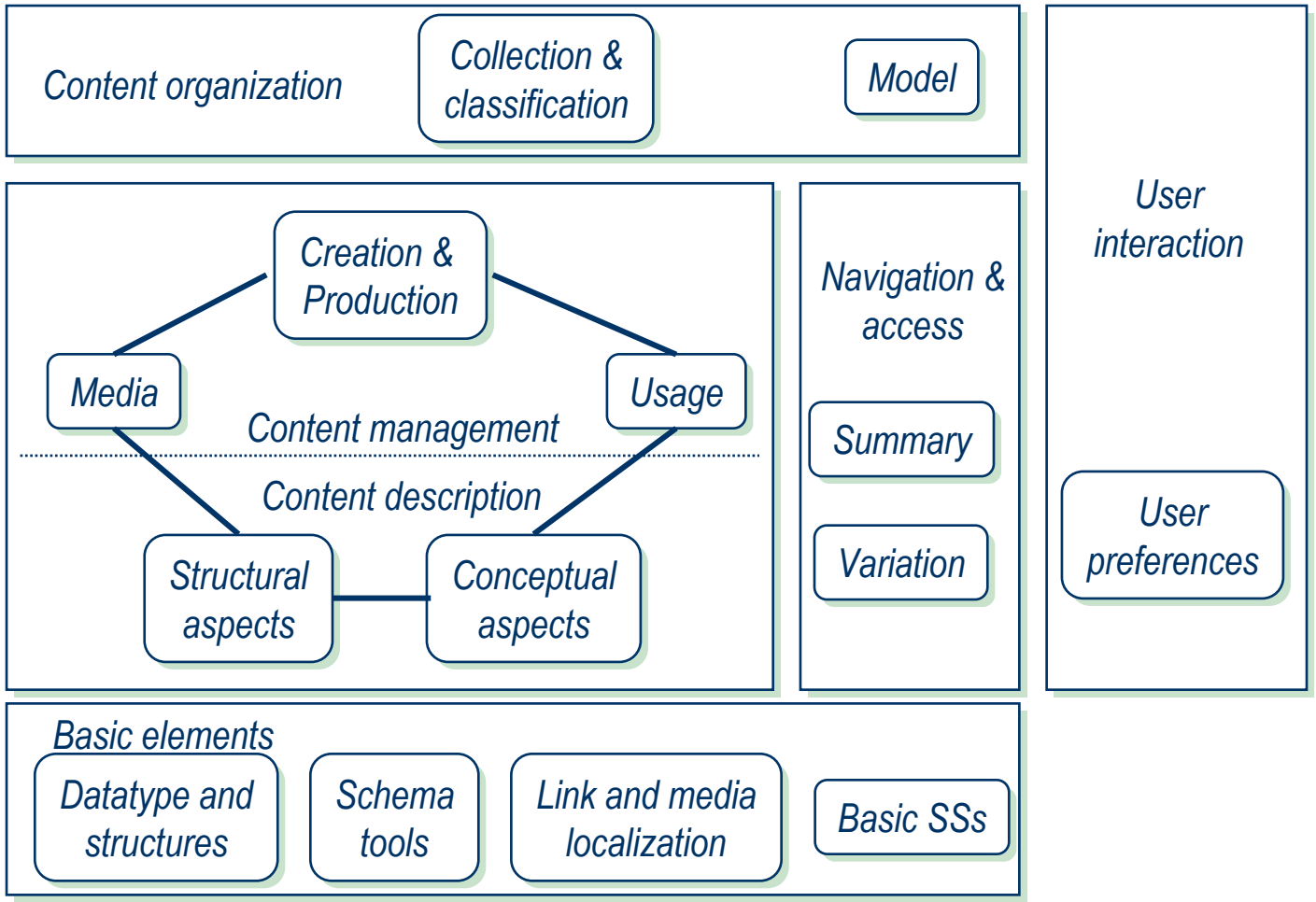
Parts of the MPEG-7 Standard

- **ISO/IEC 15938-1: Systems** – specifies the tools for preparing descriptions for efficient transport and storage (binary format), compressing descriptions, and allowing synchronization between content and descriptions
- **ISO/IEC 15938-2: Description Definition Language** – specifies the language for defining the standard set (syntax) of description tools (DSs, Ds, and datatypes) and for defining new description tools
- **ISO/IEC 15938-3: Visual** – specifies the description tools pertaining to visual content
- **ISO/IEC 15938-4: Audio** – specifies the description tools pertaining to audio content
- **ISO/IEC 15938-5: Multimedia Description Schemes** – specifies the generic description tools pertaining to multimedia including audio and visual content
- **ISO/IEC 15938-6: Reference software** – provides a software implementation of the standard
- **ISO/IEC 15938-7: Conformance testing** – specifies the guidelines and procedures for testing conformance of implementations of the standard
- **ISO/IEC 15938-8: Extraction and use of descriptions** – provides guidelines and examples of the extraction and use of descriptions

MPEG-7 Multimedia Description Schemes – MDS

- DSs are metadata structures for describing and annotating audio-visual (AV) content
- DSs provide a standardized way of describing in XML the important concepts related to AV content description and content management in order to facilitate searching, indexing, filtering, and access
- DSs are defined using the MEPG-7 Description Definition Language (DDL), which is based on the XML Schema Language, and are instantiated as documents or streams
- DSs are designed primarily to describe higher-level AV features such as regions, segments, objects, events; and other immutable metadata related to creation and production, usage, and so forth
- DSs produce more complex descriptions by integrating together multiple Descriptors and DSs, and by declaring relationships among the description components
- Dss are categorized as pertaining to the multimedia, audio, or visual domain. Typically, the multimedia DSs describe content consisting of a combination of audio, visual data, and possibly textual data
- In some cases, automatic tools can be used for instantiating the DSs, but in many cases instantiating DSs requires human assisted extraction or authoring tool

MDS Categories (1)



MDS Categories (2)

- Content organization
  - Collections: Collection Structure DS
    - Groups the audio-visual content, segments, events, or objects into collection clusters and specifies properties that are common to the elements
    - Describes collections of audio-visual content or pieces of audio-visual material, statistics and models of the attribute values of the elements and relationships among collection clusters
- Models: Model DSs provide tools for modeling the attributes and features of audio-visual content
  - Probability Model DS: provides fundamental DSs for specifying different statistical functions and probabilistic structures
  - Analytic Model DS: describe collections of examples of audio-visual data or clusters of descriptors that provide models for particular semantic classes and specifies semantic labels that indicate the classes being modeled
  - Classifier DSs describe different types of classifiers that are used to assign semantic labels to audio-visual data based

MDS Categories (3)

- Content management
  - Creation and production
    - Title, creator, creation locations, creation time, texture annotations
    - Subject, purpose, language, …
  - Media coding, storage and file formats
    - Format, compression, …
  - Content usage
    - Usage right, usage records,
- Content description
  - Structural aspects: describes the audio-visual content from the view point of its structure, such as spatial, temporal or spatio-temporal structure of the audio-visual content, or a hierarchical structure
  - Conceptual aspects: describes the audio-visual content from the viewpoint of real-world semantics and conceptual notions such as objects, events, abstract concepts and relationships
  - Structure DSs and Semantic DSs are related by a set of links, which allows the audio-visual content to be described on the basis of both content structure and semantics together

MDS Categories (4)

- Navigation and Access
  - DSs for facilitating browsing and retrieval of audio-visual content by defining summaries, partitions and decompositions, and variations of the audio-visual material
  - Summaries
    - Hierarchical summary organizes the information into successive levels, each describing the audio-visual content at a different level of detail. In general, the levels closer to the root of the hierarchy provide more coarse summaries, and levels further from the root provide more detailed summaries.
    - Sequential summary provides a sequence of images or video frames, possibly synchronized with audio, which may compose a slide-show or audio-visual skim.
  - Partitions and Decompositions
    - Describe different decompositions of the audio-visual signal in space, time and frequency
    - Describe different views of the audio-visual data, which is important for multi-resolution access and progressive retrieval
  - Variations
    - Provide information about different variations of audio-visual programs, such as summaries and abstracts; scaled, compressed and low-resolution versions; and versions with different languages and modalities – audio, video, image, text, and so forth
    - Allow the selection on the most suitable variation of an audio-visual program, which can replace the original, if necessary, to adapt to the different capabilities of terminal devices, network conditions or user preferences

MDS Categories (5)

- User interaction
  - User interaction DSs describe user preferences and usage history pertaining to the consumption of the multimedia material
  - This allows, for example, matching between user preferences and MPEG-7 content descriptions in order to facilitate personalization of audio-visual content access, presentation and consumption
- Basic elements
  - Schema tools: assist in the formation, packaging, and annotation of MPEG-7 descriptions
  - Basic datatypes: types provide a set of extended data types and mathematical structures such as vectors and matrices, which are needed by the DSs for describing AV content
  - Links & Media locations: localize pieces of content, and describe time, places, persons, individuals, groups, organizations, and other textual annotation
  - Basic tools

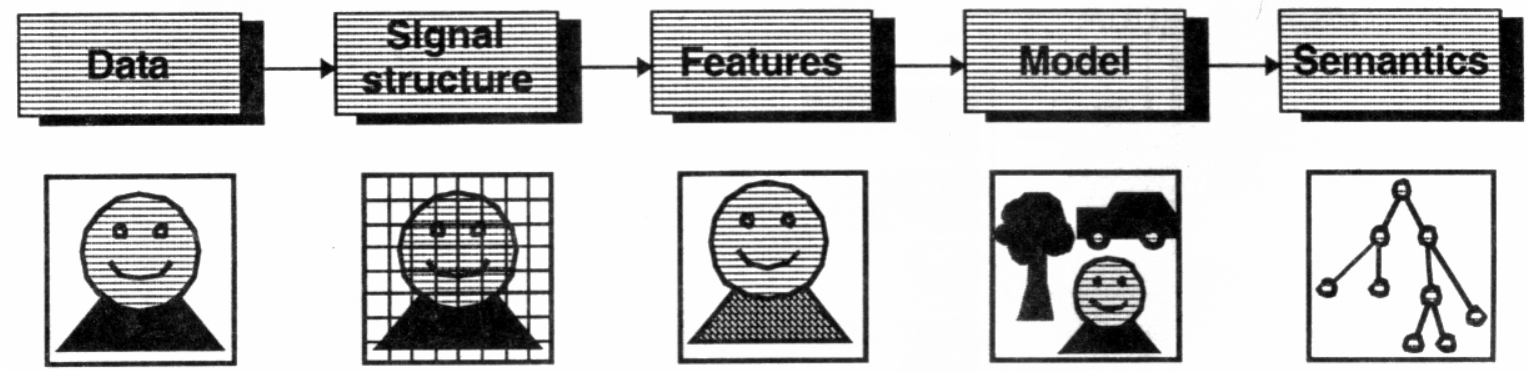
Example: MPEG-7 MDS Description

- MPEG-7 Video description: Structured Annotation DS

```
<Mpeg7 type="complete">
  <ContentDescription xsi:type="ContentEntityType">
    <MultimediaContent xsi:type="VideoType">
      <TextAnnotation>
        <StructuredAnnotation>
          <Who><Name> Sammy Sosa </Name></Who>
          <WhatObject><Name> Baseball </Name></WhatObject>
          <WhatAction><Name> Homerun </Name></WhatAction>
          <Where><Name> Chicago </Name></Where>
        </StructuredAnnotation>
      </TextAnnotation>
    </MultimediaContent>
  </ContentDescription>
</Mpeg7>
```

- Other DS's can be used to describe textual information such as title, author, producer, media format, data, address, etc.

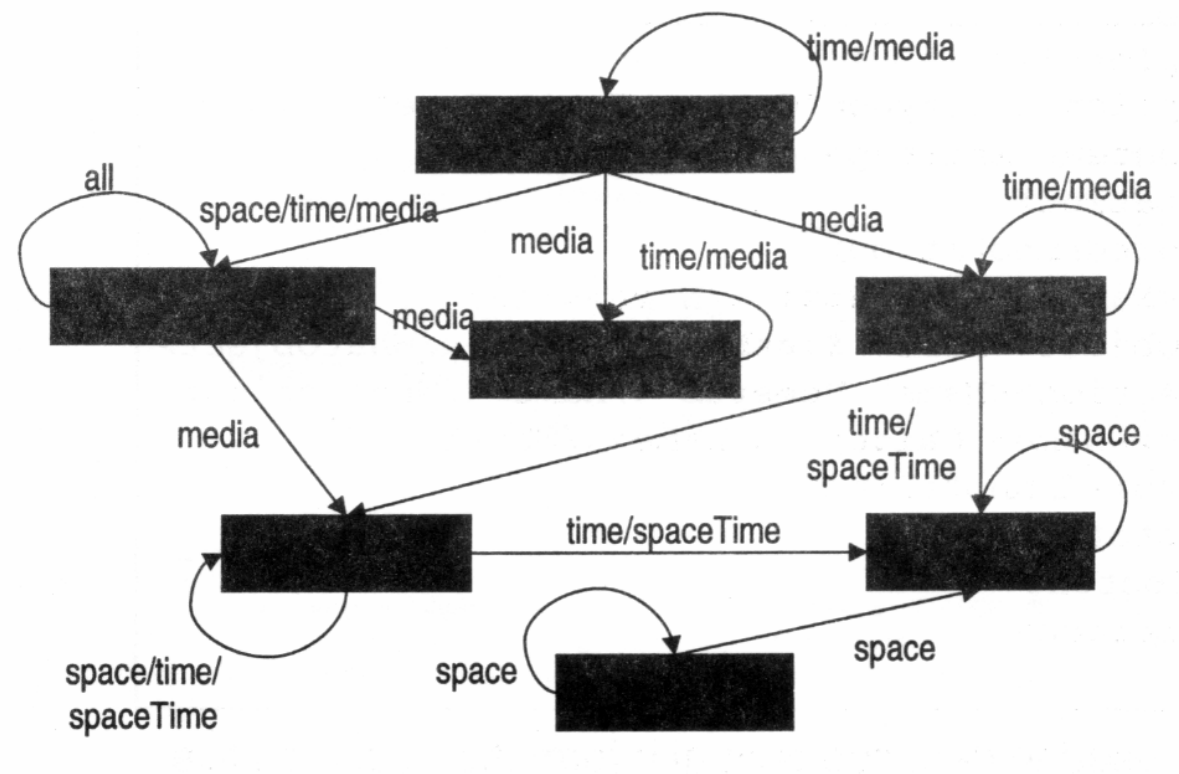
MPEG-7 for Content Description



Data	Structure	Features	Models	Semantics
Images	Regions	Color	Clusters	Objects
Video	Segments	Texture	Classes	Events
Audio	Grids	Shape	Analytic models	Actions
Multimedia	Mosaics	Motion	Probability	People
Formats	Relationships	Speech	Models	Labels
Layout	(Spatio-temporal)	Timbre	Classifiers	Relationships
		Melody		

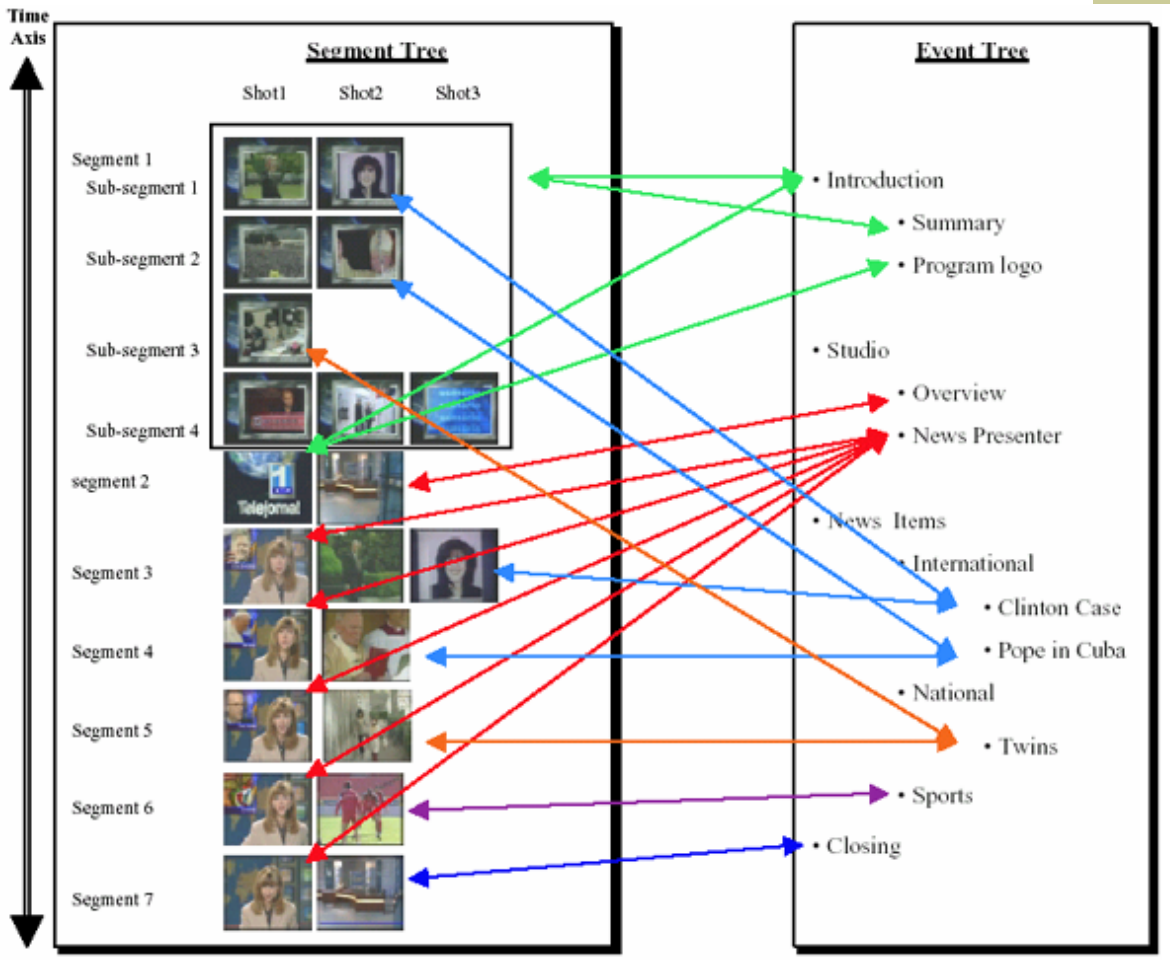
Audio-Visual Segment Decomposition

- Providing signal structure



- Spatial (image decomposition into regions)
- Temporal (video decomposition into segments)
- Spatio-temporal (video decomposition into moving regions)
- Media (video decomposition into audio and video tracks)

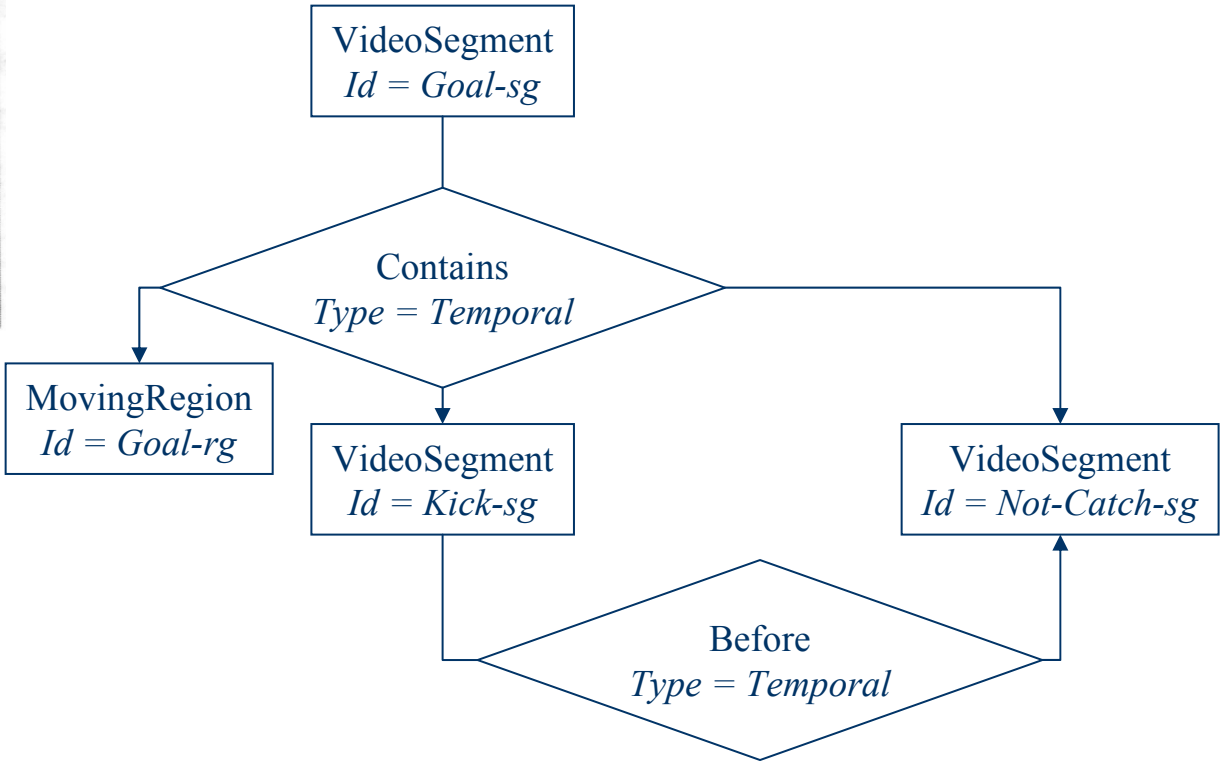
Video Segment



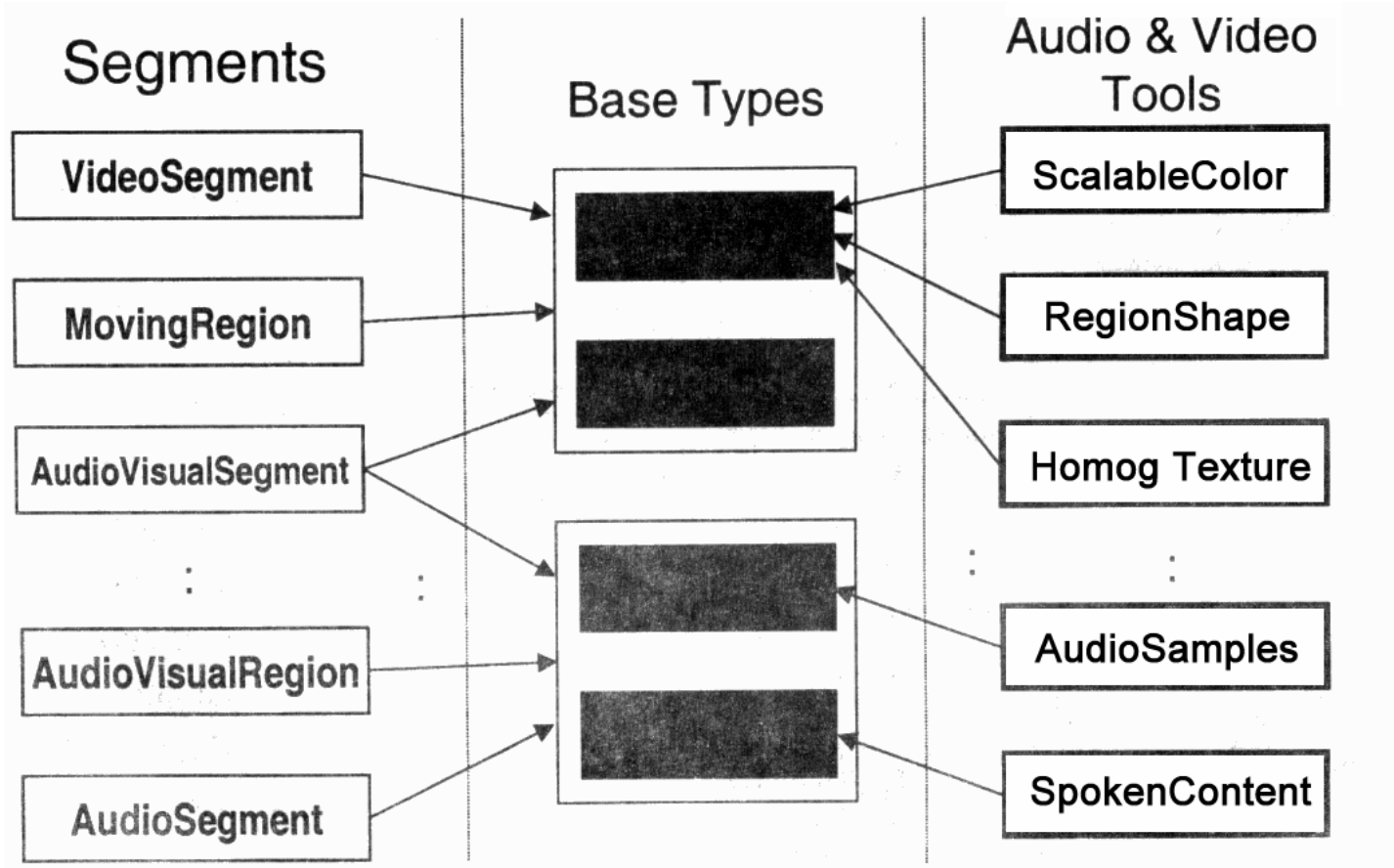
Video Segment Relationships

- Segment decomposition (video contains segments and regions)
- Temporal relations (kick before goal)

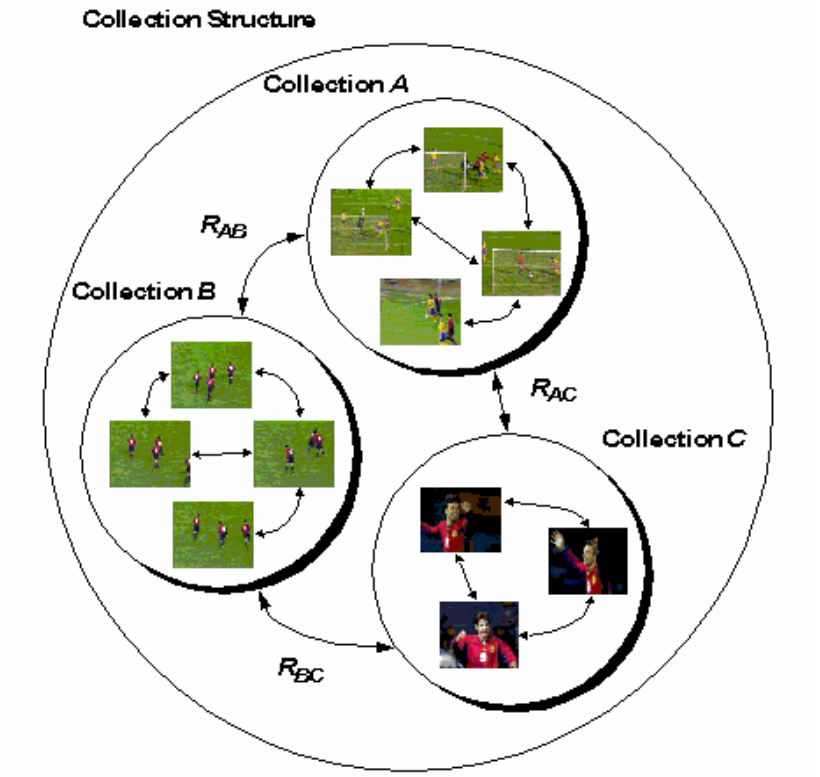
- Regions and Objects (players, goal, ball)
- Segments and events (kick, goal)



MDS Segment Interface

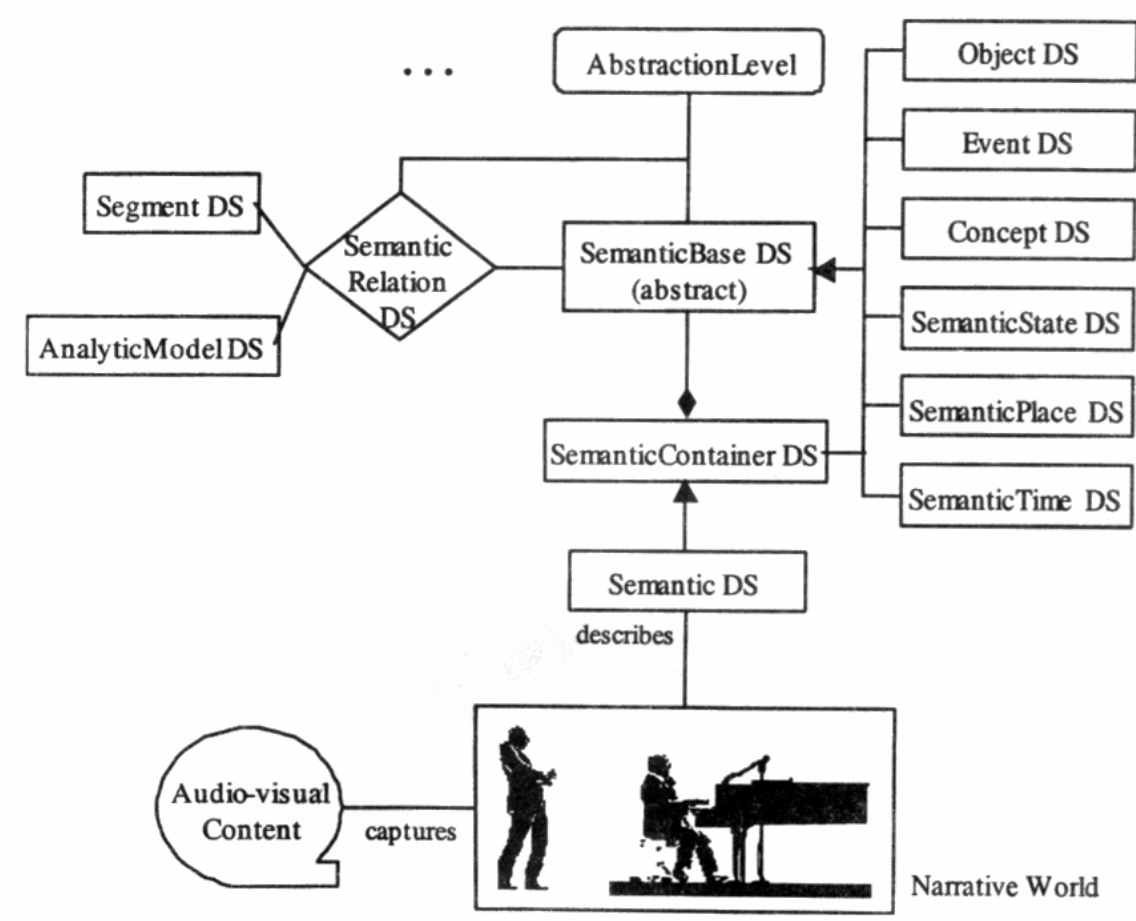


MPEG-7 MDS: Collections and Clusters



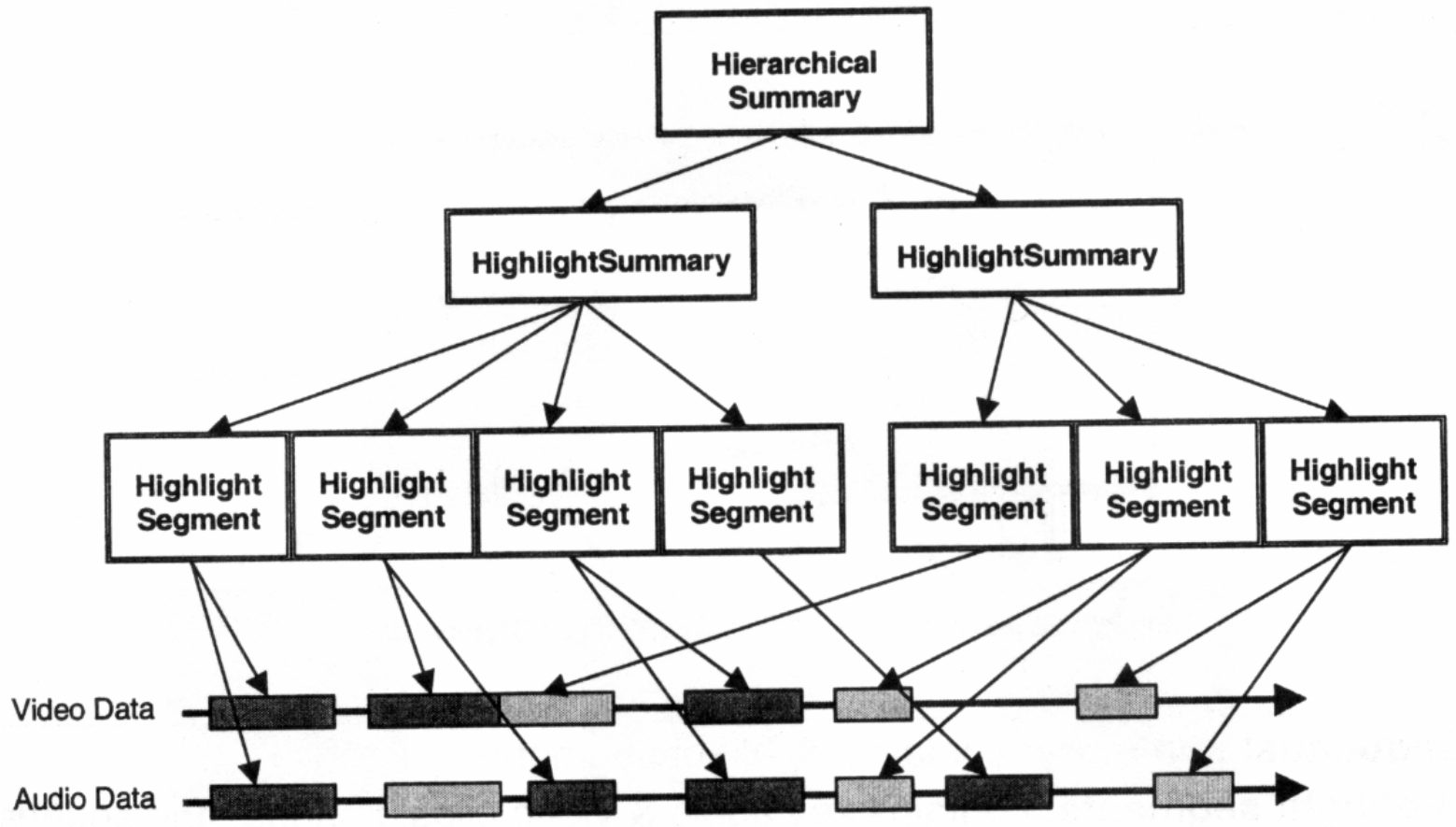
- Collection types:
  - Content collection
  - Segment collection
  - Descriptor collection
  - Concept collection
  - Mixed collection
- Nesting of collections (sub-collections)
- Relationships among collections (collection structure)

MPEG-7 MDS: Semantic DS

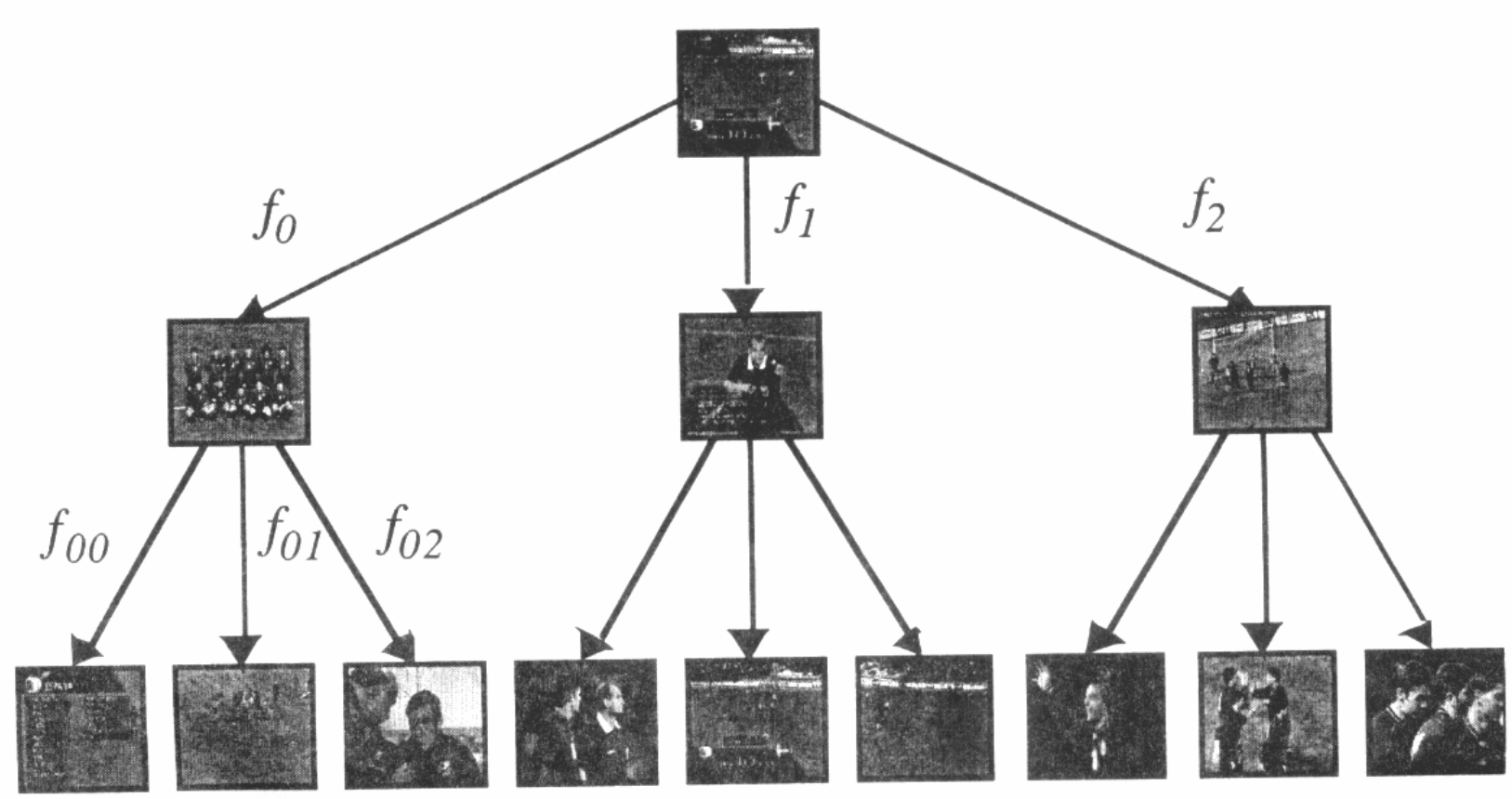


- **Semantic DS:** describes semantics of multimedia content
  - Objects
  - People
  - Events
  - Concepts
  - States
  - Places
  - Times

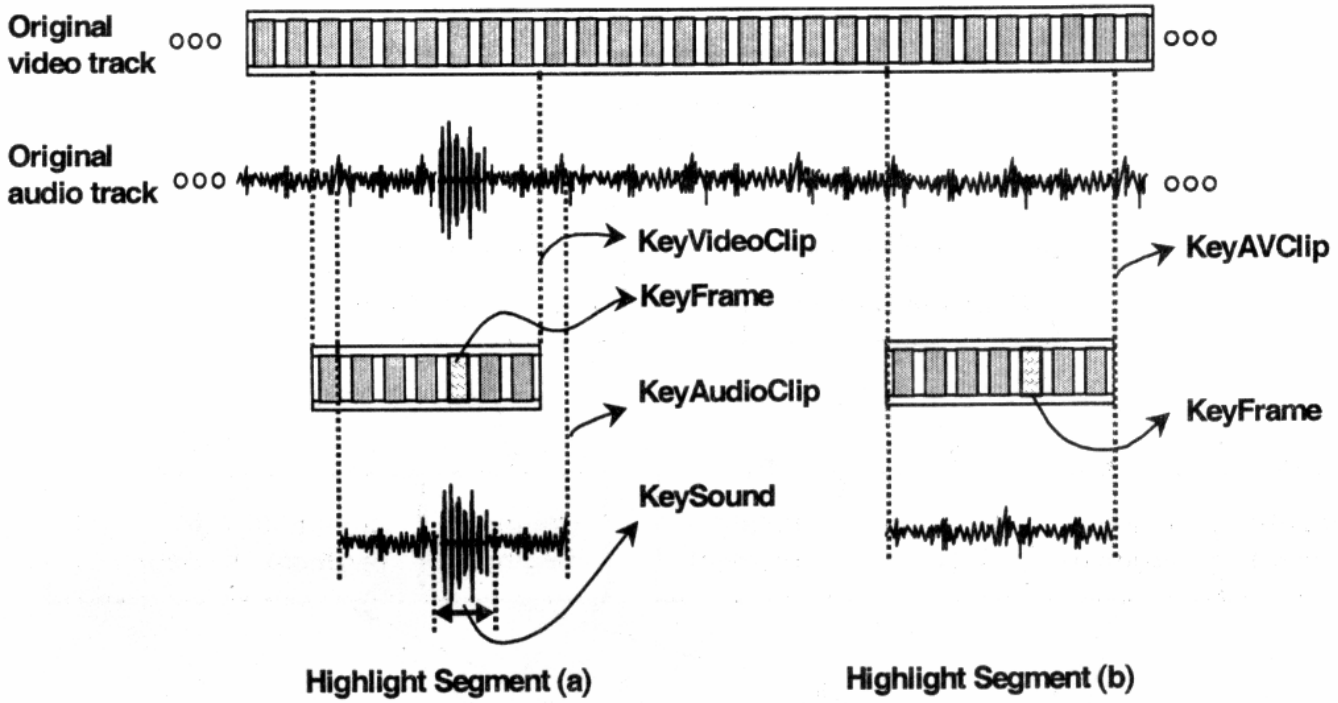
MPEG-7 MDS: Hierarchical Summary DS



Example of a Hierarchical summary of a video of a soccer game

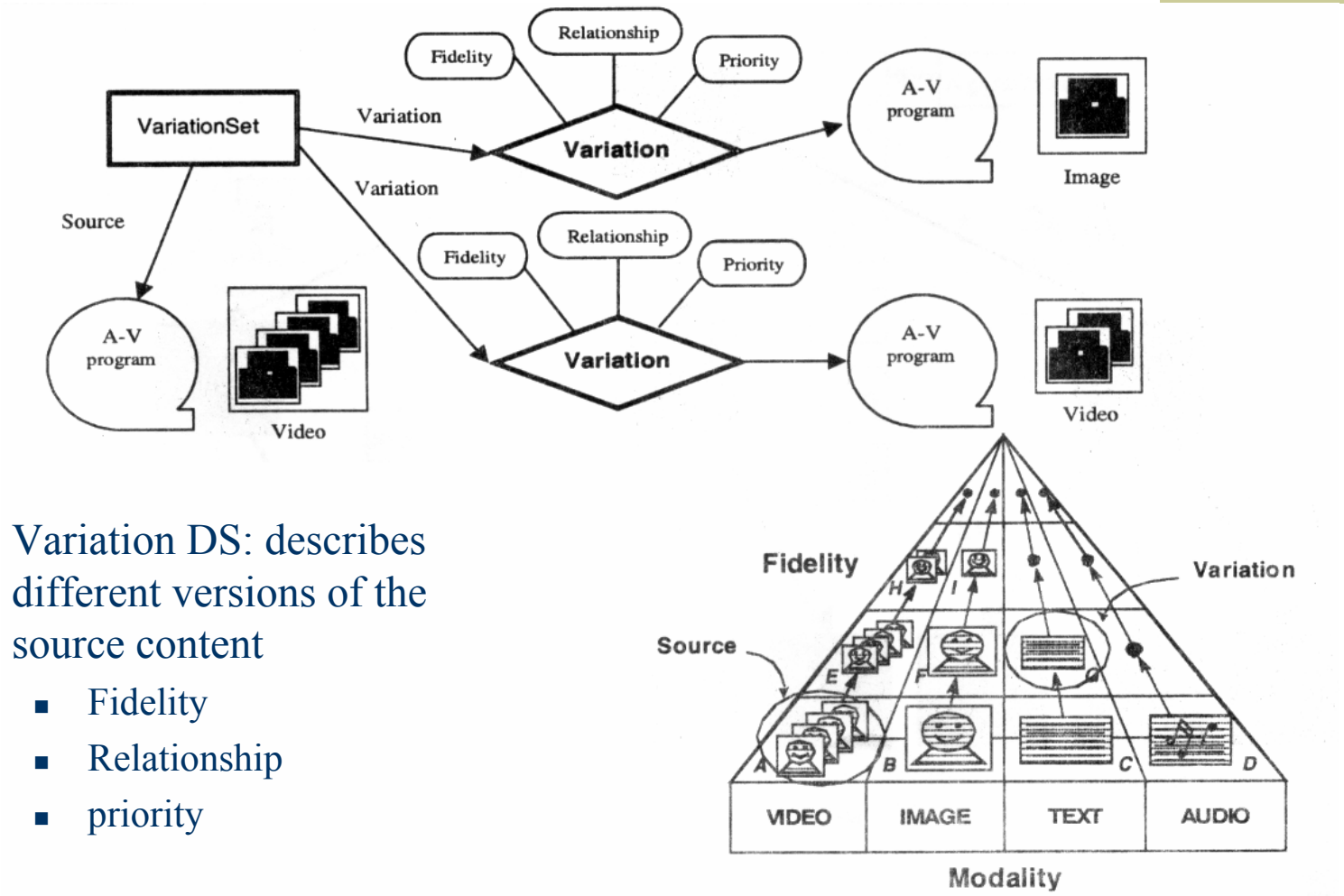


MPEG-7 MDS: Sequential Summary DS



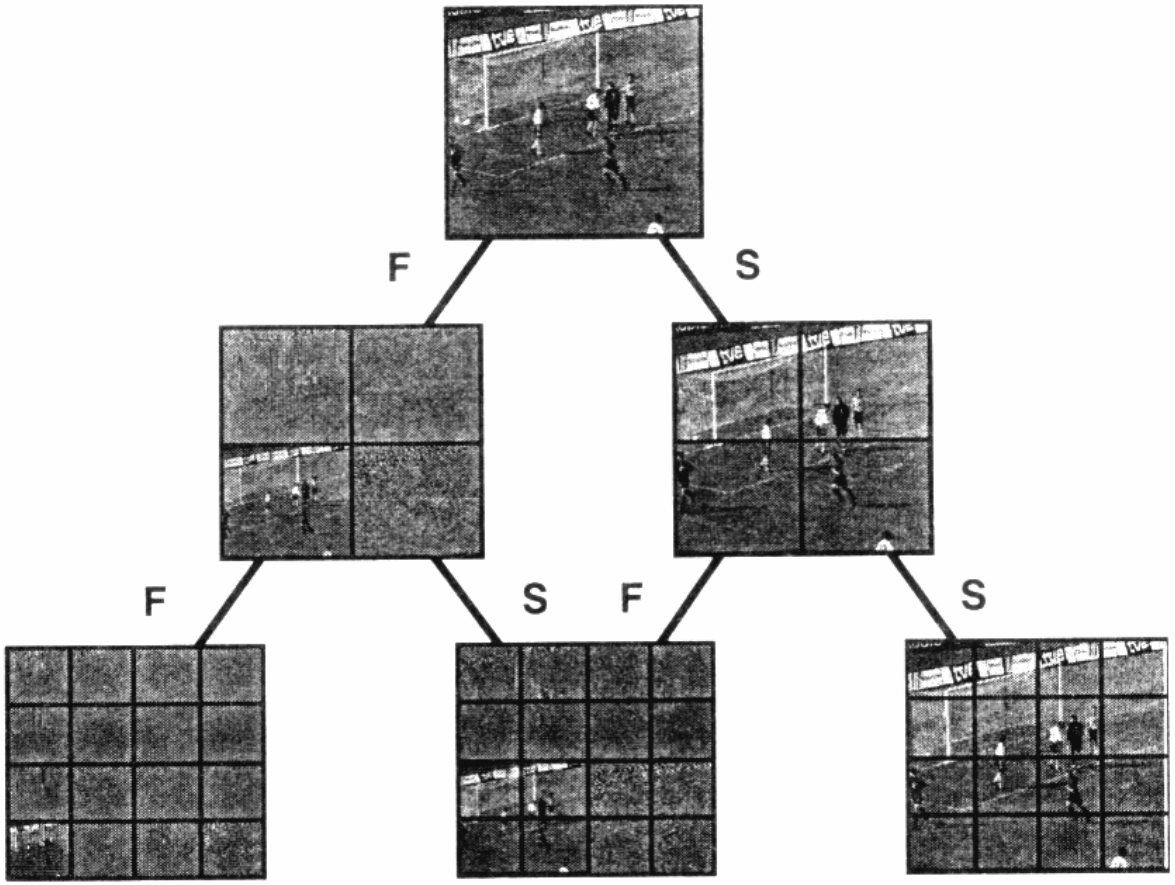
- Sequential summary: contains highlight segments
- Highlight segments: contains key audio & video clips, images, and sounds

MPEG-7 MDS: Variation DS



- Variation DS: describes different versions of the source content
  - Fidelity
  - Relationship
  - priority

Example of Partitions and Decompositions DS



Basic Structure of Visual Descriptors

- Grid Layout
  - Grid layout is a splitting of the image into a set of equally sized rectangular regions so that each region can be described separately
  - Each region can be described in terms of other descriptors such as color or texture
  - Descriptor allows to assign the sub-descriptors to all rectangular areas
- Time Series:
  - RegularTimeSeries: descriptors locate regularly within a given time span with constant intervals
  - IrregularTimeSeries: descriptors locate irregularly within a given time span
- 2D-3D Multiple View
  - Specifies a structure which combines 2D Descriptors representing a visual feature of a 3D object seen from different view angles
  - Any 2D descriptors such as contour-shape, region-shape, color or texture can be used
- Spatial 2D Coordinates
  - Defines a 2D spatial coordinate system and a unit to be used by reference in other Ds/DSs
  - Supports two kinds of coordinate systems: “local” and “integrated”
- Temporal Interpolation
  - Describes a temporal interpolation using connected polynomials
  - Can be used to approximate multi-dimensional variable values such as position of a moving object

MPEG-7 Visual Descriptors

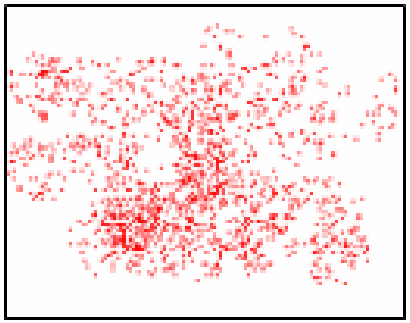
- Color
  - Color space, color quantization, dominant colors, scalable color
  - Color layout, color-structure, GoF/GoP color
- Texture
  - Homogeneous texture, edge histogram, texture browsing
- Shape
  - Region-based, contour-based, shape 3D
- Motion
  - Camera motion, motion activity, motion trajectory, parametric motion
- Localization
  - Region locator, spatio-temporal locator
- Face
  - Face recognition

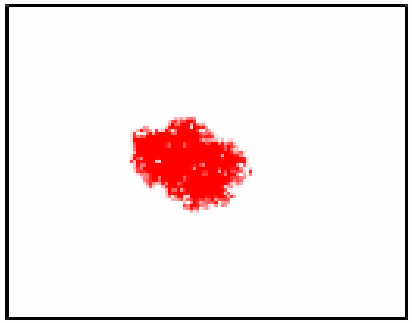
MPEG-7 Color Descriptors (1)

- Color Descriptors
  - Color space
    - Default color spaces
      - RGB for computer
      - YCrCb for video
    - Color spaces better for search
      - HSV for scalable color
      - HMMD for color structure
    - Monochrome: allow universal use in the context with descriptors
    - Linear transformation: support for additional color spaces
  - Color quantization
    - Defines a uniform quantization of a color space
    - Combined with other descriptors (dominant color) to express the meaning of the values within a color histogram

MPEG-7 Color Descriptors (2)

- Dominant color(s)
  - Useful when a small number of colors are enough to characterize a local object or an image
  - Color quantization is used to extract a small number of representing colors each region or image
  - Percentage of each quantized color in the region is calculated accordingly
  - A spatial coherency on the entire descriptor is defined and is used in similarity retrieval





MPEG-7 Color Descriptors (3)

- Scalable color
  - A color histogram in HSV color space
  - Histogram values are encoded using Haar transform
  - Binary representation is scalable in terms of number of bins used and in the number of bits per bin over a wide range of data rates
  - Number of bins can range from 16 to 256
  - Matching may be performed in either Haar or histogram domain

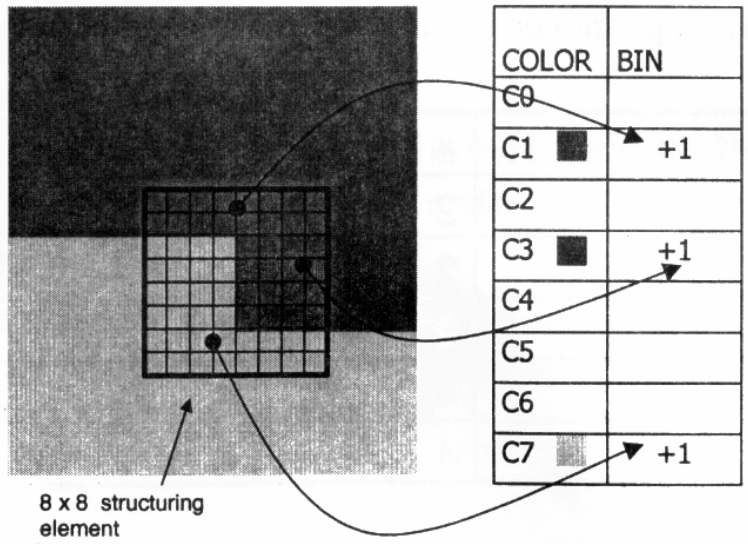
No. coeff	# bins: H	# bins: S	# bins: V
16	4	2	2
32	8	2	2
64	8	2	4
128	8	4	4
256	16	4	4

MPEG-7 Color Descriptors (4)

- Color layout
  - Effectively represents spatial distribution of colors in an image
  - Very compact representation (63 bits)
  - Uses YCbCr color space
    - Partition the image into (w/8) x (h/8) subimages
    - Calculate the “dominant color” of each sub-image
    - Compute the DCT of this 8x8 matrix of dominant colors
    - Quantize the DCT coefficients
    - Zig-zag scan of quantized DCT coefficients
  - Advantages
    - No dependency on image/video format, resolutions and bit depths
    - Small software and hardware requirement (8 bytes per image)
    - Use of frequency domain features allows users to introduce perceptual sensitivity of human vision system for similarity calculation
    - Support scalable application of the feature by controlling the number of coefficients enclosed in the descriptor

MPEG-7 Color Descriptors (5)

- Color structure
  - Uses HMMD color space quantized to 32, 64, 128 and 180 bins
  - Similar to scalable histogram, but a 8x8 structuring elements is used to compute the bin values (rather than characterizing relative frequency of single image samples, it characterized relative frequency of structuring elements of a particular color)



MPEG-7 Color Descriptors (6)

- GoF/GoP color
  - GoF/GoP: Group of Frame/Picture color descriptor extends the Scalable Color descriptor
  - Additional 2 bits allows to define how the color histogram was calculated before Haar transform is applied to it:: by average, median or intersection
  - Average histogram: averages the counter value of each bin across all frames or pictures, it is equivalent to computing the aggregate color histogram of all frames and pictures with proper normalization
  - Median histogram computes the median of the counter value of each bin across all frames or pictures, it is more robust to round-off errors and the presence of outliers in image intensity values compared to the average histogram
  - Intersection histogram to computes the minimum of the counter value of each bin across all frames or pictures to capture the “least common” color traits of a group of images

MPEG-7 Texture Descriptors (1)

- Texture Descriptors
  - Texture Browsing
    - Referred to as Perceptual Browsing Components (PBC)
    - Extracted using image convolutions with a set of masks
      - 2-D Gabor wavelet decomposition (24 filters, 6 orientations at 4 scales)
      - Convolutions can be efficiently implemented in hardware and software
    - Use filter output to characterize texture using 3 perceptual dimensions
      - Regularity, coarseness, directionality (12 bits total –compact)
  - Homogenous Texture
    - Referred to as Similarity Retrieval Components (SRC)
    - Extracted using image convolutions with a set of masks
      - Filters tuned to detect image features as different scales and orientations
      - 30 Gabor filters (6 orientations at 5 scales)
    - Descriptor specifies the normalized mean and standard deviation of these filtered outputs (computed in the frequency domain)

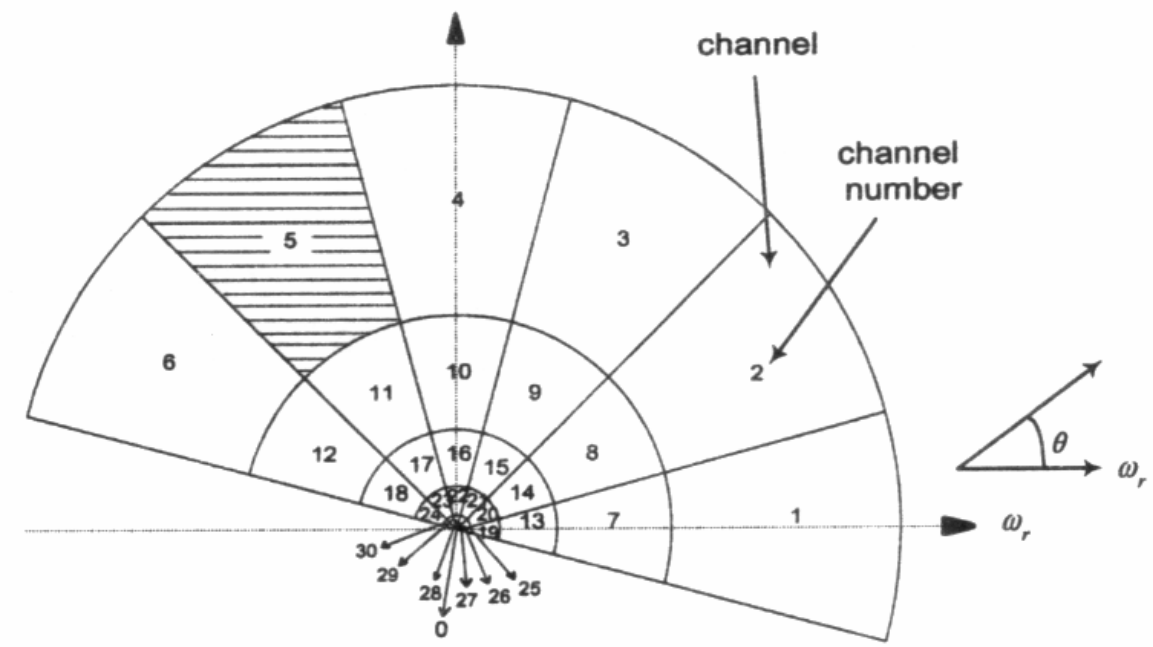
MPEG-7 Texture Descriptors (2)

- Descriptor Definition

$$T = [v_1\ v_2\ v_3\ v_4\ v_5\ \mu_1\ \sigma_1\ \mu_2\ \sigma_2\ \ldots\ldots\ \mu_M\ \sigma_M]$$

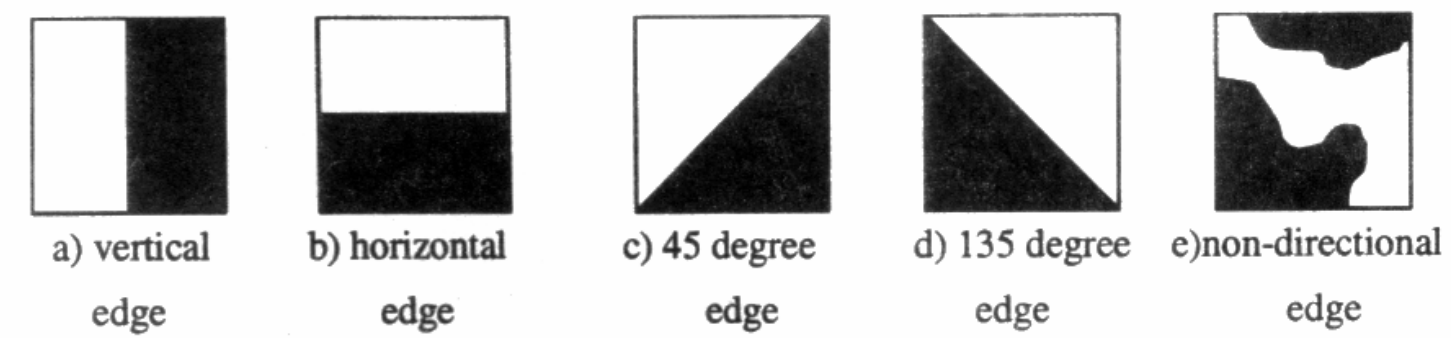
PBC SRC

- Frequency Layout



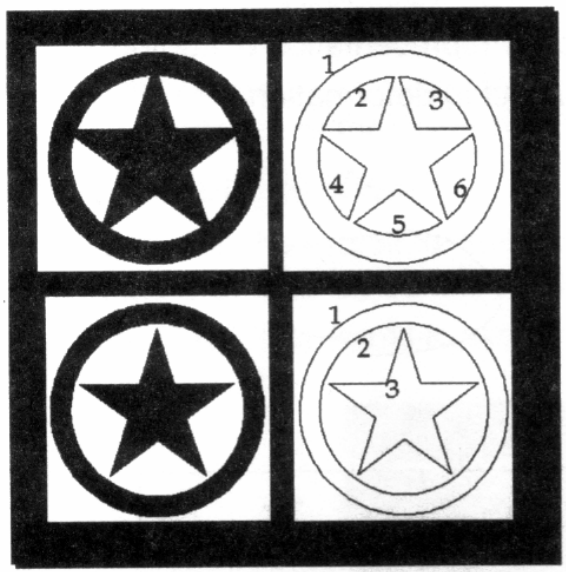
MPEG-7 Texture Descriptors (3)

- Edge Histogram
  - Describes spatial distribution of four edges and one non-directional edge in the image
  - Comparison of above edge types in 3 levels of localization: global, semi-global and local
  - Useful for indexing and retrieval of natural images with subjective objects due to information of edge distribution

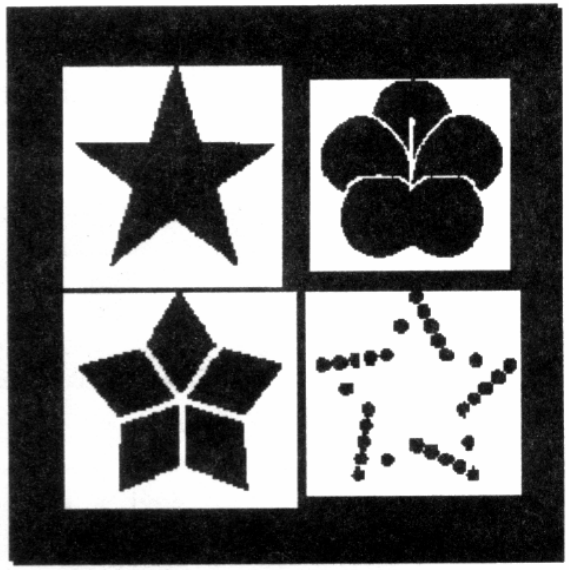


MPEG-7 Shape Descriptors (1)

- Shape Descriptors



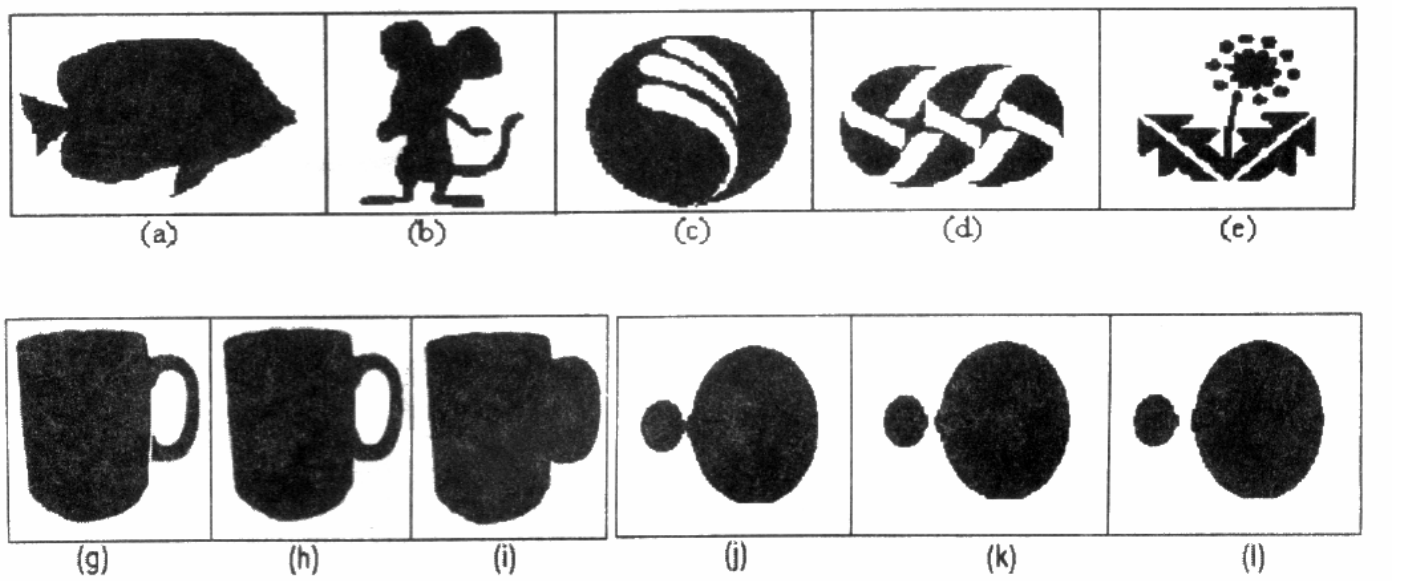
Contour-based
shape descriptor



Region-based
shape descriptor

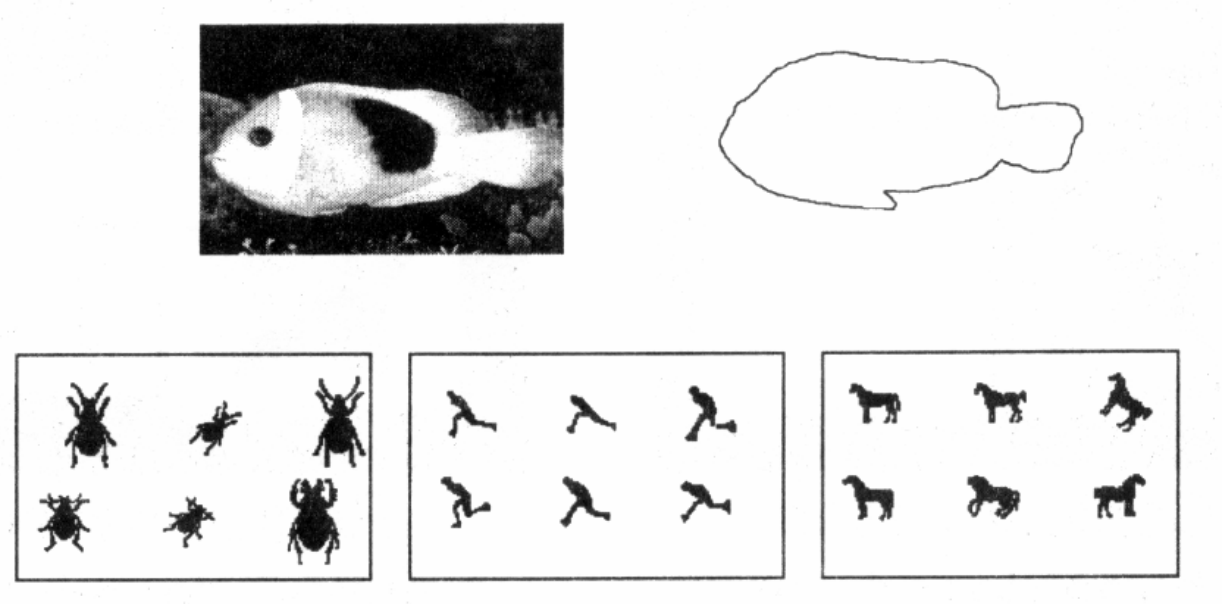
MPEG-7 Shape Descriptors (2)

- Region-based shape
  - An object may consist of either a single region or a set of regions with some holes, region-based shape is used to describe these kind of shapes
  - Robust to minor deformation along the boundary (g, h same, but l different)
  - Small size, fast extraction time and matching



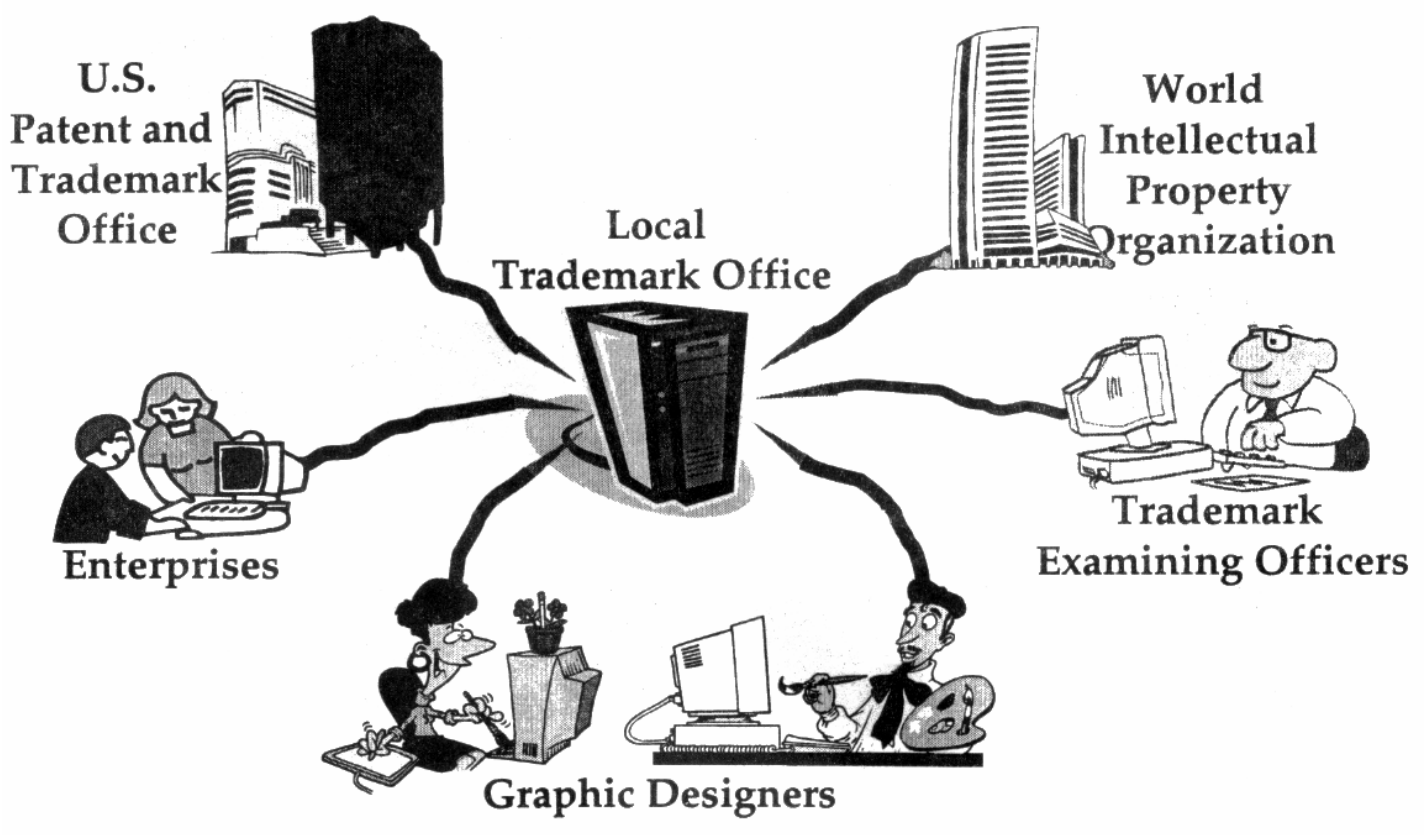
MPEG-7 Shape Descriptors (3)

- Contour-based shape
  - Captures contour characteristics of an object boundary
  - It is based on the curvature scale space representation of contour
  - Performs very well when shat is subject to deformation



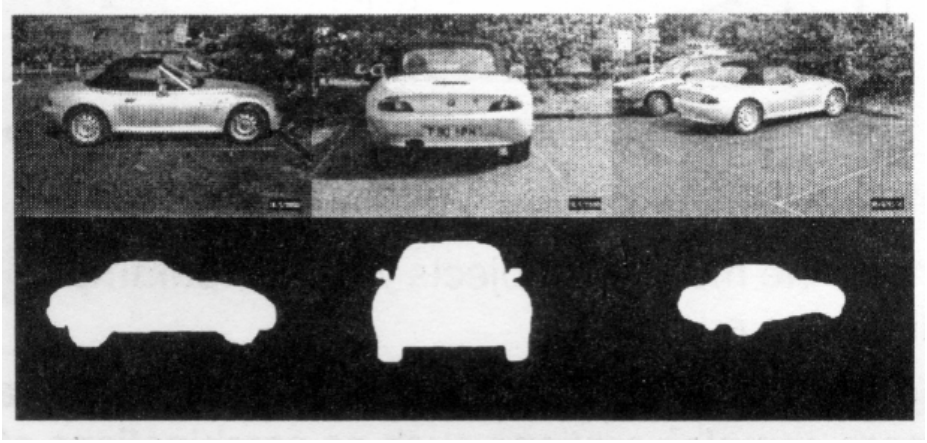
MPEG-7 Shape Descriptors (4)

- Example Application: Trademark Registration



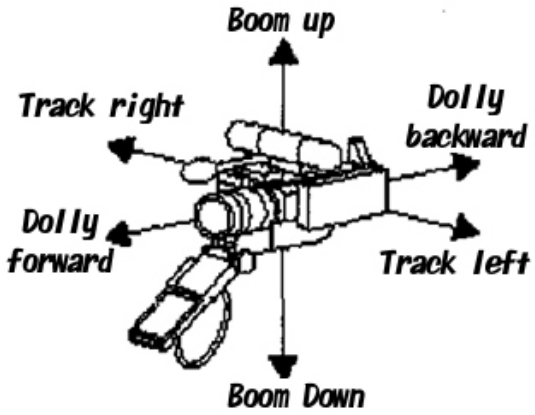
MPEG-7 Shape Descriptors (5)

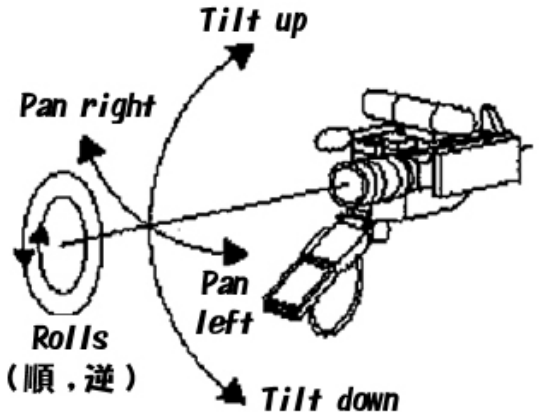
- 3D shape
  - Descriptor of 3D mesh models that exploits local attributes of the 3D surface (principal curvature) at each point or shape index
  - A shape spectrum is produced, which is a histogram for shape indices calculated over the entire mesh
- 2D-3D Multiple View
  - Specifies the syntax to which combines 2D descriptors representing a visual feature of a 3D object seen from different view angles
  - The descriptor forms a complete 3D view-based representation of the object



MPEG-7 Motion Descriptor (1)

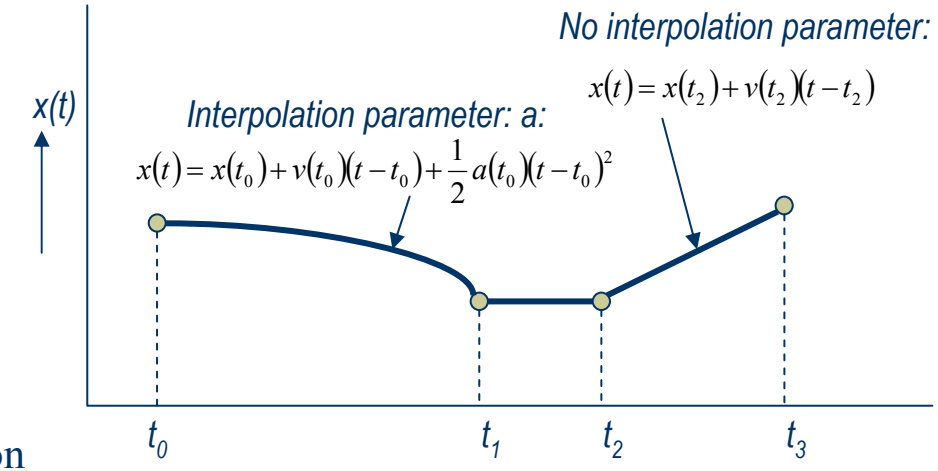
- Motion descriptors
  - Camera motion
    - Describes the global motion induced by the camera for given video
    - It supports the following basic camera operations: fixed, panning (horizontal rotation), tracking (horizontal transverse movement), zooming, dollying (translation along the optical axis), and rolling





MPEG-7 Motion Descriptor (2)

- Motion Trajectory
  - Used to define the localization, time, and space of an object
  - Consists of a set of key points along with a set of optional interpolating functions used to describe the path of the object between two key points



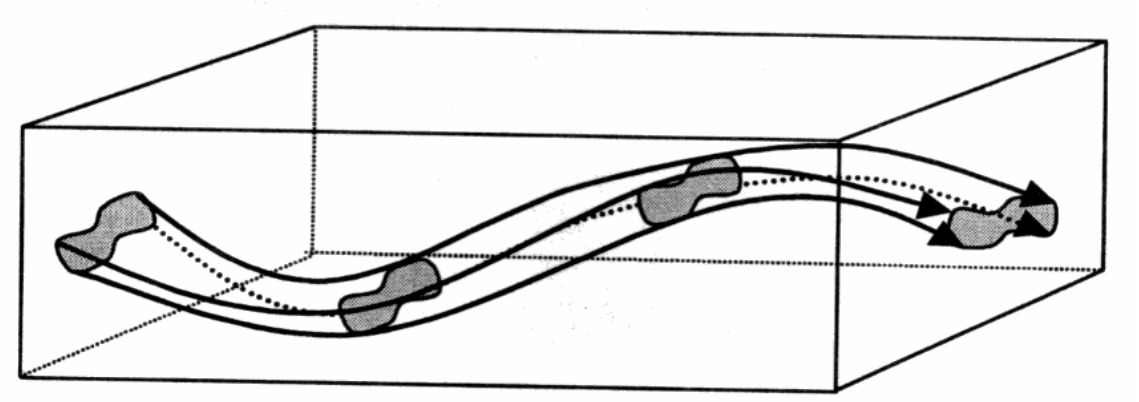
- Parametric motion
  - Describes the motion of objects with 2-D parametric models such as for global motion and spite
  - Can be used to efficiently describe the simple translation, rotation, zooming and more complex motions such as combinations of above types of motion

MPEG-7 Motion Descriptor (3)

- Motion Activity
  - Able to express the activity of a given video/shot as a slow sequence, fast paced sequence, action sequence, etc…
  - Can be applied to fast-browsing, dynamic video summarization, content-based querying etc
  - Very simple and only contains four attributes:
    - Intensity – high value indicates high activity, while low value of indicates low activity
    - Direction – expresses dominant direction of the activity, if any
    - Spatial Distribution – indicates whether the activity is spread across many regions or restricted to one large region. It is an indication of the number and size of "active" regions in a frame
    - Temporal Distribution – expresses the variation of activity over the duration of the video segment/shot

MPEG-7 Localization Descriptors

- 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, such as moving object regions
  - Provides localization functionality for such regions



MPEG-7 Face Descriptor

- Face Descriptor
  - Represents the projection of a face vector onto a set of 49 eigenvectors, commonly referred to as eigenfaces
    - Obtain normalized face image (46x56, luma only = 2576 elements)
    - Arange pixels into 1D face vector (Λ) via column-wise raster scan
    - Project 1D face vector onto space defined by normative basis function (U), where Ψ is the mean face vector; normatively specified

$$W = U^T \left(\Lambda - \Psi \right)$$

    - 49 features are the normalized, clipped and linearly quantized

MPEG-7 Visual Extensions (1)

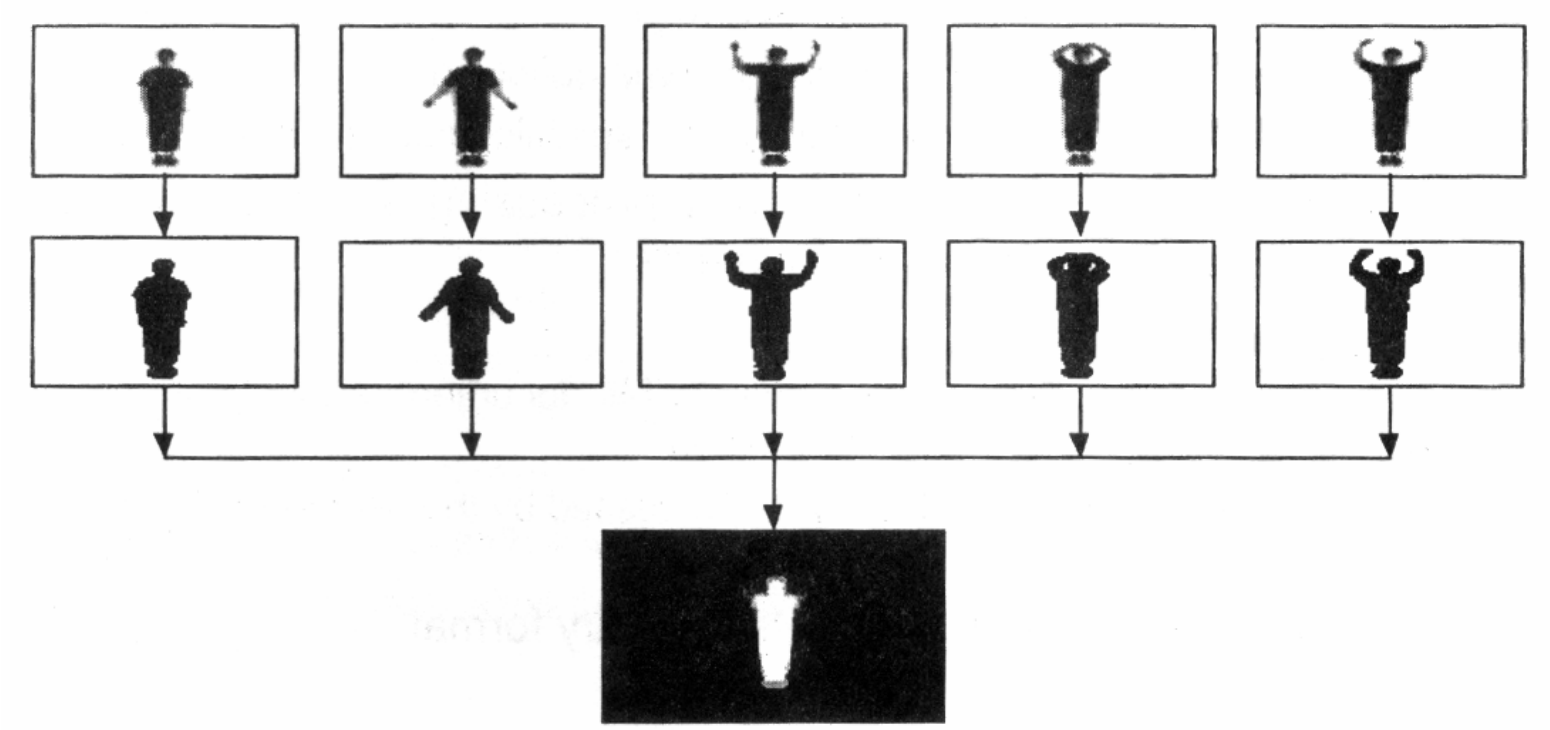
- Color Temperature
  - This descriptor specifies the perceptual temperature feeling of illumination color in an image for browsing and display preference control purposes
  - Four perceptual temperature browsing categories are provided
    - Hot: 1667K ~ 2250K
    - Warm: 2251K ~ 4170K
    - Moderate: 4171K ~ 8060K
    - Cool: 8061K ~ 25000K
  - The range of color temperature values for each category is the uniformly divided into 64 sub-ranges for finer classification

MPEG-7 Visual Extensions (2)

- Shape sequence
  - The descriptor can describe shape variations in terms of Shape Variation Map and the statistics of the region shape description of each binary shape image in the collection
  - Shape Variation Map consists of Static Shape Variation and Dynamic Shape Variation
    - Static Shape Variation corresponds to 35 quantized ART coefficients on a 2-dimentional histogram of group of shape images whereas
    - Dynamic Shape Variation does to the inverse of the histogram except the background
  - For the statistics, a set of standard deviations of 35 coefficients of the Region Shape descriptor, which is defined in ISO/IEC 15938-3 are used
  - Its main functionality is to retrieve similar collection of shape images regardless of their sequence or the number of frames in each collection
  - When they are in sequences, it can be used for retrieval of shape sequence represented by a set of frames in video segments – in terms of the similar shape variation due to the object movement

MPEG-7 Visual Extensions (3)

- Process of generation Shape Variation Map



MPEG-7 Description Definition Language

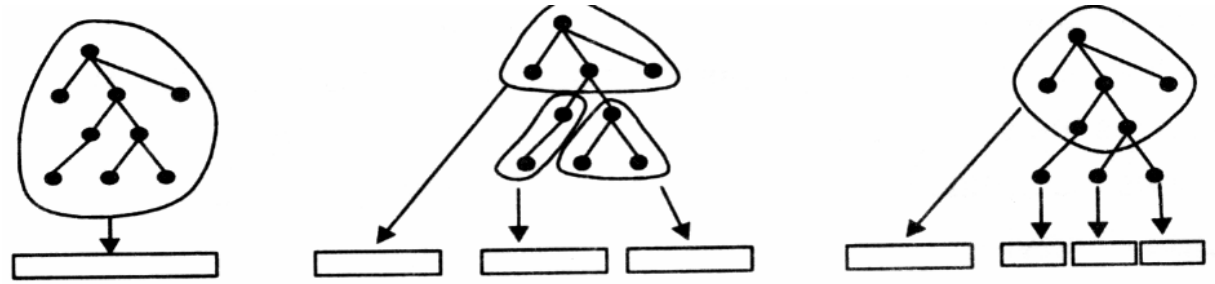
- Summary
  - DDL is a language that allows the creation of new Description Schemes (DSs) and Descriptors (Ds)
  - DDL is a XML schema language based on the XML by W3C (May 2001)
  - DDL provides the solid descriptive foundation to create DSs and Ds
  - DDL defines the syntactic rules to express and combine DSs and Ds
- XML Schema Overview
  - XML Schema Structures
    - Primary components: schema- wrapper round the definitions and declarations, simple type definitions, complex definitions, attribute definitions, element declarations
    - Secondary components: attribute group definitions, identity definitions, named group definition and notation declaration
    - Third components: substitution groups, annotations, wildcards
  - XML Schema Datatype
    - A set of built-in primitive and derived datatypes and mechanism to define own datatypes
  - MPEG-7 Extensions to XML Schema
    - Array and matrix dattypes both fixed size and parametrerized size
    - Built-in primitive time datatypes; basic TimePoint and basic Duration

Binary Format for MPEG-7 (BiM)

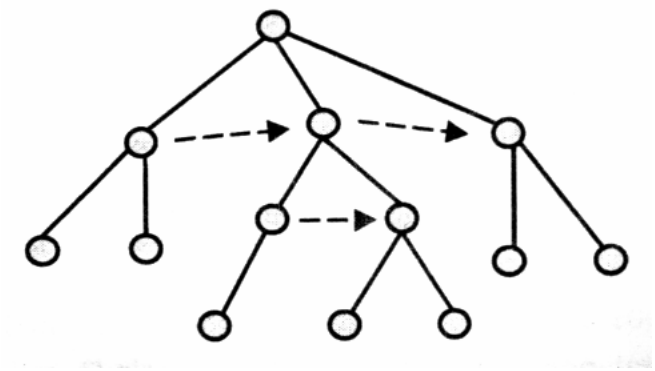
- BiM is designed for transmission and storage
- BiM encoders and decoders can deal with any XML language
- BiM is a schema oriented encoding scheme
  - The binary format is deduced from the schema definition
  - There is no need to define coding tables or a specific encoding mechanism
- BiM is a pre-parsed format, very few extra processing is needed on the decoder to validate the received documents
- BiM is a typed binary format
  - At the encoder side, every components of XML document (attribute, elements, leaf notes) is mapped to a type information
  - The typed binary format can directly be processed by the decoder without string conversion
- BiM is backward and forward compatible binary format

Binary Format for MPEG-7 (BiM)

- BiM allow a parameterized transmission of XML document



- BiM decoder can be efficiently processed at binary level



eXperimentation Model: XM (1)

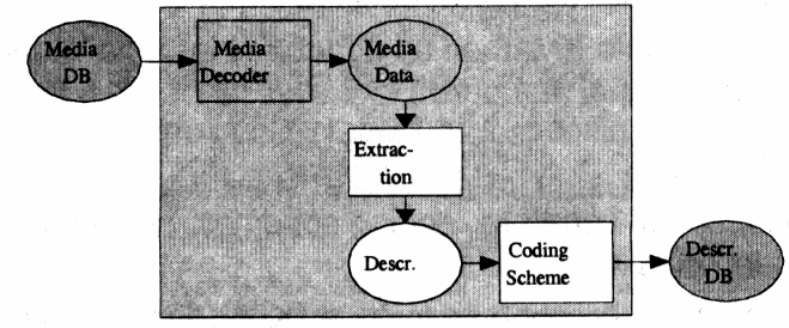
- XM is the reference model for the MPEG-7 standard
  - Same purpose as the Test Model (TM) in MPEG-2
  - Confirm the stability of the standard; basis to conduct core experiments
  - Technology first goes into XM, then as it matures into the Working Draft (WD) and ultimately into the final standard
- Normative Components:
  - Descriptors, description schemes, coding schemes, description definition language, system tools, etc.
  - In MPEG-2, normative components include DCT, 2D translational motion vectors, 16x16 macroblock, VLC tables
- Non-Normative Components:
  - Feature extraction, search engine, segmentation, etc.
  - In MPEG-2, non-normative tools include motion estimation, rate control, MD mode decision.

eXperimentation Model: XM (2)

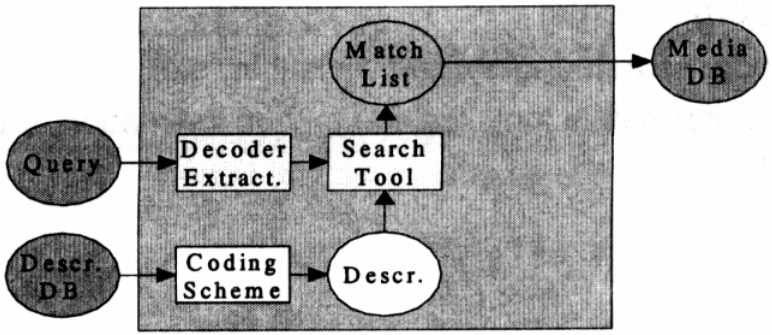
- Extraction vs. Client applications
  - Applications are related to one particular descriptor or description scheme
  - Extraction applications: creating descriptors or description schemes
  - Client applications: using descriptors or description schemes under test
- Modularity of the XM software
  - Compiling the complete framework (all Ds and DSs) into one program results in an executable of more than 100 Mbytes of size – too big
  - All applications are built from modules to support
    - Partial compilation to allow to use only one signal D or DS
    - Combination of a subset of Ds or DSs
    - Reuse of code, for example, a DS is built in an hierarchical way from other Ds and DSs
  - Modules include:
    - Media decoder class
    - Multimedia data class
    - Extraction tool class (only for extraction applications)
    - Descriptor class
    - Coding Scheme class, and
    - The search tool class (only for client applications)

eXperimentation Model: XM (3)

- Applications types in XM software
  - Extraction from Media: description is extracted from the media input data

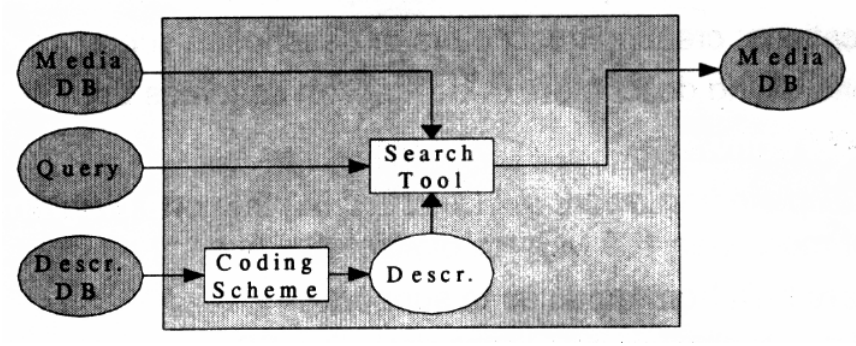


  - Search & Retrieval application: a sorted database is created from Ds and query

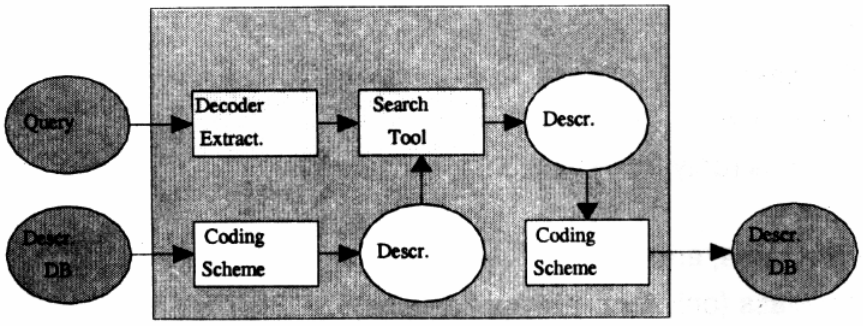


eXperimentation Model: XM (4)

- Media Transcoding Application: a transcoded media database is created from the original media database, corresponding descriptions and query

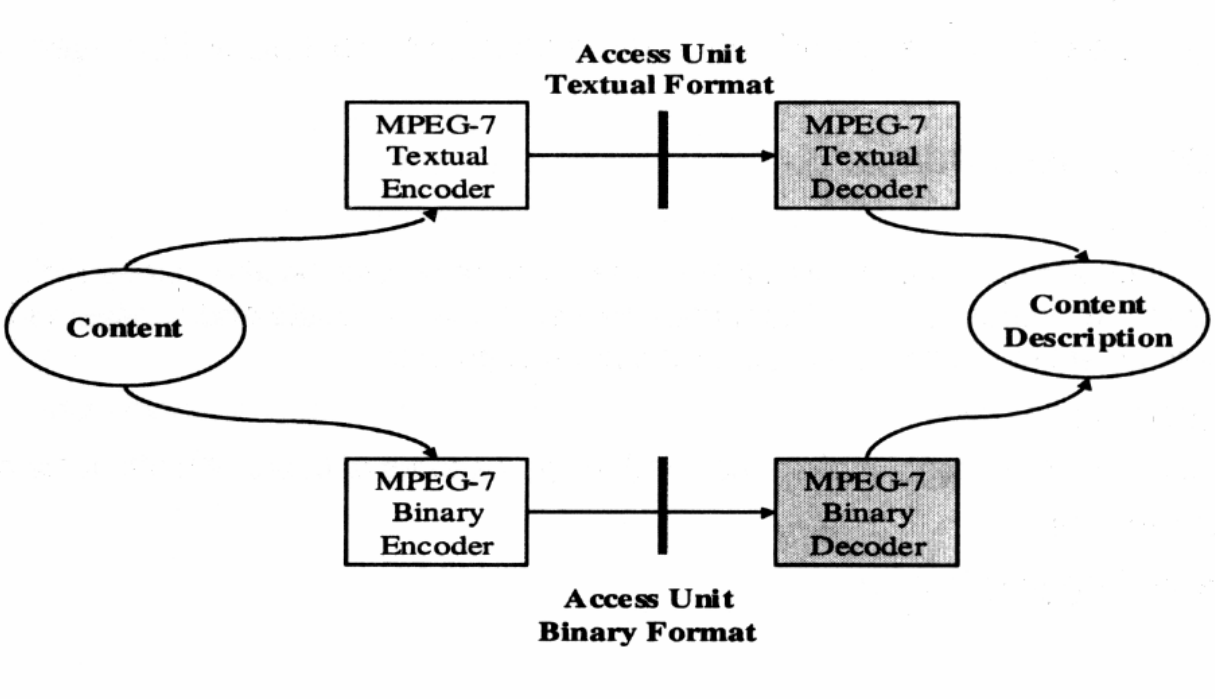


- Description filtering Applications: a description is created, filtering the input description with respect to a query



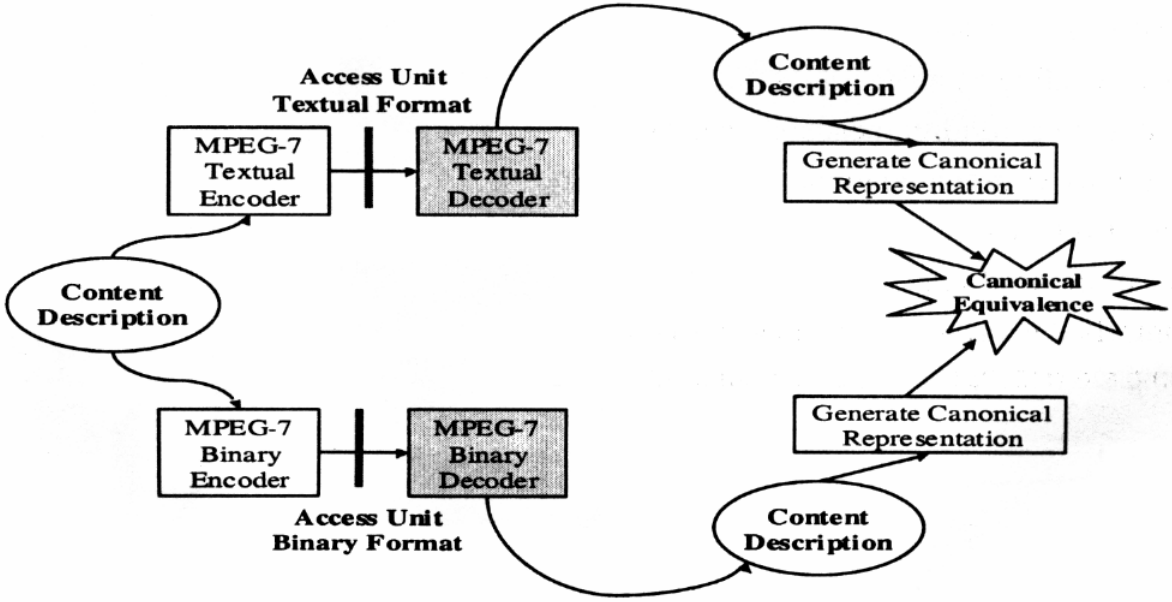
MPEG-7 Conformation Testing (1)

- Normative interface: two normative interfaces
  - Texture format interface: describes the format for the textual access units
  - Binary format interface: describes the format of the binary access units



MPEG-7 Conformation Testing (2)

- Validation of the standard
  - Describes how to prove the equivalence of two representations, texture and binary
  - Two canonical descriptions are generated from the reconstructed content descriptions, they should be equivalent



MPEG-7 Profiles and Levels (1)

- Profile
  - It is a set of tools providing a set of functionalities for a certain class of applications
  - A new profile should be defined if it provides a significantly different set of functionalities
  - It is no yet clear how many and which dimensions of profiling will be used for MPEG-7
- Profiles under consideration
  - Simple profile
    - Application areas include limited textual metadata (e.g., Title, Author, Copyright, Abstract, Keywords, URL or Asset Identifier, etc.) used to locate and subsequently access an entire multimedia asset or temporal segments
  - User description profile
    - To describe the personal preferences and usage patterns of users of multimedia content
  - Summary profile
  - Audiovisual Logging Profile

MPEG-7 Profiles and Levels (2)

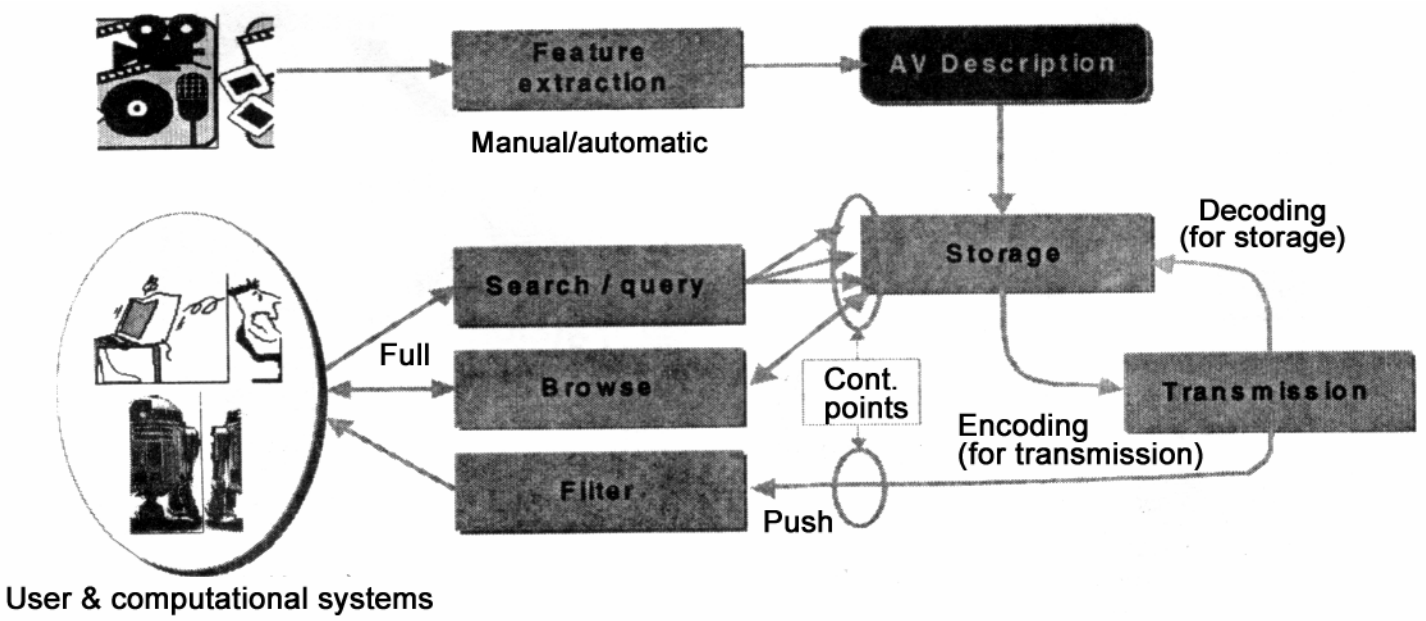
- Levels
  - Levels limit the complexity associated to a certain profile @ level
  - A new level should be defined if it is associated to a significantly different implementation complexity
- Levels may reflect tradeoffs in complexity. Possible dimensions of complexity are:
  - Access unit us none
  - BiM encoding vs none
  - Linking (e.g. id/idref, headers) vs none
  - Hierarchy vs graph relations
  - Size of schema
  - Size of descriptions
  - Controlled vocabulary vs unconstrained
  - Template descriptions vs unconstrained

MEPG-7 Application Areas

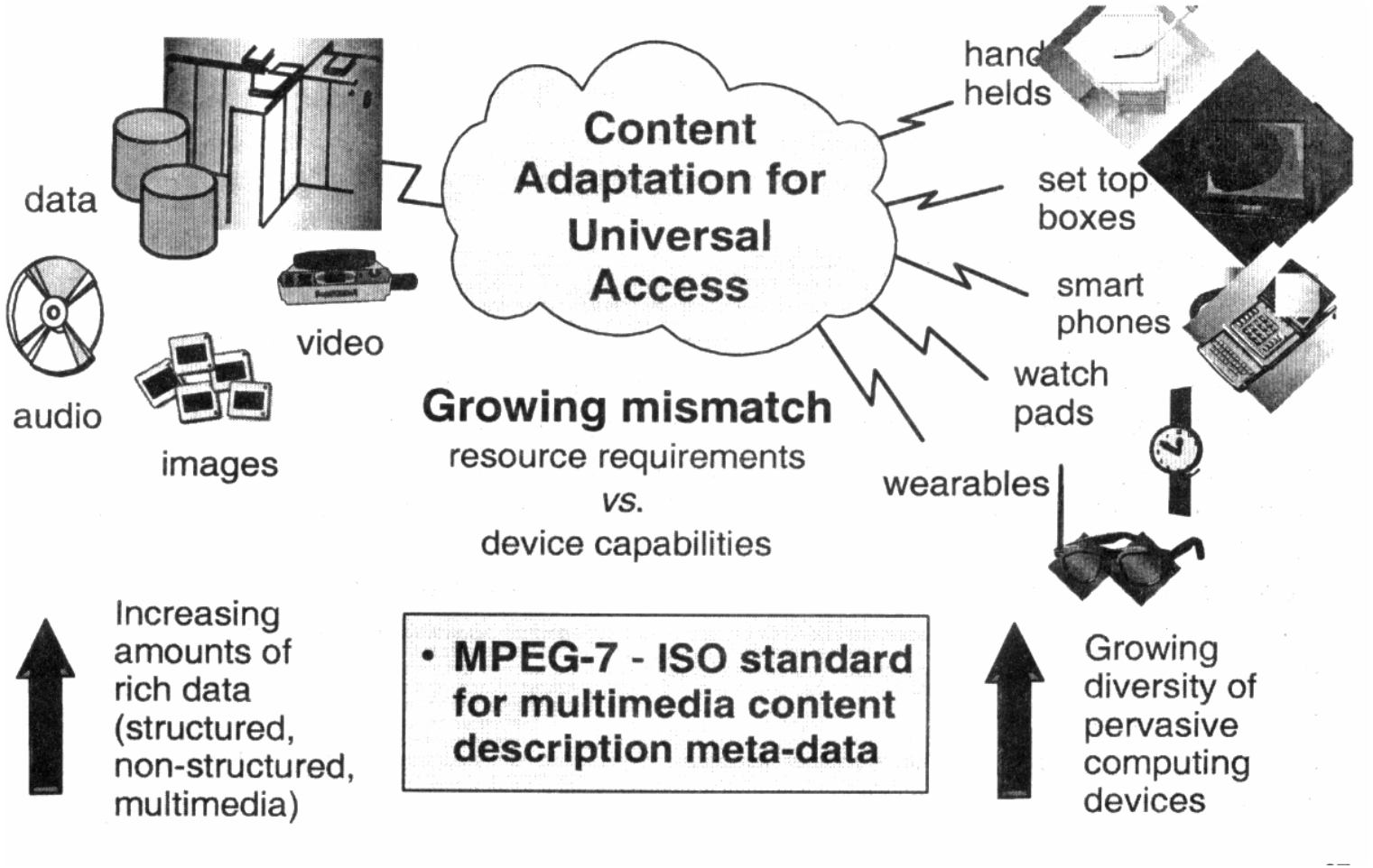
- Architecture, real estate, and interior design (e.g., searching for ideas)
- Broadcast media selection (e.g., radio channel, TV channel)
- Cultural services (history museums, art galleries, etc.)
- Digital libraries (e.g., image catalogue, musical dictionary, bio-medical imaging catalogues, film, video and archives)
- E-Commerce (e.g., personalized advertising, on-line catalogues, directories of e-shops)
- Education (e.g., repositories of multimedia courses, multimedia search for support material)
- Home Entertainment (e.g., systems for the management of personal multimedia collections, including manipulation of content, e.g. home video editing, searching a game karaoke)
- Investigation services (e.g., human characteristics recognition, forensics)
- Journalism (e.g. searching speeches of a certain politician using his name, his voice or his face)
- Multimedia directory services (e.g. yellow pages, Tourist information, Geographical information systems)
- Multimedia editing (e.g., personalized electronic news service, media authoring)
- Remote sensing (e.g., cartography, ecology, natural resources management)
- Shopping (e.g., searching for clothes that you like)
- Social (e.g. dating services)
- Surveillance (e.g., traffic control, surface transportation, non-destructive testing in hostile environments)

Application Scenario

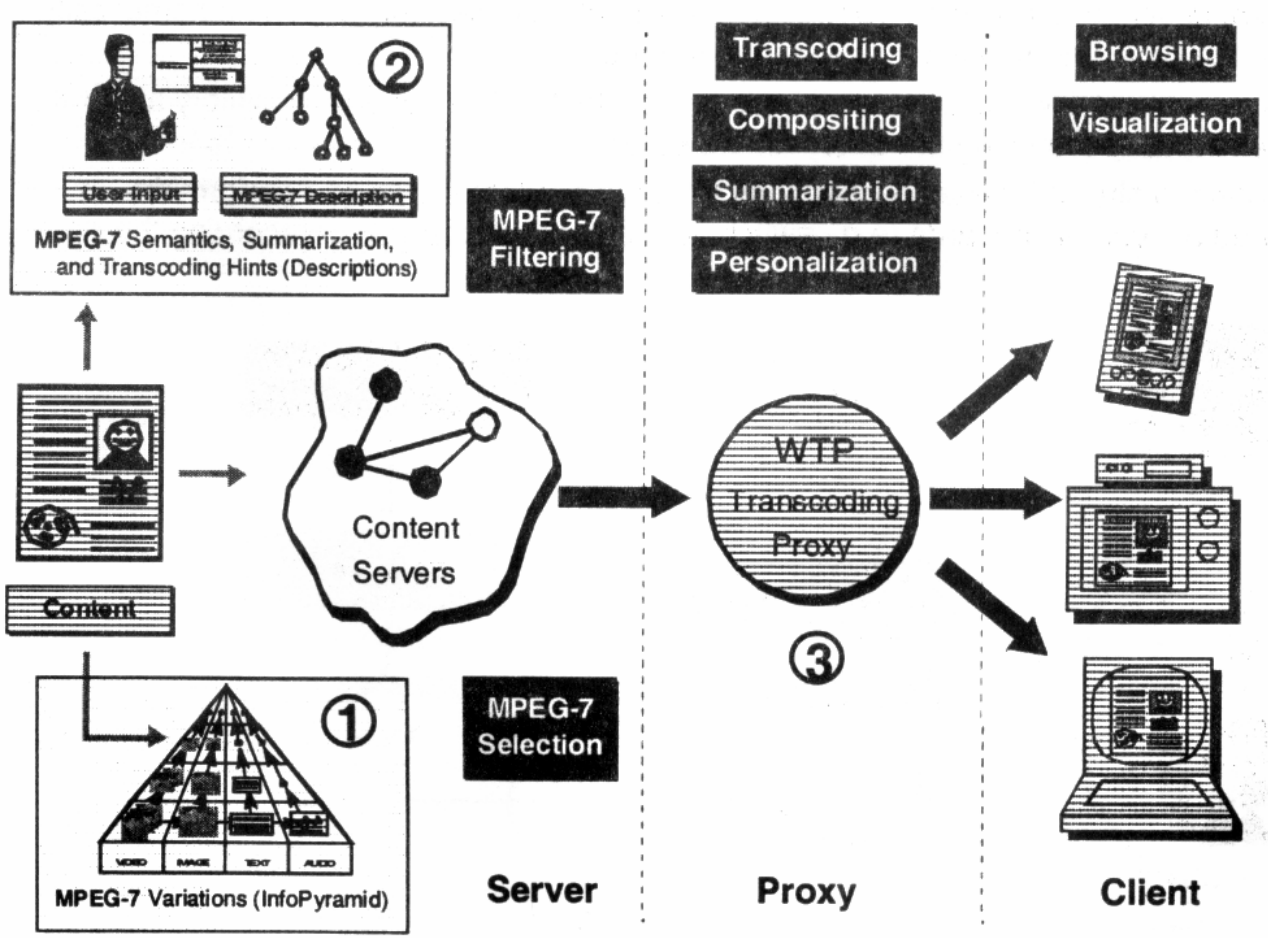
- Audio-visual descriptors obtained from multimedia contents manually or semi-automatically extraction, the descriptors may be stored or streamed
- Pull scenario: client applications will submit queries to the descriptions repository and will receive a set of descriptions matching the query for browsing
- Push scenario: a filter will select descriptions from the available ones and perform the programmed actions afterward (e.g., switching a broadcast channel or recording the described stream)



Application: Universal Multimedia Access



Application: Universal Multimedia Access



Additional Applications

- Mobile phone
  - Information retrieval, e.g., news, from mobile terminal
  - Push new information, e.g., pictures, to personal database
  - M-commerce, tele-shopping
- Personal video recording (PVR) systems
  - Smart access techniques that would enable consumer video devices like DVD to go beyond simple playback
  - Filter and summarize favorite programs according to user prefers
- Content delivery
  - Collect information about the content to improve transcoding quality and reduce complexity (in server or terminal)
- Surveillance, remote sensing, visual control equipment
- Database, server