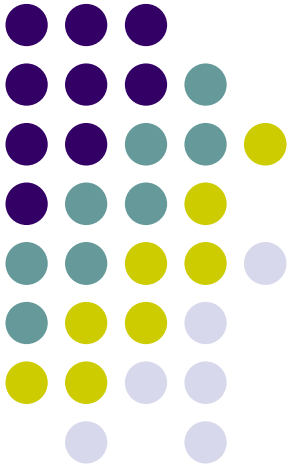
Lecture 7: Introduction to Multimedia Content Description

Reji Mathew & Jian Zhang
NICTA & CSE UNSW
COMP9519 Multimedia Systems
S2 2009



THE UNIVERSITY OF NEW SOUTH WALES
SYDNEY · AUSTRALIA

NICTA

Outline

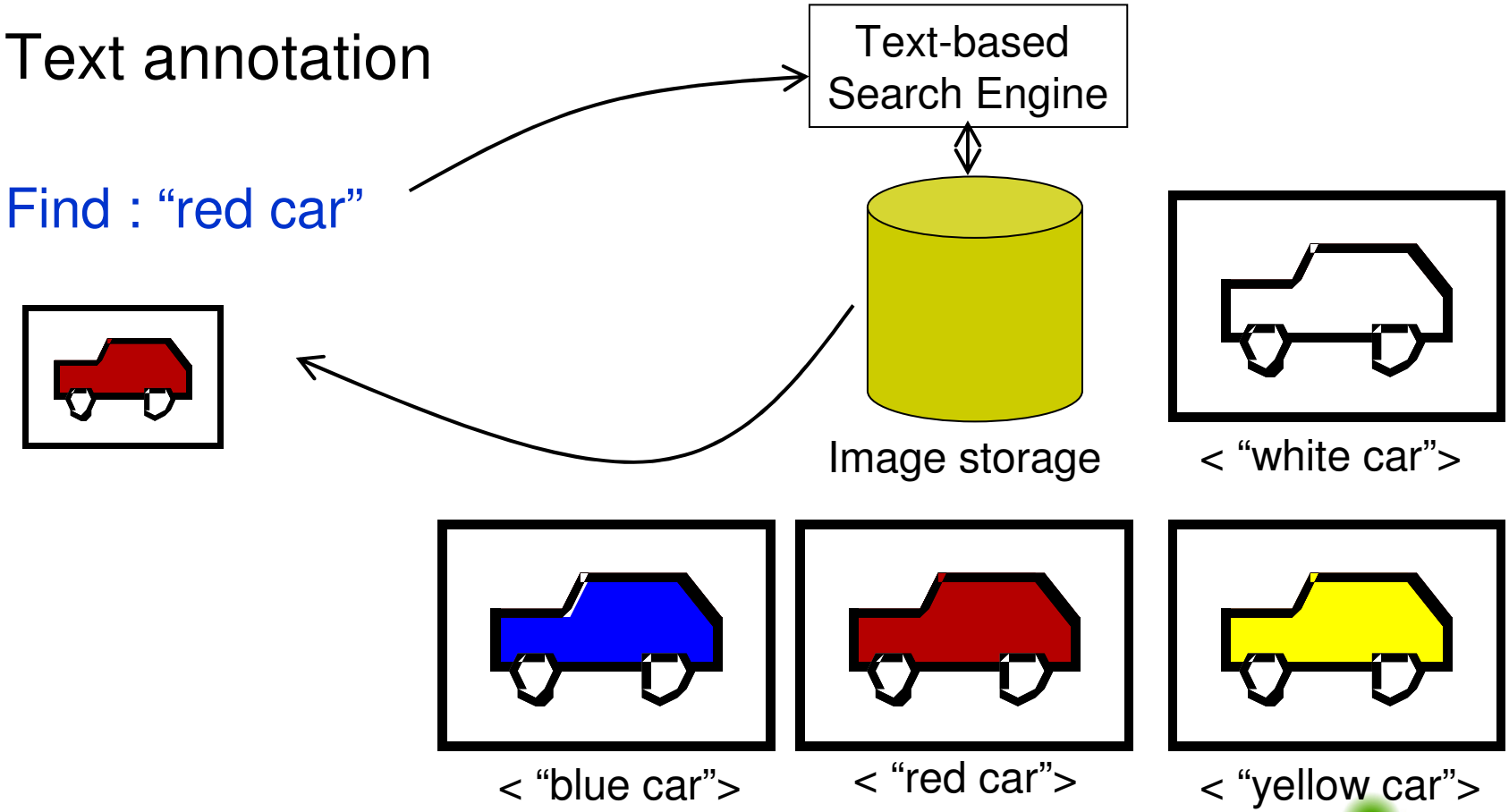
- Why do we need to describe multimedia content ?
  - Low level descriptors
  - High level descriptors
- Why standardize description of multimedia ?
- Application areas
- International Standard : MPEG-7
  - Overview
  - Multimedia Descriptions Schemes (MDS)
  - Visual and Audio descriptors – examples
  - System – Queries of video, image database

Describing multimedia content

- Explosion in the availability of digital media content
  - Individuals now creators and producers of content
  - Digital cameras, increased storage capability, internet …
- Large collections of media items
  - Images, video, animation, audio recordings, …..
- Problem :
  - How to search and discover multimedia contents ?
  - How to index video and audio sequences ?
  - How to easily browse contents ?

Describing multimedia content

- Searching and Discovering content
- Text annotation



Describing multimedia content

- Text based annotation is not always suitable
  - Requires manual description to label content
  - Not suitable for large collections of content
  - Subjective, description may vary from person to person
- Desirable to have objective features to describe multimedia contents
  - Objective features can be automatically generated
  - Examples – colour histogram, level of motion in video
- Framework still required for textual descriptions
  - High level or semantic descriptions and relationships
  - Example – photo of “two people shaking hands”

Describing multimedia content

- Text based annotation is not always suitable
- Desirable to have objective features to describe multimedia content
- Framework still required for textual descriptions

- A need exists for an architecture
  - That can integrate low-level and high-level descriptors
  - Able to describe content from many application domains
  - Rich set of descriptions

- **MPEG-7** : *multimedia content description interface*

Outline

- Why do we need to describe multimedia content ?
  - Low level descriptors
  - High level descriptors
- Why standardize description of multimedia ?
- Application areas
- International Standard : MPEG-7
  - Overview
  - Multimedia Descriptions Schemes (MDS)
  - Visual and Audio descriptors – examples
  - System – Queries of video, image database

Content Description Standard

- MPEG-7 : multimedia content description interface
  - An international standard for descriptions and description systems
- Goal of MPEG-7 Standard
  - Allow interoperable searching, indexing, filtering and access of multimedia content
  - Enable interoperability among devices that deal with multimedia content description
- Why standardize ?
  - To enable interoperability
  - Examples :
    - Search across different repositories
    - Content exchange between different databases

MPEG-7 Introduction

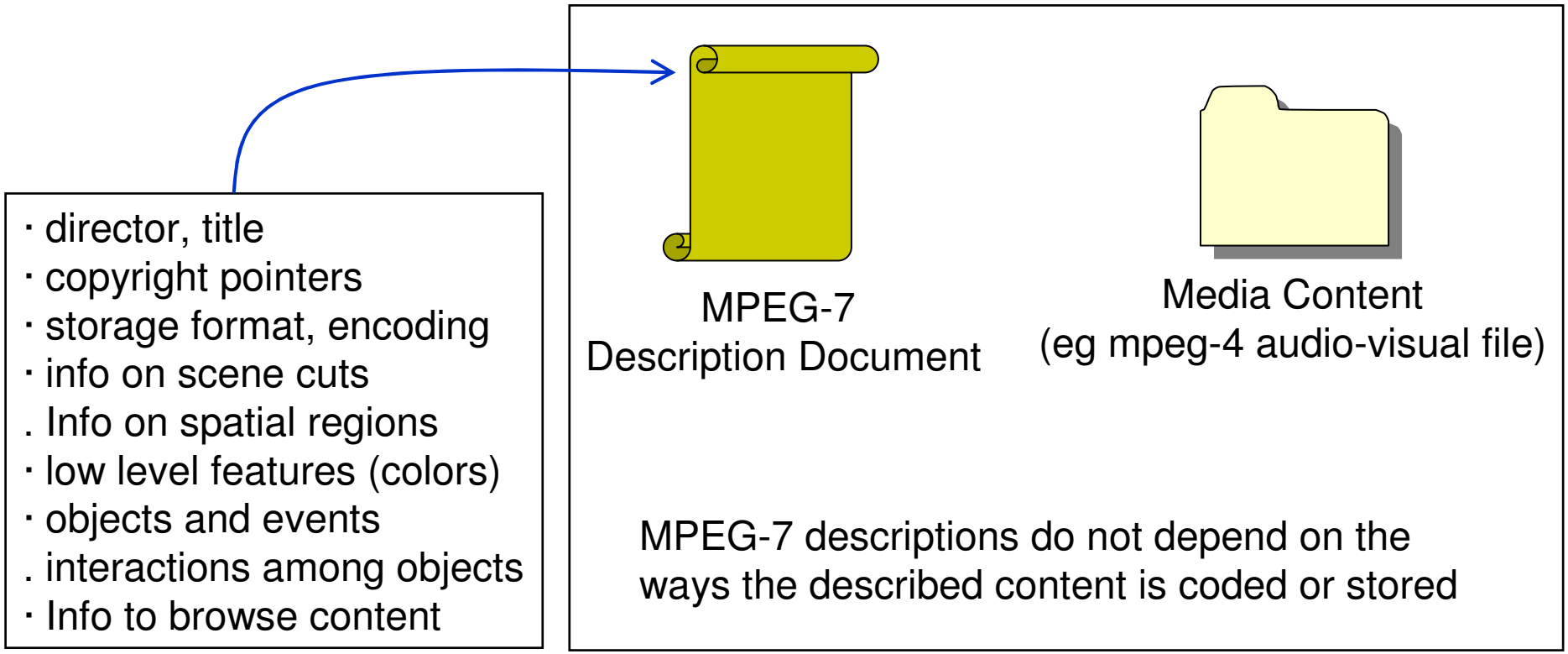
- The MPEG-7 descriptions of content that may include:
  - Information describing creation & production process of the content
    - director, title, short feature movie
  - Information related to the usage of the content
    - copyright pointers, usage history, broadcast schedule
  - Information of the storage features of the content
    - storage format, encoding
  - Structural information on spatial, temporal or spatio-temporal components of the content
    - scene cuts for video, segmented regions for image

MPEG-7 Introduction

- The MPEG-7 descriptions of content that may include: (continued)
  - Information about low level features in the content
    - colors, textures, sound timbres, melody description
  - Conceptual information of the reality captured by the content
    - objects and events, interactions among objects
  - Information about how to browse the content in an efficient way
    - summaries, variations
  - Information about collections of objects.
  - Information about the interaction of the user with the content
    - user preferences, usage history

MPEG-7 Introduction

- MPEG-7 enables description of content from several viewpoints



Application Areas

- Digital libraries
  - Searching through bio-medical imaging catalogues
  - Play a few notes on a keyboard and retrieve similar music segments from musical repository
- Journalism
  - Search radio archives based on name of a politician
- Home Entertainment
  - Search digital photo collection based on an example image
  - Search based on an example colour or sketch
- Surveillance
  - Store detected events for searching / indexing
  - Example : accompany surveillance video with metadata of locations and time of detected motion regions

MPEG-7 : Normative Elements

- Four types of normative elements

- Descriptors (D): describe individual features of multimedia content
  - Describe low-level features : colour, motion, audio energy
  - Describe high-level features of semantic object

- Description Schemes (DS) : descriptions by integrating together multiple descriptors and description schemes
  - Combining D and DS within more complex structures
  - Defining relationships between D and DS

- Description Definition Language (DDL) : used to define D and DS, an extension of the XML Schema language.

- System Tools : binary coded representation for efficient storage and transmission, ….

MPEG-7 : Normative Elements

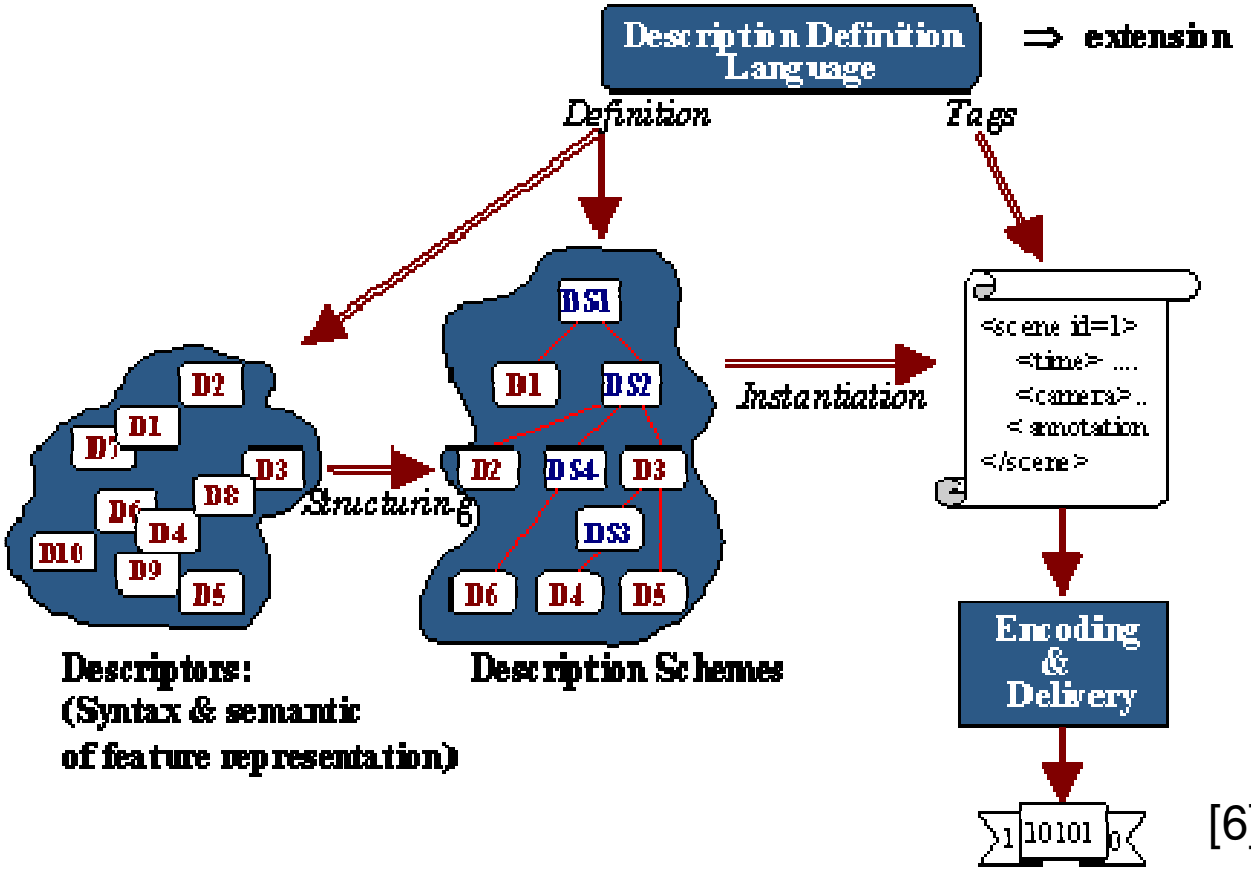
- Four types of normative elements (continued)

- Descriptors (D)
- Description Schemes (DS)

→ Description Tools

- Description Definition Language (DDL)
  - Based on XML Schema Language
  - Consists of
    - XML Schema Structural Components
    - XML Schema Data Types
    - MPEG-7 Specific Extensions
  - System Tools

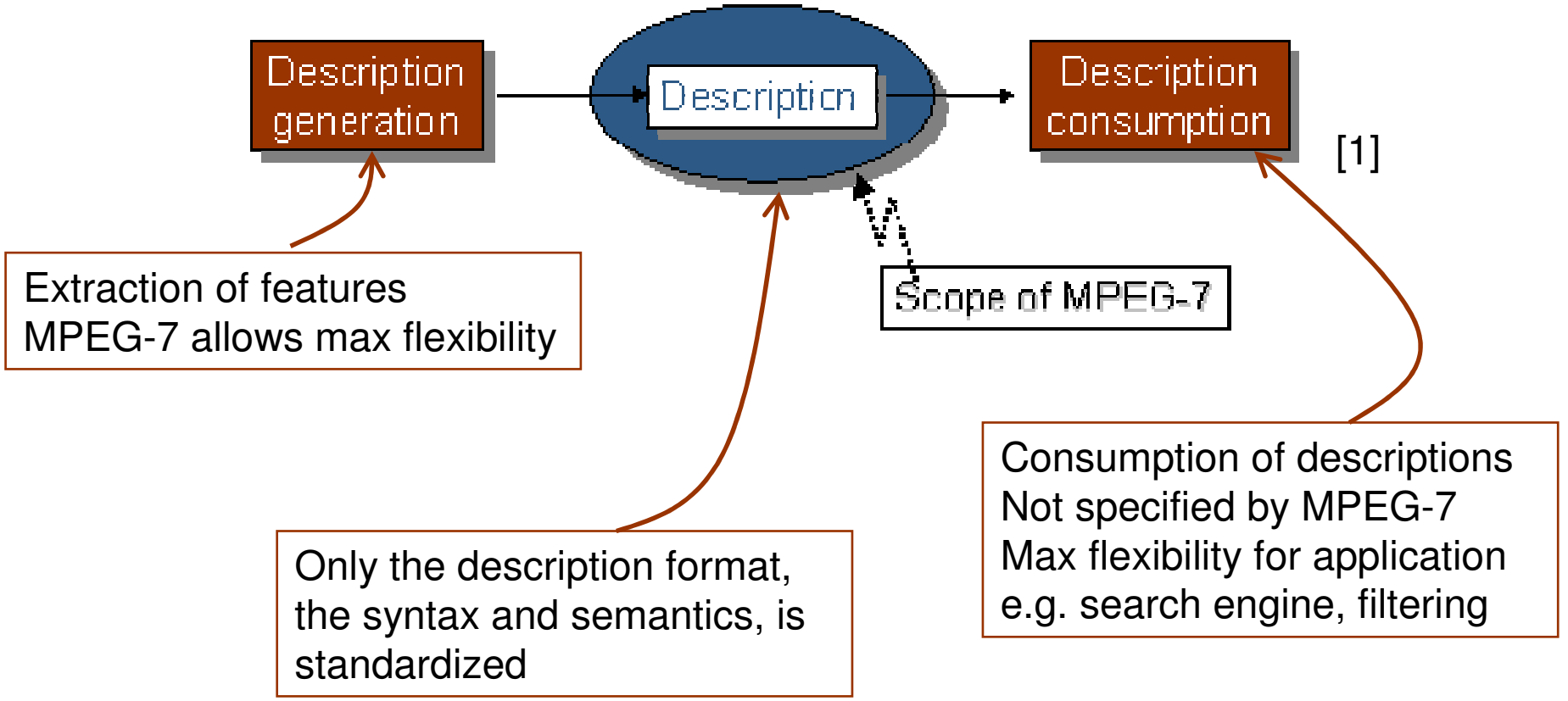
MPEG-7 : Normative Elements [6]



[6]

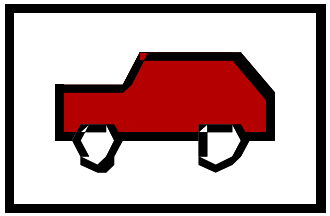
MPEG-7 allows to create descriptions, which is a set of instantiated Description Schemes and their corresponding Descriptors and to deploy the descriptions using System tools.

MPEG-7 : Scope


Description generation
Description
Description consumption
[1]
Extraction of features
MPEG-7 allows max flexibility
Scope of MPEG-7
Only the description format, the syntax and semantics, is standardized
Consumption of descriptions
Not specified by MPEG-7
Max flexibility for application
e.g. search engine, filtering

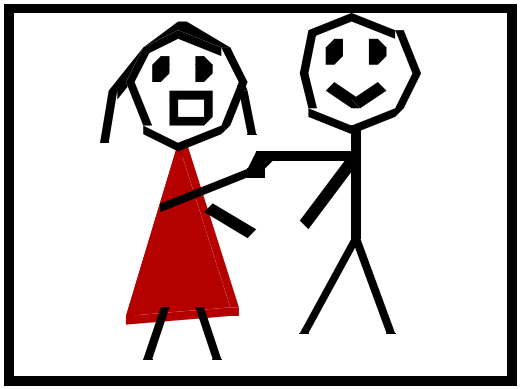
MPEG-7 : Example [6]

```
<Mpeg7>
  <Description xsi:type="SemanticDescriptionType">
    <Semantics>
      <Label>
        <Name> Car </Name>
      </Label>
      <Definition>
        <FreeTextAnnotation>
          Four wheel motorized vehicle
        </FreeTextAnnotation>
      </Definition>
      <MediaOccurrence>
        <MediaLocator>
          <MediaUri> image.jpg </MediaUri>
        </MediaLocator>
      </MediaOccurrence>
    </Semantics>
  </Description>
</Mpeg7>
```



MPEG-7 : Example

- MPEG-7 description of the event of handshake between two people:
  - See next slide, example taken from [6]



[6]

```
<Mpeg7>                                                    [6]
  <Description xsi:type="SemanticDescriptionType">
    <Semantics>
      <Label>
        <Name> Shake hands </Name>
      </Label>
      <SemanticBase xsi:type="AgentObjectType" id="A">
        <Label href="urn:example:acs">
         <Name> Person A </Name>
        </Label>
      </SemanticBase>
      <SemanticBase xsi:type="AgentObjectType" id="B">
        <Label href="urn:example:acs">
          <Name> Person B </Name>
        </Label>
      </SemanticBase>
      <SemanticBase xsi:type="EventType">
        <Label><Name> Handshake </Name></Label>
        <Definition>
          <FreeTextAnnotation> Clasping of right hands by two people </FreeTextAnnotation>
        </Definition>
        <Relation type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:agent" target="#A"/>
        <Relation type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:accompanier“target="#B"/>
      </SemanticBase>
    </Semantics>
  </Description>
</Mpeg7>
```

MPEG-7 Parts

- Systems : the tools needed to prepare MPEG-7 descriptions for efficient transport and storage and the terminal architecture.
- Description Definition Language : the language for defining the syntax of the MPEG-7 Description Tools and for defining new Description Schemes.
- Visual : the Description Tools dealing with (only) Visual descriptions.
- Audio : the Description Tools dealing with (only) Audio descriptions.
- Multimedia Description Schemes : the Description Tools dealing with generic features and multimedia descriptions.

MPEG-7 Parts

(continued)

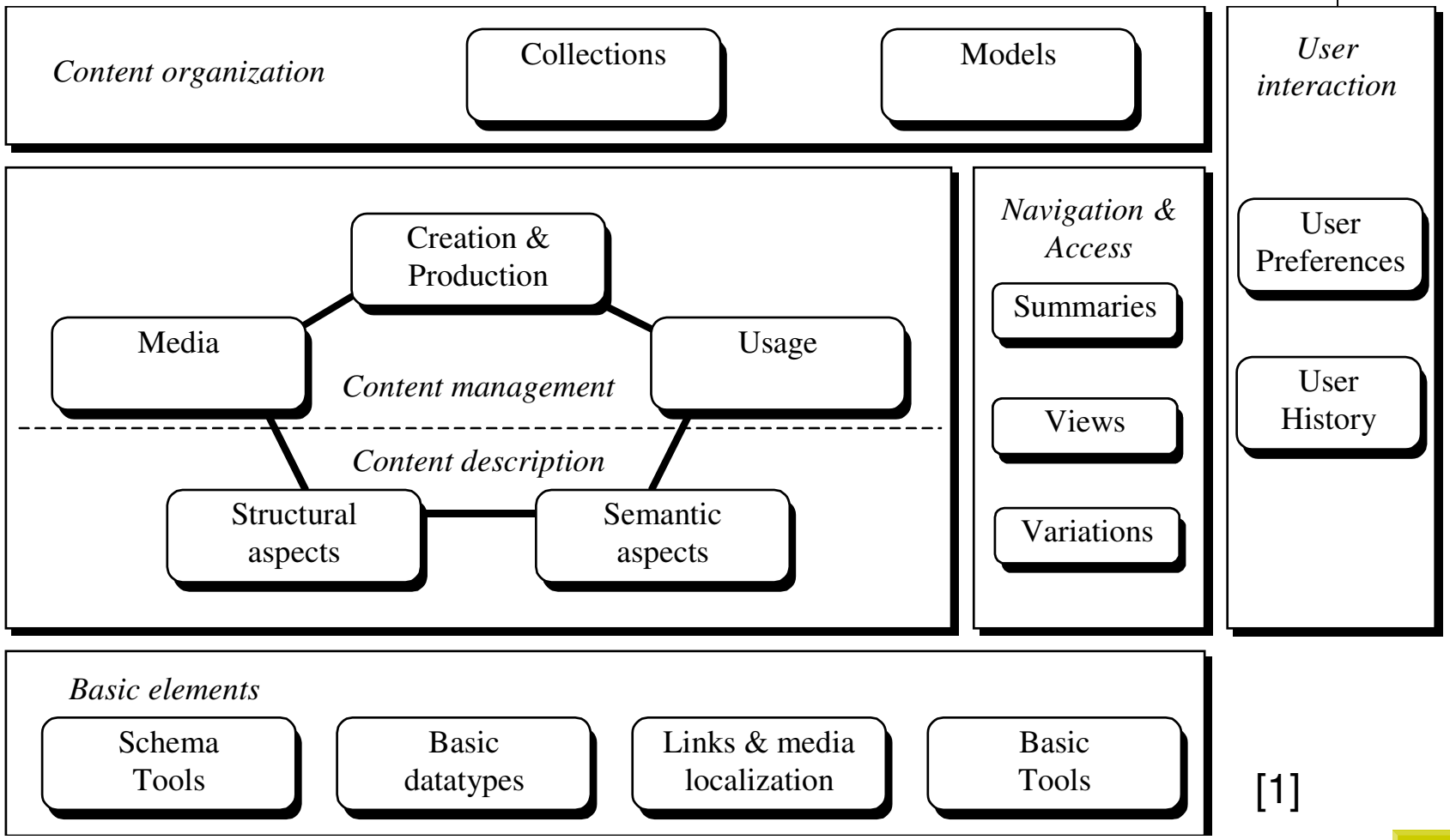
- Systems
- Description Definition Language
- Visual
- Audio
- Multimedia Description Schemes

- Reference Software : a software implementation of relevant parts of the MPEG-7 Standard with normative status.
- Conformance Testing : guidelines and procedures for testing conformance of MPEG-7 implementations
- Extraction and use of descriptions – informative material about the extraction and use of some of the Description Tools.

MPEG-7 : MDS

- Multimedia Description Schemes (MDS)
  - Description Tools dealing with generic features and multimedia descriptions
    - Metadata structures for describing and annotating multimedia content
    - NOT specific to image, video or audio but general to multimedia content.
  - MDS is organized into the following areas
    - Basic Elements
    - Content Description
    - Content Management
    - Content Organization
    - Navigation and Access
    - User Interaction

MPEG-7 : MDS [1]


Content organization
Collections
Models
User interaction
Creation & Production
Media
Usage
Content management
Content description
Structural aspects
Semantic aspects
Navigation & Access
Summaries
Views
Variations
User Preferences
User History
Basic elements
Schema Tools
Basic datatypes
Links & media localization
Basic Tools

[1]

MPEG-7 : MDS

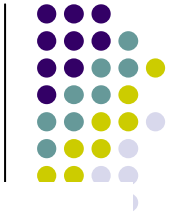
- Basic Elements
  - Essentials of multimedia content description
  - Used repeatedly in descriptions of multimedia content
  - Used by other parts of MPEG-7 (Visual and Audio)
  - Examples
    - Schema Tools: XML-like Language
    - Basic data types : for describing matrices
    - Linking and localization tools : link MPEG-7 descriptions to media
    - Basic Tools : graph tool to represent relation
      : text annotation
      : description schemes for describing people & places

  More Info :
  - www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm
  - *Introduction to MPEG-7, Multimedia Content Description Interface*, John Wiley & Sons, 2002

MPEG-7 : MDS

- Description Schemes for **Content Management**
  - Creation Info : Title, creators, creation location & dates, genre category, age classification, ...
  - Usage Info : Usage rights, ...
    Links to rights holders & rights management
  - Media description : compression, coding and storage format of multimedia content

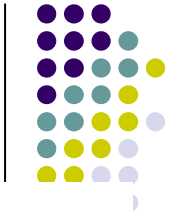
MPEG-7 : MDS – Content Management



```
<CreationInformation>
 <Creation>
  <Creator>
   <Role><Name xml:lang="en">Photographer</Name></Role>
   <Agent xsi:type="PersonType">
    <Name>
     <GivenName>Seungyup</GivenName>
    </Name>
   </Agent>
  </Creator>
  <CreationCoordinates>
   <Location>
    <Name xml:lang="en">Columbia University</Name>
    <Region>us</Region>
   </Location>
   <Date>
    <TimePoint>1998-09-19</TimePoint>
   </Date>
  </CreationCoordinates>
 </Creation>
</CreationInformation>
```

XML example of content management descriptions

MPEG-7 : MDS – Content Management

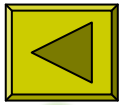


```
<MediaFormat>
  <Content href=" urn:mpeg:mpeg7:cs:ContentCS:2001:1">
    <Name xml:lang="en">image</Name>
  </Content>
  <FileFormat href=" urn:mpeg:mpeg7:cs:FileFormatCS:2001:1">
    <Name xml:lang="en">jpeg</Name>
  </FileFormat>
  <VisualCoding>
    <Format colorDomain="color"
      href=" urn:mpeg:mpeg7:cs:VisualCodingFormatCS:2001:1">
      JPEG
    </Format>
    <Frame height="480" width="704"/>
  </VisualCoding>
</MediaFormat>
```

XML example of content management descriptions

MPEG-7 : MDS – Content Description

- Description Schemes for **Content Description**
  - content description tools describe the **structure** and **semantics** of multimedia data
  - Structure : segments – will explore this first
    Describe :Objects in image, video shot, audio segment ……
  - Semantics : describing *semantic entities* in the narrative world
    Describe :People, Actions, Concepts,
    Relation between people and actions,
    actions and concepts….

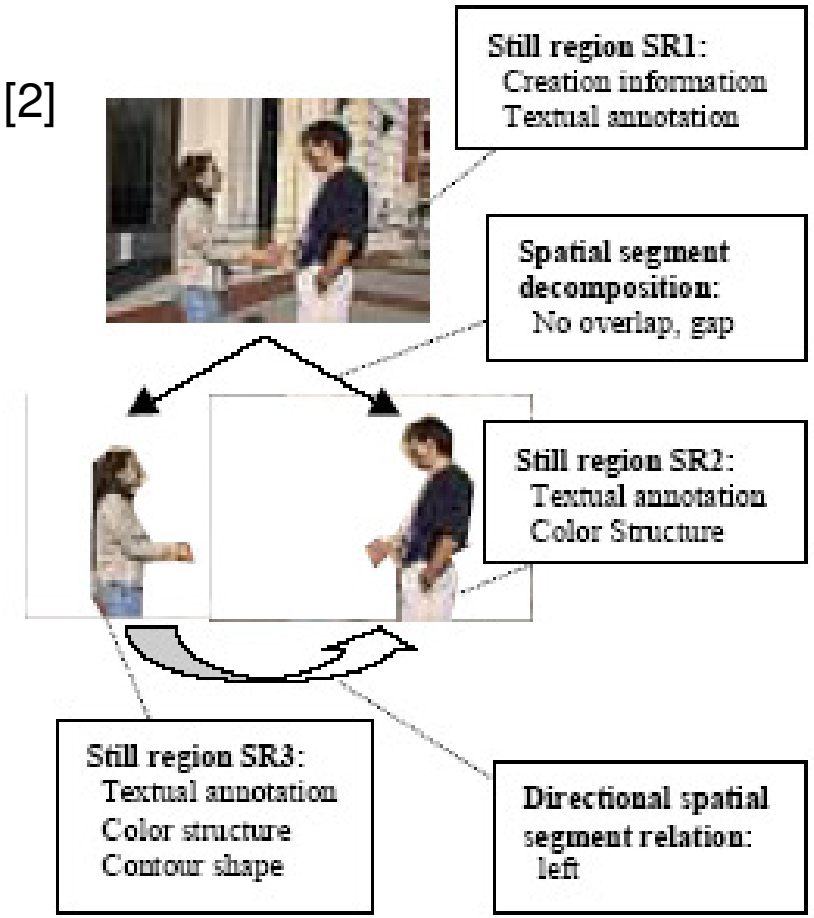


MPEG-7 : MDS – Content Description

- Description Schemes for **Content Description**
  - Describe structure of content
    - Describes content by using the notion of Segments
    - Image regions, video frames, audio segments
    - Describe segments : using low level descriptors, text annotation,…
    - Example
    - A single image decomposed into a set of segments (or regions)
    - Each image region can then be further described using other tools
    - A single video / audio clip can be decomposed into a set temporal segments
      - E.g. Segment a video clip – into video shots

MPEG-7 : MDS – Content Description

[2]



Describing Structure using StillRegion segments (spatial portions)

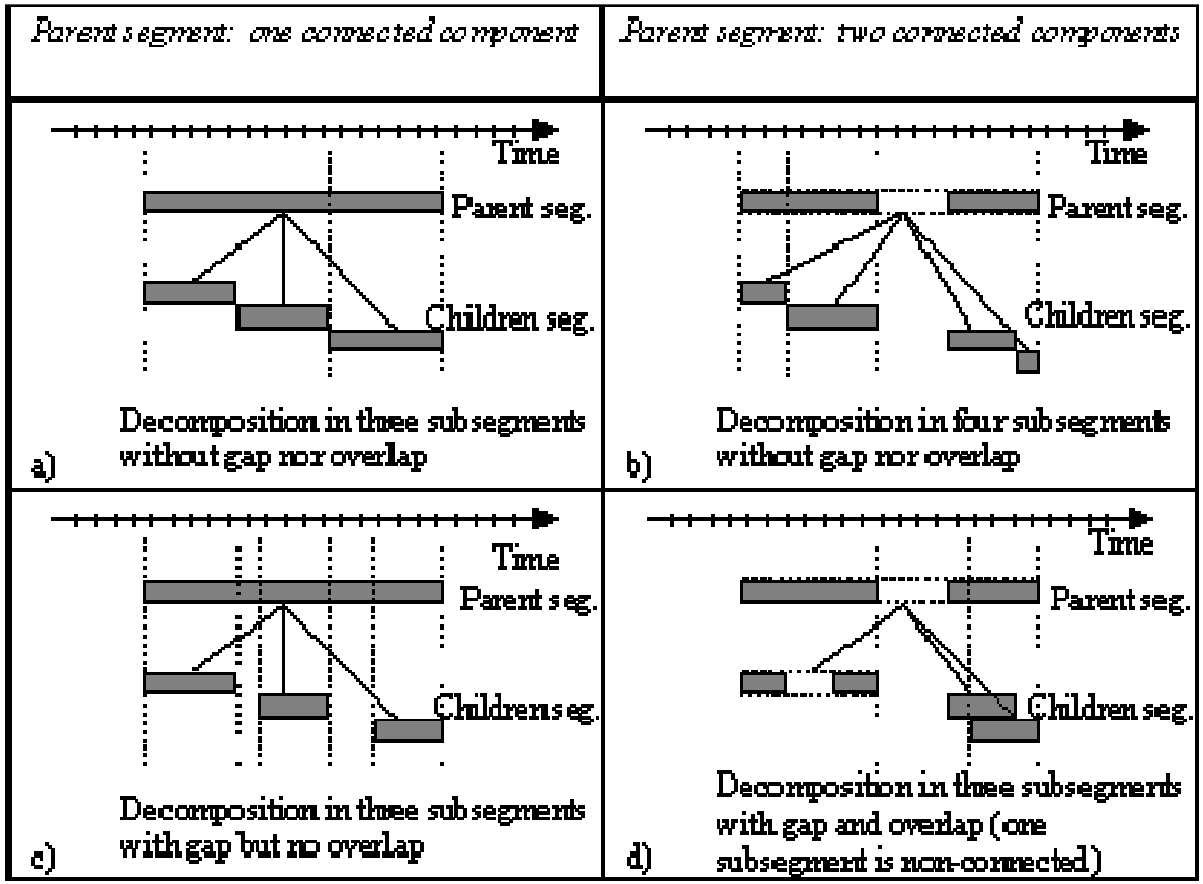
Decompose the image (SR1) into two segments corresponding to the two people in the image (SR2 and SR3).

Further describe segments using colour feature and text annotation

Describe spatial relation between SR2 and SR3

MPEG-7 structural relations include left (spatial), precedes (temporal), ….

MPEG-7 : MDS – Content Description



[1]

Similarly, temporal portions of video can constitute segments

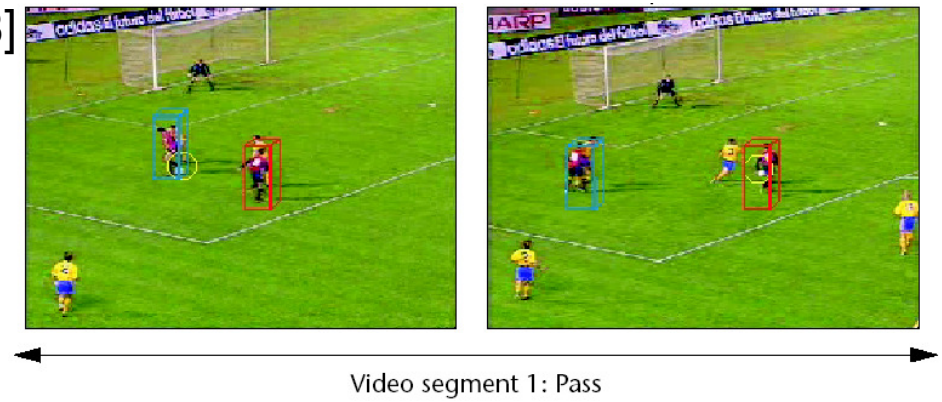
Decompose one video clip into segments, with or without overlap.

Each segment can then be described further (e.g. using level of motion in video and text annotations).

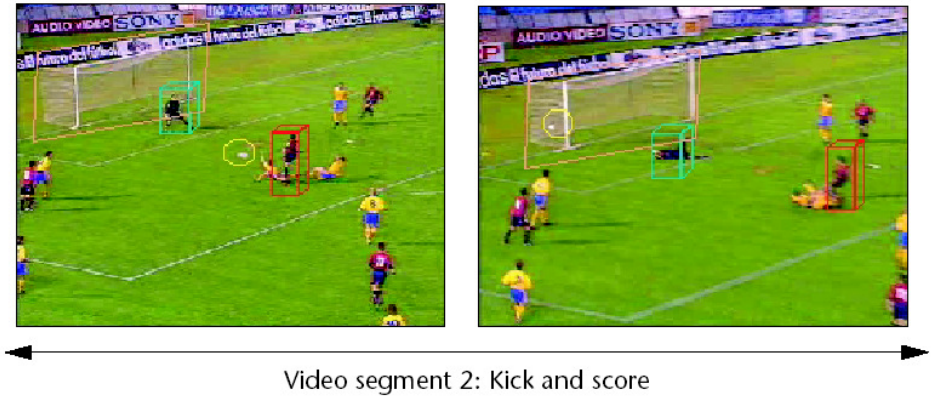
Suited to video shot boundary detection and indexing

MPEG-7 : MDS – Content Description [3]

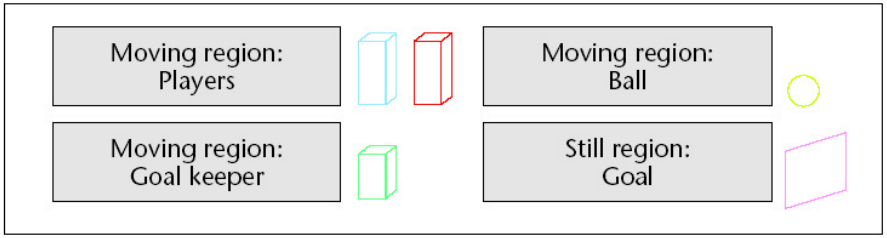
[3]



Video segment 1: Pass



Video segment 2: Kick and score



Spatio-temporal segments or moving regions

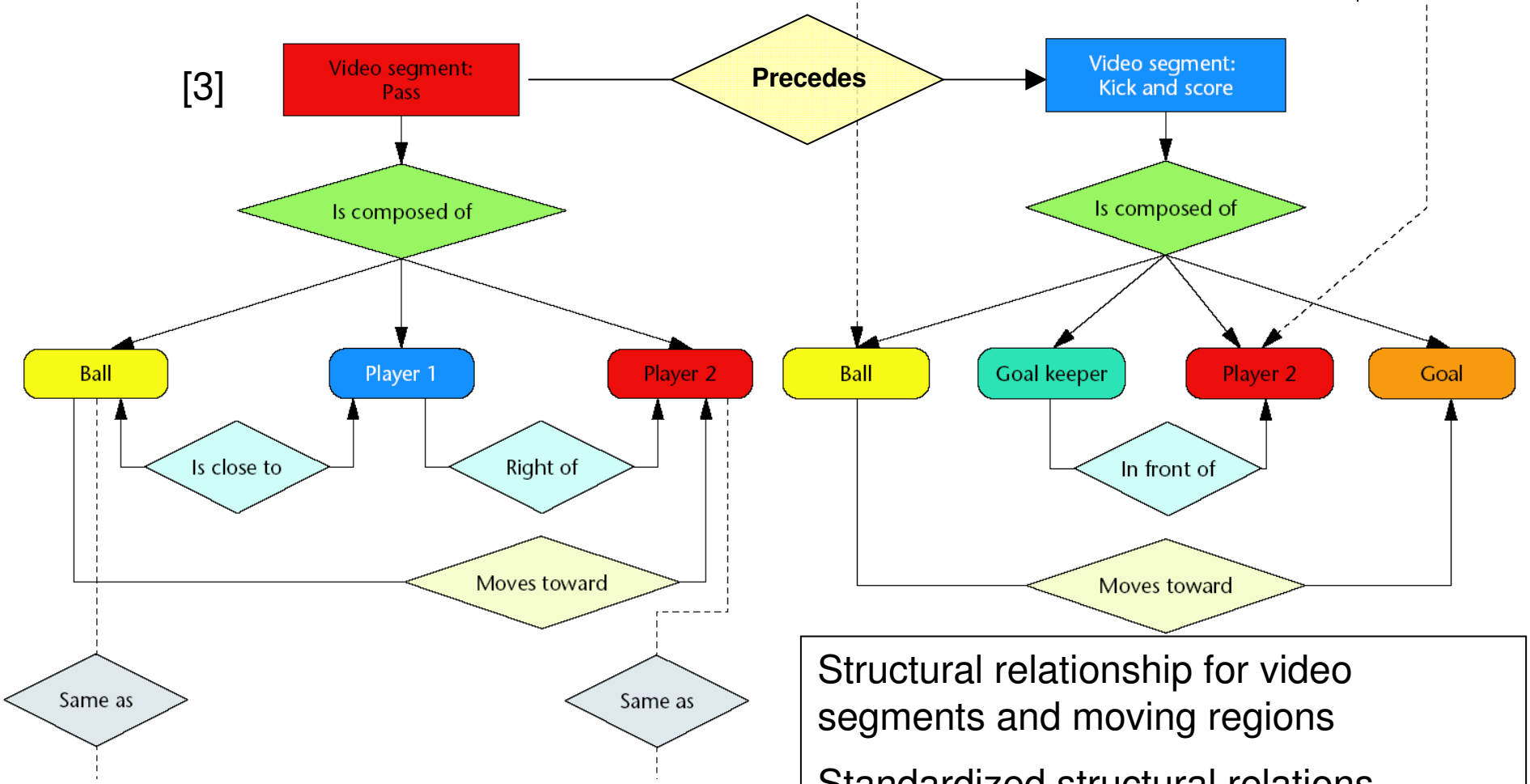
Decompose video segment into various moving regions (spatio-temporal segments).

Further descriptions of moving regions possible

Structural relation tools to describe more general segment structures

Example – segment relationship graph, see next slide

MPEG-7 : MDS – Content Description [3]

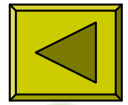


Structural relationship for video segments and moving regions

Standardized structural relations
As well as non-normative relations

MPEG-7 : MDS – Content Description

- Description Schemes for **Content Description**
  - content description tools describe the **structure** and **semantics** of multimedia data
  - Structure : segments
  - Semantics : describing *semantic entities* in the narrative world
    - Examples include :
      - agent objects (eg person or a group of people)
      - *events :* perceivable event that takes place in time and space in the narrative world
      - *semantic relation* : describe general relations between entities
        “hasAgentOf” - initiates the action of an event
        “hasAccompanierOf”- object that is a join agent in an event



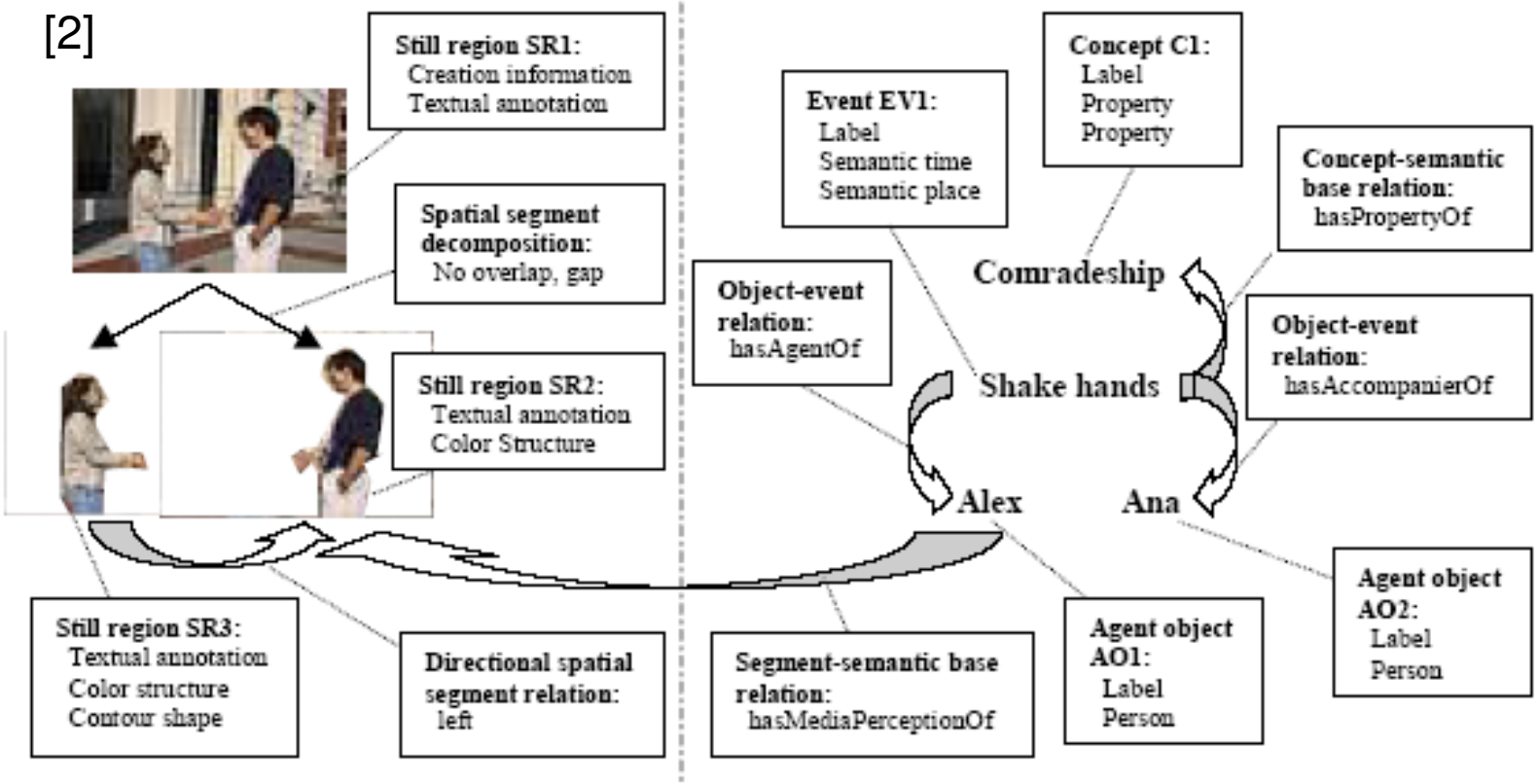
MPEG-7 : MDS – Content Description

- Structure : describing structure of content

- Semantics : describing semantics and concepts
  - Semantic description scheme
    - Objects (person, car, …)
    - Events (perceivable occurrence)
    - Abstract concepts
    - Relationships
    -
  - Multimedia content can be described by both content structure and semantics
    - Related together by a set of links

  - Example : see next slide

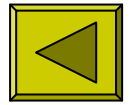
MPEG-7 : MDS – Content Description [2]



Structure Description and Semantic Description

MPEG-7 : MDS – Navigation & Access

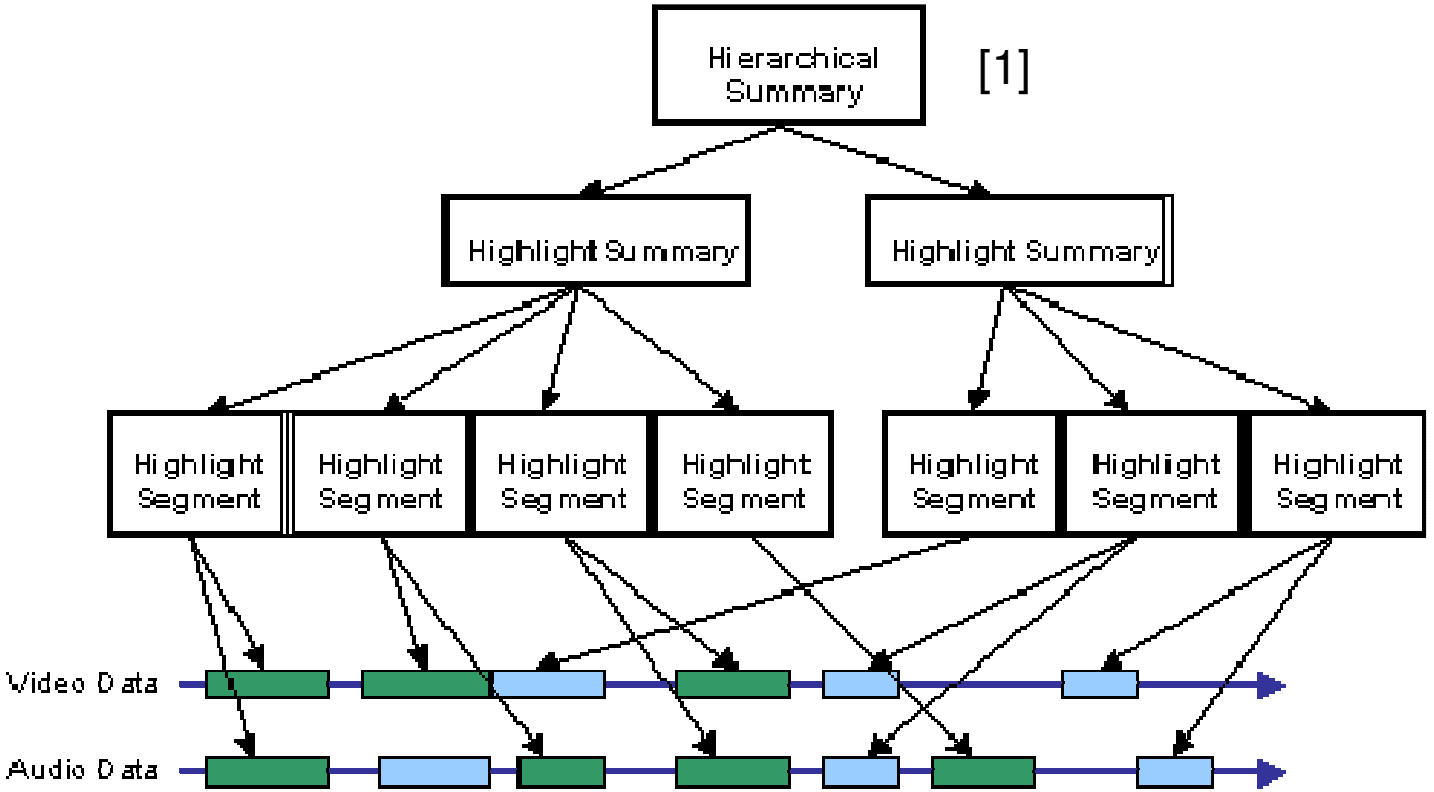
- Description schemes for enabling browsing & retrieval
  - Summarization tools :
    - Summarize a long video clip to highlight important segments
    - Allows fast browsing of content
    - Example – highlights of a soccer game (just shots at goals)
  - View tools :
    - Different partitions and decompositions of image, video & audio
  - Variations :
    - Describe different variations of content available
    - Example – low resolution version, video only version, ….



MPEG-7 : MDS – Navigation & Access

- Description schemes for enabling browsing & retrieval
  - Summarization tools :
    - Summarize a long video clip to highlight important segments
    - Allows fast browsing of content
    - Example – highlights of a soccer game (just shots at goals)
  - Example
    - Describe summarizations of video content that allows fast and flexible browsing
  - Applications include
    - summarizing sports events,
    - Key frames or shots of long periods of surveillance video

MPEG-7 : MDS – Navigation & Access



MPEG-7 enables that above summary (of audio-video content) to be captured in XML format

Outline

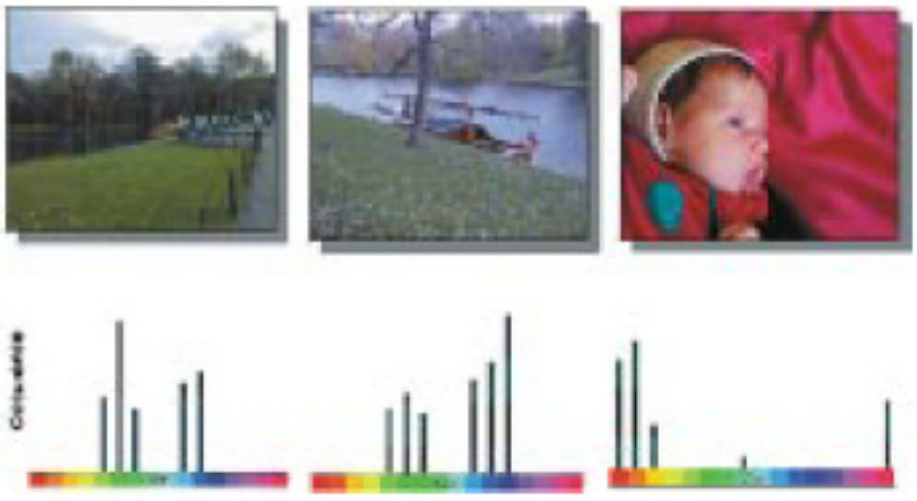
- Why do we need to describe multimedia content ?
  - Low level descriptors
  - High level descriptors
- Why standardize description of multimedia ?
- Application areas
- International Standard : MPEG-7
  - Overview
  - Multimedia Descriptions Schemes (MDS)
  - Visual and Audio descriptors – examples
  - System – Queries of video, image database

MPEG-7 : Visual

- Descriptors and Description Schemes exclusively for visual information

- Descriptors : describe low-level features of visual content, such as colour, texture, motion, ….

- Example : colour histogram [4]



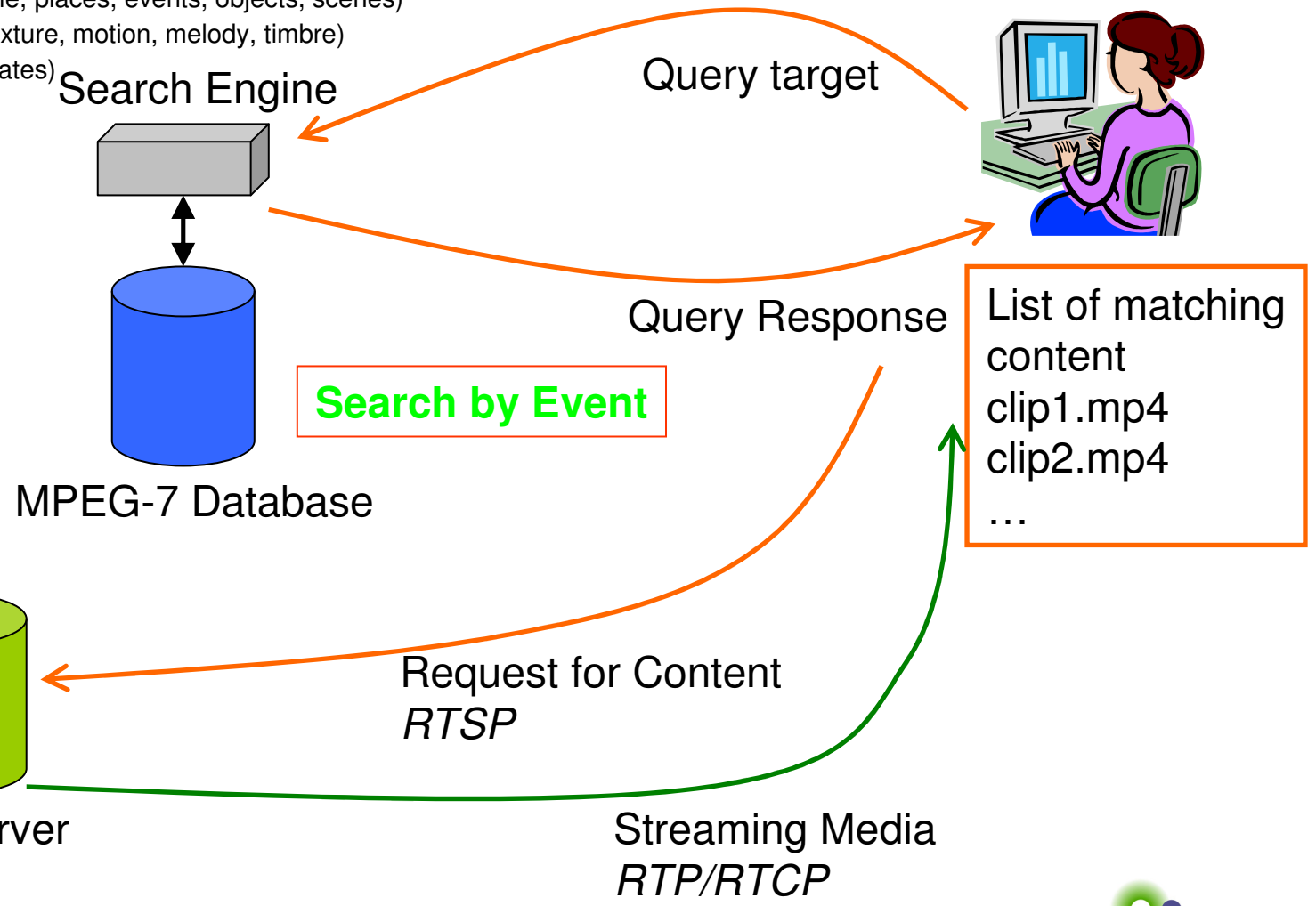
[4]

MPEG-7 : Audio

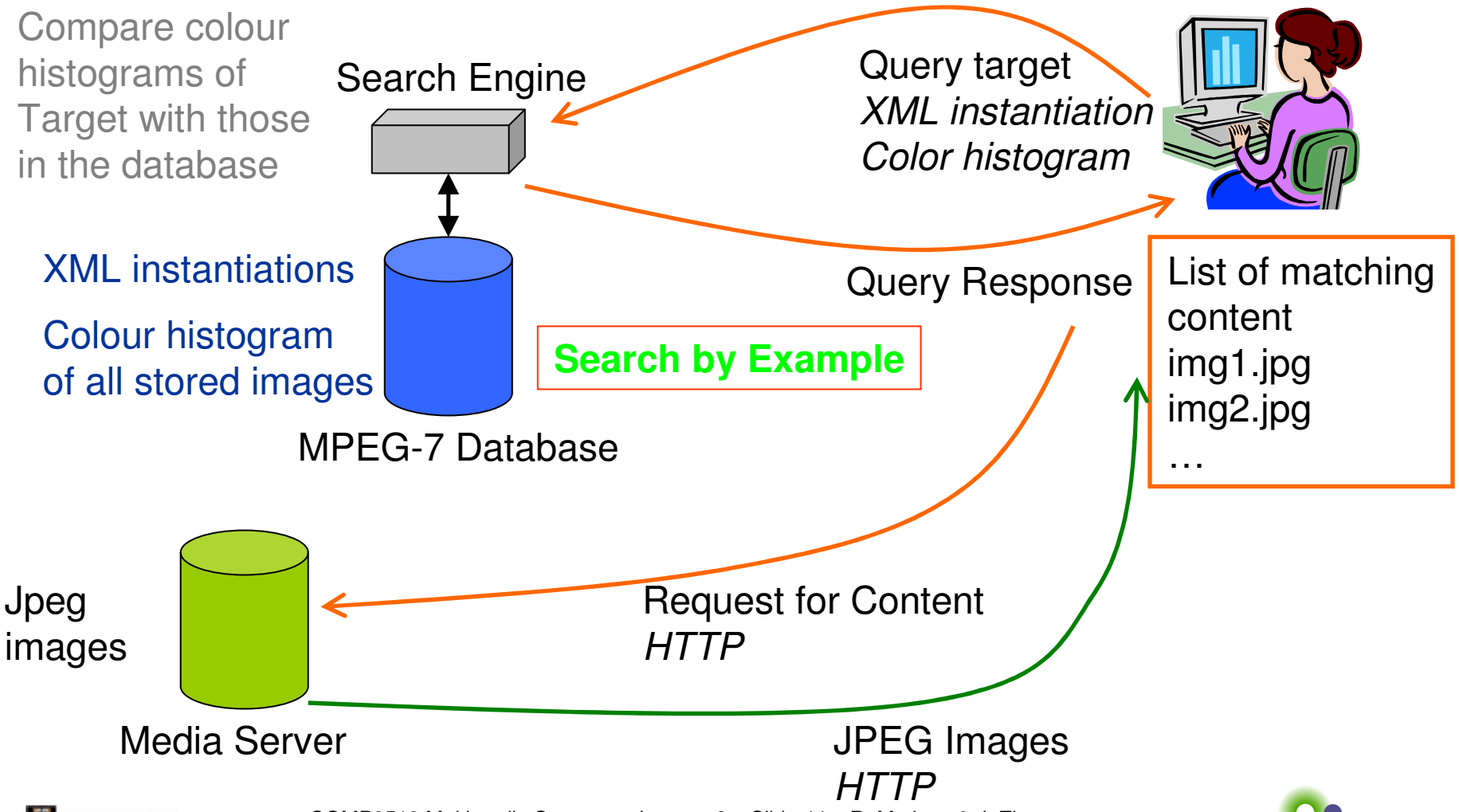
- Descriptors and Description Schemes exclusively for audio information
- Low Level Descriptors : describe low-level features of audio content, such as instantaneous waveform and power values, power spectrum and spectral features, ……
- High Level Descriptors : application-specific tools

MPEG-7 System : Client Server architecture

- **MPEG-7 Indexing & Searching:**
  - Semantics-based (people, places, events, objects, scenes)
  - Content-based (color, texture, motion, melody, timbre)
  - Metadata (title, author, dates)



MPEG-7 System : Client Server architecture



MPEG 7 System : Search by Example

- Colour Histogram Descriptor
  - Used in the demonstration system
  - Example below, showing part of the XML document

```
<VisualDescriptor xsi:type="ScalableColorType" numOfCoeff="32"
  numOfBitplanesDiscarded="0">
        <Coeff> 62  17  -127  47  -8  13  22  30  -31  -33  3  13
                -25  -11  13  20  2  -13  -1  3  -11  -10  1  6  2
                -1  0  0  -9  5  1  -4
        </Coeff>
</VisualDescriptor>
```

MPEG-7 : Demonstration

- MMVC Demonstration
  - Select Image region
  - Calculate Colour Histogram for selected region
  - Generate XML instantiation
  - Submit target to search engine
  - Perform matching between histograms of target and stored content
  - Return list of best matching content (in order)
  - Retrieve images from Image Database

Outline

- Why do we need to describe multimedia content ?
  - Low level descriptors
  - High level descriptors
- Why standardize description of multimedia ?
- Application areas
- International Standard : MPEG-7
  - Overview
  - Multimedia Descriptions Schemes (MDS)
  - Visual and Audio descriptors – examples
  - System – Queries of video, image database

References and Further Reading

1. www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm

2. *Introduction to MPEG-7, Multimedia Content Description Interface*, John Wiley & Sons, 2002

3. *Lecture 14, MPEG-7*, SIMS 202: Information Organization and Retrieval, Prof. Ray Larson & Prof. Marc Davis UC Berkeley SIMS, www.sims.berkeley.edu/academics/courses/is202/f03/

4. *The MPEG-7 Visual Standard for Content Description—An Overview*, Thomas Sikora, IEEE Trans. on Circuits and Systems for Video Technology, Vol. 11, No. 6, June 2001

5. *MPEG-7: The Generic Multimedia Content Description Standard, Part 1*," IEEE MultiMedia, vol. 09, no. 2, pp. 78-87, April-June 2002

6. MPEG-7 Multimedia Content Description Standard, John R. Smith, Pervasive Media Management Group, IBM, January 8, 2003