Lecture 8: Introduction to Multimedia Content Description

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 content?
  - How to index long video and audio sequence?
  - How to easily browse content?

Describing multimedia content
- Searching and Discovering content
- Text annotation

Find: “red car”

Image storage

<“red car”>  <“blue car”>  <“yellow car”>  <“white car”>
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 content
  - 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

(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: Scope

- Extraction of features
  - MPEG-7 allows max flexibility
- Consumption of descriptions
  - Not specified by MPEG-7
  - Max flexibility for application e.g., search engine, filtering

Only the description format, the syntax and semantics, is standardized.
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]

```
<Mpeg7/>

<Mpeg7>
  <Description xsi:type="SemanticDescriptionType">
    <Semantics>
      <Label>
        <Name> Shake hands </Name>
      </Label>
      <SemanticBase xsi:type="AgentObjectType" id="A">
        <Label href="urn:example:acs">
          Person A
        </Label>
      </SemanticBase>
      <SemanticBase xsi:type="AgentObjectType" id="B">
        <Label href="urn:example:acs">
          Person B
        </Label>
      </SemanticBase>
      <SemanticBase xsi:type="EventType">
        <Label><Name> Handshake </Name></Label>
        <Definition>
          Clasping of right hands by two people
        </Definition>
      </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]

<table>
<thead>
<tr>
<th>Content organization</th>
<th>Collections</th>
<th>Models</th>
<th>User interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Media</td>
<td>Content management</td>
<td>Usage</td>
<td>Creation &amp; Production</td>
</tr>
<tr>
<td>Media</td>
<td>Content description</td>
<td>Usage</td>
<td>Content description</td>
</tr>
<tr>
<td>Structural aspects</td>
<td>Semantic aspects</td>
<td>Usage</td>
<td>Structural aspects</td>
</tr>
</tbody>
</table>

Basic elements

- Schema Tools
- Basic datatypes
- Links & media localization
- Basic Tools

User interaction

- Navigation & Access
  - Summaries
  - Views
  - Variations

- User Preferences
  - User History
  - Views

More Info:

- [www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm](http://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

---

**Creation Information**

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

---

**Media Format**

```xml
<MediaFormat>
  <Content href="urn:mpeg:mpeg7:cs:ContentCS:2001:1">
    <Name xml:lang="en">image</Name>
  </Content>
    <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>
```

---

**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….
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

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), ….

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).

Suitable to video shot boundary detection and indexing

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

- **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 – 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 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]

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)
- “sounds like”, “looks like”

MPEG-7 System: Search by Example
- Compare colour histograms of Target with those in the database
- XML instantiations Colour histogram of all stored images
- Request for Content HTTP
- Jpeg images
- MPEG-7 Database
- Streaming Media RTP/RTCP

MPEG-7 Database
- Query target
- Query Response
- List of matching content clip1.mp4, clip2.mp4, ...
- Media Server
- Request for Content RTSP
- Search Engine
- XML instantiation Colour histogram
- Jpeg images
- Request for Content HTTP
- MPEG-7 Database
- Streaming Media RTP/RTCP
- Search Engine
- Query target XML instantiation Colour histogram
- List of matching content img1.jpg, img2.jpg, ...
- Media Server
- Request for Content HTTP
- JPEG Images
- Streaming Media HTTP
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](http://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/is203/](http://www.sims.berkeley.edu/academics/courses/is202/is203/)
6. MPEG-7 Multimedia Content Description Standard, John R. Smith, Pervasive Media Management Group, IBM, January 8, 2003