Outline

- Objective, goals, requirements and applications, basic components of the MPEG-7 standard
- Systems tools and Description Definition Language
- Multimedia Description Schemes
- Visual Tools
- Audio Tools
- Relation with other standards
Motivation

- 300 million photos uploaded per day

- 60 hours of video uploaded every minute,
- more than 3 billion views per day,
- over 3 billion hours of video watched each month
Motivation

- The multimedia context:
  - More information is in digital form and is on-line
  - AV content covers: still pictures, audio, speech, video, graphics, 3D models, etc.
  - AV content is available at all bitrates and on all networks.
  - Increasing number of multimedia applications, services.

- Necessity of describing content:
  - Increasing amount of information.
  - More needs to have “information about the content”.
  - Difficult to manage (find, select, filter, organize, etc) content.
  - User: human or computational systems.
Visual information retrieval

- Manual/automatic annotation
  - Assign keywords to an image
  - Multi-class classification by machine learning

- Content-based image retrieval (CBIR)
  - Based on the content of the image
  - Image analysis to extract (high-dimensional) feature vectors
    - Color, texture, shape
    - Motion, shot detection, camera operation
  - Low-level versus high-level features: bridging the semantic gap

- MPEG-7
  - Standardized content-based description
MPEG Standards

- **MPEG-1 – ISO/IEC 11172 (1993):**
  - Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s

- **MPEG-2 – ISO/IEC 13818 (1995):**
  - Generic coding of moving pictures and associated audio information

- **MPEG-4 – ISO/IEC 14496 (1999):**
  - Coding of audio-visual objects

- **MPEG-7 – ISO/IEC 15938 (2002):**
  - Multimedia content description interface

- **MPEG-21 – ISO/IEC 21000 (2001):**
  - Multimedia framework
- 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 representation pyramid

- MPEG-1, MPEG-2, and MPEG-4 represent the content itself (‘the bits’).
- MPEG-7 should represent information about the content (‘the bits about the bits’).

The pyramid illustrates the progression from pixel-based representation to semantic descriptor extraction, culminating in MPEG-7 as the ultimate level for content representation.
Objective of MPEG-7

- Standardize content-based description for various types of audiovisual information
  - Enable fast and efficient content searching, filtering and identification
  - Describe several aspects of the content (low-level features, structure, semantic, models, collections, creation, etc.)
  - Address a large range of applications (⇒ user preferences, universal media access)

Types of audiovisual information:
- Audio, speech
- Moving video, still pictures, graphics, 3D models
- Information on how objects are combined in scenes
Type of applications

Pull Applications: “Search /Browsing”
Example: Search engines for Internet and databases

Push Applications: “Filtering”
Example: Broadcast of video, Interactive TV

- Universal Multimedia Access: Adapt delivery to network / terminal characteristics
- Specialized Professional and Control Applications
Example of application areas

- Storage and retrieval of audiovisual databases (image, film, radio archives)
- Broadcast media selection (radio, TV programs)
- Surveillance (traffic control, surface transportation, production chains)
- E-commerce and Tele-shopping (searching for clothes / patterns)
- Remote sensing (cartography, ecology, natural resources management)
- Entertainment (searching for a game, for a karaoke)
- Cultural services (museums, art galleries)
- Journalism (searching for events, persons)
- Personalized news service on Internet (push media filtering)
- Intelligent multimedia presentations
- Educational applications
- Bio-medical applications
Example of queries

- **Text (keywords):**
  - Find AV material with subject corresponding to some keywords

- **Semantic description:**
  - Find AV material corresponding to a specified semantic

- **Image as an example:**
  - Find an image with similar characteristics (global or local)

- **A few notes of music:**
  - Find corresponding musical pieces or movies

- **Low level features (example: motion):**
  - Find video with specific object motion trajectories
Relation content / description

- Description may be separated from the content

- Description may be multiplexed with the content
Type of description

- Information about the content: recording date & conditions, title, author, copyright, coding format, classification etc.

- Information present in the content: combination of low level and high level descriptors

High level description:
- Efficient and powerful
- Lack of flexibility

Low level description:
- Generic and flexible
- Intelligent / efficient search engine

Diagram:
- Indexing Feature extraction
- MPEG-7 enabled Database
- Search Retrieval

Efficiency  |  High level Recognition process  |  Low level Recognition process  |  No restriction on the search
The description generation (feature extraction, indexing process, annotation & authoring tools,...) and consumption (search engine, filtering tool, retrieval process, browsing device, ...) are non normative parts of MPEG-7.

The goal is to define the minimum that enables interoperability (syntax and semantic of description tools).
MPEG-7: The Workplan

**Collaboration:**
- Common work
- Core experiments
- eXperimentation Model
- Requirements

**Competition:**
- Individual work
- Definition of the scope and requirements

Timeline:
- 1996: Call for proposals
- 1998: Working draft
- 1999: Committee draft
- 2000: Final committee draft
- 2001: Draft international standard, International standard
Information Flow

- **Feature extraction**: Manual / automatic
- **AV Description**
- **Search / query**: Pull
- **Browse**
- **Filter**
- **Storage**
- **Transmission**: Encoding (for transmission)
- **Decoding (for transmission)**

**User & computational systems**

The content and its description may also be multiplexed
MPEG-7 elements

- **Descriptors (D):**
  - to represent Features. Descriptors define the syntax and the semantics of each feature representation.

- **Description Schemes (DS):**
  - to specify the structure and semantics of the relationships between their components, which may be both Ds and DSs.

- **Description Definition Language (DDL):**
  - to allow the creation of new DSs and, possibly, Ds and to allows the extension and modification of existing DSs.

- **System tools:**
  - to support multiplexing of descriptions, synchronization of descriptions with content, transmission mechanisms, file format, etc.
MPEG-7 working areas

Descriptive Definition

Tags

Instantiation

Description Definition Language

Encoding & Delivery

Structuring

Descriptive Schemes

Description Schemes

Descriptors:
(Syntax & semantic of feature representation)
Parts of the MPEG-7 Standard

- Part 1: Systems
- Part 2: Description Definition Language
- Part 3: Visual
- Part 4: Audio
- Part 5: Multimedia Description Schemes
- Part 6: Reference Software
- Part 7: Conformance testing
- Part 8: Extraction and use of MPEG-7 descriptions (TR)
- Part 9: Profiles and levels
- Part 10: Schema definition
- Part 11: MPEG-7 profile schemas
- Part 12: Query format
Outline

- Objective, goals, requirements and applications, basic component of the MPEG-7 standard
- Systems tools and Description Definition Language
- Multimedia Description Schemes
- Visual Tools
- Audio Tools
- Relation with other standards
System tools

- **System tools:**
  - Two formats:
    - Textual: XML
    - Binary: Binary format for MPEG-7 data (BiM)
  - Rationale:
    - XML is human readable, but very verbose
    - Not appropriate for bandwidth efficient storage or transmission
    - BiM allows bandwidth efficient binary representation, dynamic update and flexible streaming
    - High compression ratio
    - MPEG-7 XML file and corresponding BiM stream result in identical descriptions
System tools

- Streaming and delivery:
  - Split the description in pieces
  - Encapsulate the pieces of description in “access units”
  - Transmit the access units
  - Dynamic description
System tools

Relation XML - BiM:

- Access Unit Textual Format
  - MPEG-7 Textual Encoder
  - MPEG-7 Textual Decoder
  - Content Description
  - Generate Canonical Representation
  - Canonical Equivalence
  - Generate Canonical Representation
  - Content Description

- Access Unit Binary Format
  - MPEG-7 Binary Encoder
  - MPEG-7 Binary Decoder
MPEG-7 terminal architecture
Mr Robert Smith
15 rue Lacepede
75005 Paris

Dear Mr Doe, .....
<letter>
  <header>
    <name>Mr Robert Smith</name>
    <address>
      <street>15 rue Lacepede</street>
      <city>Paris</city>
    </address>
  </header>
  <text>Dear Mr Doe, .....</text>
</letter>
<letter>
  <header>
    <name>Mr Robert Smith</name>
    <address>
      <street>15 rue Lapecepede</street>
      <city>Paris</city>
    </address>
  </header>
  
  <text>Dear Mr Doe, .....</text>
</letter>
Description Definition Language

- **Definition of the Ds and DSs:**
  - XML Schema + MPEG-7 extensions

- **Instantiation (description):**
  - XML

- **Allow to define new entities**

![Diagram showing the relationship between DDL, DS, and D entities. The diagram illustrates that some DS and D entities are in the standard, while others are defined with DDL. The DDL entity links to DS, which further links to D entities. Some D entities are marked as not in the standard and defined with DDL.](image-url)
DDL: Schema definition

**DDL**
- XML-oriented, extends XML Schema (W3C)
- define MPEG-7 data models
- can be used to extend MPEG-7 if needed

**XML Schema:**
- Datatypes, Simple and Complex types
- Elements, Inheritance, Abstract types

**MPEG-7 extensions: do not break an XML Schema parser**
- Array and Matrix datatype
- Enumerated datatypes for MimeType, CountryCode, RegionCode, CurrencyCode and CharacterSetCode
- Typed references
Outline

- Objective, goals, requirements and applications, basic component of the MPEG-7 standard
- Systems tools and Description Definition Language
- Multimedia Description Schemes
- Visual Tools
- Audio Tools
- Relation with other standards
Content Management & Description

- Format, Coding, Instances, Identification, Transcoding Hint etc. *(Several instances)*
- Title, Creator, Creation location & date, Purpose, Classification, Genre etc. *(Author generated)*
- Rights holder, Access rights, Usage Record, Financial aspects etc. *(Evolution)*

**Creation & production**

**Media**

**Content management**

**Content description**

**Structural aspects**

- Viewpoint of the structure: Segments
  - Spatial / temporal structure
  - Audio, video low-level Ds
  - Elementary semantic information.

**Conceptual aspects**

**Link & media localization**

**Basic DSs**
Examples of Segments

Video segments

Still regions

Moving regions

Audio segments
### Examples of Segments

<table>
<thead>
<tr>
<th>Mosaic</th>
<th>3D still region</th>
<th>Image text</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Mosaic Image" /></td>
<td><img src="image2.png" alt="3D still region Image" /></td>
<td><img src="image3.png" alt="Image text Image" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ink segment</th>
<th>Multimedia segment</th>
<th>Analytic clips/transitions</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image4.png" alt="Ink segment Image" /></td>
<td><img src="image5.png" alt="Multimedia segment Image" /></td>
<td><img src="image6.png" alt="Analytic clips/transitions Image" /></td>
</tr>
</tbody>
</table>
Specific description of segments

Video segments
- Color
- Camera motion
- Motion activity
- Mosaic

Still regions
- Color
- Shape
- Position
- Texture

Moving regions
- Color
- Motion trajectory
- Parametric motion
- Spatio-temporal shape

Audio segments
- Spoken content
- Spectral characterization
- Music: timbre, melody
Example of Segment trees

SR1:
- Creation/Usage meta information
- Media description
- Textual annotation
- Color histogram, Texture

SR2:
- Shape
- Color histogram
- Textual annotation
  No gap, no overlap

SR3:
- Shape
- Color histogram
- Textual annotation

SR4:
- Shape
- Color histogram
- Textual annotation

SR5:
- Shape
- Textual annotation

SR6:
- Shape
- Color histogram
- Textual annotation

SR7:
- Color histogram
- Textual annotation
  No gap, no overlap

SR8:
- Color histogram
- Textual annotation
Graph

**Goal:**
- The segment DS allows the construction of tree structures
  - Efficient for access, retrieval, compression.
  - Lack of flexibility
- Graph structure to improve flexibility

**Outline of the approach:**
- Definition of entity nodes representing segments
- Definition of relationships: space, time, visual
Video Segment 1: Dribble & Kick

- Player
- Ball
- Goalkeeper

Is composed of

- Moves toward
- Is close to

Video Segment 2: Goal Score

- Ball
- Player
- Goalkeeper
- Goal

Is composed of

- Moves toward
- Left of

Direction scientifique
Viewpoint of conceptual notions
- Events, objects, abstract concepts, and their relation
Navigation and Access

Efficient support of: discovery, browsing, navigation, visualization / sonification. Two navigation modes: Hierarchical & sequential

Substitution of the original content
Adaptation to terminal, network, or user preferences
Universal Multimedia Access: Adapt delivery to network and terminal characteristics (QoS)
**Content Organization**

- **Content organization**
  - Description and organization of collection of documents
  - Media
  - Structural aspects

- **Collection & Classification**

- **Analytic Model**
  - Analytic model and classifiers: Definition of cluster, classes and models to associate a semantic label to a set of data.
  - Probability Model: (Gaussian & State Transition)
  - Compact representation of features & descriptors.

- **Datatype & structures**
- **Schema tools**
- **Link & media localization**
- **Basic DSs**
Outline

- Objective, goals, requirements and applications
- Basic component of the MPEG-7 standard
- Systems tools and Description Definition Language
- Multimedia Description Schemes
- Visual Tools
- Audio Tools
- Relation with other standards
MPEG-7 Visual part contains 25 Ds/DSs

- Basic Elements (2)
- Localization (2)
- Containers (3)
- Color (7)
- Texture (3)
- Shape (3)
- Motion (4)
- Face (1)
Dominant Color(s):
- 1-8 dominant colors in image / region
- color space, quantization, dominant color(s) value(s)
- variance of color value, percentage of pixels of this color, spatial coherency of color repartition

Color Content (histogram):
- Color histogram in HSV color space, encoded by Haar transform
  - scalable in number of coefficients kept for representation
  - scalable in number of bits per coefficients
  - lower end: 60 bits, very fast matching
Color (2)

- **Color content + coherence of repartition:** ColorStructure
  - Histogram of structuring elements that contain a particular color
    - Enhanced retrieval (in conjunction with HMMD)

- **Color content + its layout:** ColorLayout
  - Based on the DCT coefficients. (size: about 160 bits)
    - Layout sensitive retrieval, sketch-to-image matching

- **Color content of Group of Pictures / Frames:** GoFGoPColor
  - Aggregation of color histograms (average, median, etc)
    - Clustering of data for browsing / retrieval
Characterization of homogeneous textures:
- Low-level: *HomogeneousTexture* retrieval
- High-level: *TextureBrowsing* browsing

Characterization of structures in generic images:
- edges content and layout: *EdgeHistogram*
**Texture (2)**

<table>
<thead>
<tr>
<th>HomogeneousTexture</th>
<th>TextureBrowsing</th>
</tr>
</thead>
<tbody>
<tr>
<td>In the frequency domain:</td>
<td></td>
</tr>
<tr>
<td>• Decomposition into 30 channels (5 scales, 6 angles) using Gabor filters</td>
<td></td>
</tr>
<tr>
<td>• Energy and energy deviation</td>
<td></td>
</tr>
<tr>
<td>• Main direction(s)</td>
<td></td>
</tr>
<tr>
<td>• Regularity (1 to 4)</td>
<td></td>
</tr>
<tr>
<td>• Coarseness (1 to 4)</td>
<td></td>
</tr>
</tbody>
</table>

**EdgeHistogram**

- Fixed size: 240 bits

- Edge types histograms on 16 sub-images
**Shape (1)**

- **2D:** *ContourShape and RegionShape*

- **3D:** *Shape3D*
### Contour Based: \textit{ContourShape}

Curvature Scale Space:
- curvature points importance and relative positions
- Variable size: < 15 Bytes

#### Shape3D
- Based on 3D meshes
- Histogram of 3D shape indexes (Koenderink) representing local curvature properties of the 3D surface

### Region Based: \textit{RegionShape}

- Angular Radial Transf. (ART) moments
- Fixed size: 17.5 Bytes
Motion (1)

**MotionActivity**
- browsing, repurposing

**CameraMotion**
- browsing, high level queries

**MotionTrajectory**
- retrieval, high level queries

**ParametricMotion**
- mosaic, retrieval
Motion Activity:
- Intensity of motion (1 to 5)
- Main direction(s)
- Spatial and temporal distribution

Camera Motion:

- Boom up
- Track left
- Dolly forward
- Dolly backward
- Track right
- Boom down
- Roll
- Pan right
- Pan left
- Tilt up
- Tilt down
- Pan right
- Pan left
- Roll
- Tilt up
- Tilt down
Motion Trajectory:

- Interpolations
- Keypoints

Queries:
- similarity
- high level

Parametric Motion:
- translational
- rotation/scaling
- affine
- planar perspective
- parabolic
Face

Face Characterization: FaceRecognition

- Size: 238 bits
- Based on eigenfaces (vector of 56*46 values, extracted from normalized faces)
  - 49 basis vectors which span the space of possible face vectors
  - projection of the face vector on the 49 eigenfaces
Outline

- Objective, goals, requirements and applications
- Basic component of the MPEG-7 standard
- Systems tools and Description Definition Language
- Multimedia Description Schemes
- Visual Tools
- Audio Tools
- Relation with other standards
Audio descriptors

- **Low level audio features:**
  - Waveform and spectrum envelops
  - Power, spectrum centroid, spectrum spread
  - Fundamental frequency, harmonicity
  - Independent spectral component representation

- **Spoken content**
  - Lattice of hypothesis

- **Music:**
  - Timbre, Melody
  - Genre

- **Silence description**
Use of description tools

- Library of tools!
- The description tools are presented on the basis of the functionality they provide.
- In practice, they are combined into meaningful sets of description units.
- Furthermore, each application will have to select a sub-set of descriptors and DSs.
- DDL can be used to handle specific needs of the application.
MPEG-7 XML image description

Unstructured news image
MPEG-7 XML image description

<StillRegion id = "news">
  Title
</StillRegion>
MPEG-7 XML image description

Spatial decomposition

<StillRegion id = “news”>
  <SegmentDecomposition
decompositionType = “spatial”>
    <StillRegion id = “background”>
      <StillRegion id = “speaker”>
        <StillRegion id = “topic”>
        </SegmentDecomposition>
    </StillRegion>
  </SegmentDecomposition>
</StillRegion>
MPEG-7 XML image description

<StillRegion id = “news”>
  <SegmentDecomposition
decompositionType = “spatial”>
    <StillRegion id = “background”>
      <DominantColor> 10 10 250 </DominantColor>
    </StillRegion>
    <StillRegion id = “speaker”>
      <StillRegion id = “topic”>
        </SegmentDecomposition>
    </StillRegion>
  </SegmentDecomposition>
</StillRegion>
<StillRegion id = “news”>
  <SegmentDecomposition decompositionType = “spatial”>
    <StillRegion id = “background”>
    <StillRegion id = “speaker”>
      <TextAnnotation>
        <FreeTextAnnotation> Journalist Judite Sousa </FreeTextAnnotation>
      </TextAnnotation>
      <SpatialMask>
        <Poly>
          <CoordsI> 80 288 100 200 ... 352 288 </CoordsI></Poly>
        </Poly>
      </SpatialMask>
      <StillRegion id = “topic”>
    </SegmentDecomposition>
  </StillRegion>
</StillRegion>
<StillRegion id="news">
  <SegmentDecomposition decompositionType="spatial">
    <StillRegion id="background">
      <DominantColor>10 10 250</DominantColor>
    </StillRegion>
    <StillRegion id="speaker">
      <TextAnnotation>
        <FreeTextAnnotation>Journalist Judite Sousa</FreeTextAnnotation>
      </TextAnnotation>
      <SpatialMask>
        <Poly>
          <CoordsI>5 25 10 20 15 15 10 10 5 15</CoordsI>
        </Poly>
      </SpatialMask>
    </StillRegion>
    <StillRegion id="topic">
      <TextAnnotation>
        <FreeTextAnnotation>Clinton’s affair</FreeTextAnnotation>
      </TextAnnotation>
    </StillRegion>
  </SegmentDecomposition>
</StillRegion>
MPEG-7 camera

The MPEG-7 camera describes a scene in terms of semantic objects and of their properties
• Image analysis: segmentation, change detection, and tracking (implemented on the camera DSP).
• MPEG-7 coder: scene description represented using MPEG-7 (XML).
• MPEG-7 decoder: MPEG-7 description is parsed. Extraction of the information related to the specific applications.
MPEG-7 camera

<!-- ################################################################## --!>
<!-- DDL output for object 4 -->
<!-- ################################################################## --!>
<Object id="4">
  <RegionLocator>
    <BoxPoly> Poly </BoxPoly>
    <Coords1> 237 222 </Coords1>
    <Coords2> 230 252 </Coords2>
    <Coords3> 240 286 </Coords3>
    <Coords4> 308 287 </Coords4>
    <Coords5> 312 284 </Coords5>
  </RegionLocator>
  <DominantColor>
    <ColorSpace> YUV </ColorSpace>
    <ColorValue1> 143.4 </ColorValue1>
    <ColorValue2> 123.3 </ColorValue2>
    <ColorValue3> 128.2 </ColorValue3>
  </DominantColor>
  <HomogeneousTexture>
    <TextureValue> 9.02 </TextureValue>
  </HomogeneousTexture>
  <MotionTrajectory>
    <TemporalInterpolation>
      <KeyFrame> 100 </KeyFrame>
      <KeyPos> 268.6 251.7 </KeyPos>
      <KeyFrame> 101 </KeyFrame>
      <KeyPos> 262.8 241.0 </KeyPos>
      ...
      <KeyFrame> 138 </KeyFrame>
      <KeyPos> 192.9 79.0 </KeyPos>
    </TemporalInterpolation>
  </MotionTrajectory>
</Object>
MPEG-7 camera for video surveillance

😊 Privacy: in surveillance applications persons feel uncomfortable to be filmed. Only the behavior of the persons are transmitted.
😊 Checking intentions in surveillance: deduce intentions by studying how a person moves.
😊 Extract various statistics without revealing identity of people
Relation with other standards

- SMPTE
  - Material Exchange Format (MXF)
  - Metadata dictionary, KLV encoding
- European Broadcast Union
- Dublin Core Metadata Initiative
- W3C: XML Schema, NewsML etc.
- TV AnyTime Application
- JPEG
  - JPSearch
Conclusions

**MPEG-7:**
- AV content description for interoperable applications

**Description Definition Language:**
- XML Schema (flexibility) + Binary version (efficiency)

**Description Schemes and Descriptors:**
- Library of description tools
- Covers a wide range of generic needs
- Structural aspects close to Signal / Image Processing (segment trees and graph).
- Low-level Descriptors characterize Segments.
Further information

- Major MPEG-7 documents are public:
  - MPEG Home page: http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm
  - Public documents: http://mpeg.chiariglione.org/working_documents.htm
  - Also check:
    - http://www.m4if.org/resources.php#Section40