Overview of MPEG-7

Dr Junqing Yu
Huazhong University of Sci & Tech
10/29/2016

Outline of contents

- Introduction
- Basic Components
- Content Description
- Audiovisual (AV) Descriptions
- Multimedia Description Schemes
- XM and Applications
- More Information

Terms

MPEG-7 – Multimedia Content Description Interface
MPEG-7 Standard No. ISO/IEC 15938

From MPEG-1 to MPEG-7

- MPEG-3, ever defined, but abandoned
- MPEG-5 and -6, not defined

Why is MPEG-7 needed

- Digital audiovisual information increasing
  - more and more available contents
  - all kinds of sources of information
- Use of the digital audiovisual information
  - description of the contents
  - fast search of the contents
Why do we need MPEG-7?

**Need**
- Fast & Accurate Access
- Personalized Content Production and Consumption
- Content Management
- Automation
- Visual
- Audio
- Sketch

**Support for Advanced Query**

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

**Scope of MPEG-7**
- 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.

**MPEG-7 Normative Interfaces**

**Abstract representation of possible applications using MPEG-7**

- Feature Extraction
- MPEG-7 Description
- Search Engine
- MPEG-7 Scope:
  - Description Schemes (DSs)
  - Descriptors (Ds)
  - Language (DDL)
  - Ref: MPEG-7 Concepts
- Abstract representation of possible applications using MPEG-7
Example: Content description

MPEG-7 Systems
- Defines
  - the terminal architecture and the normative interfaces.
  - how descriptors and description schemes are stored, accessed and transmitted
  - tools that are needed to allow synchronization between content and descriptions

Reference Software: the XM
- XM implements
  - MPEG-7 Descriptors (Ds)
  - MPEG-7 Description Schemes (DSs)
  - Coding Schemes
  - DDL

MPEG-7 Conformance
- Includes the guidelines and procedures for testing conformance of MPEG-7 implementations

Outline of contents
- Introduction
- Basic Components
- Content Description
- Audiovisual (AV) Descriptions
- Multimedia Description Schemes
- XM and Applications
- More Information
Main elements of MPEG-7

- **Descriptors (D)**: representations of features, that define the syntax and the semantics of each feature representation (low-level).
- **Description Schemes (DS)**: that specify the structure and semantics of the relationships between their components, which may be both Ds and DSs (high-level).
- **A Description Definition Language (DDL)**: based on XML Schema, to allow the creation of new DSs and Ds, and to allow the extension and modification of existing DSs.
- **System tools**: to support multiplexing of descriptions, synchronization issues, transmission mechanisms, coded representations, management and protection of intellectual property.

Description Definition Language

- **Description Definition Language (DDL)** is a language that define what description is valid, and allows the creation of new Description Schemes and Descriptors. It also allows the extension and modification of existing Description Schemes.
  - DDL is used to define a set of formal rules:
    - ordering of the elements
    - occurrences of elements
  - XML + MPEG-7 extensions

XML Basics

- **XML** - eXtensible Markup Language
  - Not a marking language
  - Used to define a set of rules, a meta-language

XML Basics Example

```xml
<?xml version="1.0"?>
<!--一个简单的XML文档-->
<message>
  <to>Student</to>
  <from>Teacher</from>
  <subject>Introduction to XML</subject>
  <body>Welcome to XML!</body>
</message>
```

XML: Base for DDL

- **Why choose XML as the base for the DDL?**
  - The popularity of XML
  - The interoperability with other standards in the future
- **Why XML should be extended for MPEG-7?**
  - SGML > XML
  - Structural extensions
  - Datatype extensions

DDL parser

DD license is a software to check if a description is valid

Integration of MPEG-7 into XML

```xml
<seq begin="20s" dur="10s">
  <img id="Image1" dur="5s">
    <MP7: annotation>
      <Who>Fernando Morientes</Who>
      <WhatAction>Spain vs. Sweden soccer match</WhatAction>
    </MP7: annotation>
  </img>
  <img id="Image2" dur="2s"/>
</seq>
```
Descriptor

- **Definition**
  A Descriptor (D) is a representation of a Feature. A Descriptor defines the syntax and the semantics of the Feature representation.

- **Notes**
  A descriptor allows an evaluation of the corresponding feature via the descriptor value. It is possible to have several descriptors representing a single feature.

- **Examples**
  For example, for the color feature, possible descriptors are: the color histogram, the average of the frequency components, the motion field, the text of the title, etc.

Descriptor Value

- **Definition**
  A Descriptor Value is an instantiation of a Descriptor for a given data set (or subset thereof).

- **Notes**
  Descriptor Values are combined via the mechanism of a Description Scheme to form a Description.

Descriptor Example

```xml
<VisualDescriptor xsi:type="DominantColorType">
  <SpatialCoherency>31</SpatialCoherency>
  <Value>
    <Percentage>31</Percentage>
    <Index>255 0</Index>
    <ColorVariance>0 0</ColorVariance>
  </Value>
</VisualDescriptor>
```

Description Scheme

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

- **Examples**
  A movie, structured as scenes and shots, including some textual descriptors at the scene level, and color, motion and some audio descriptors at the shot level.

- **Note**
  Ds contain only basic data types, and does not refer to others D or DSs.

Example

DS: XML Scheme & Extensions

- **XML Scheme**
  - Data types
  - Simple and Complex types
  - Elements
  - Inheritance, Abstract types

- **MPEG-7 extensions**
  - Array and Matrix datatype
  - Enumerated datatypes for MimeType, CountryCode, RegionCode, CurrencyCode and CharacterSetCode
  - Typed references
Description Scheme Example

< DSType name="MovingRegion">
  <attribute name="TemporalConnectivity" type="Segment"/>
  <attribute name="SpatialConnectivity" type="boolean"/>
  <attribute name="ColorSpace" minOccurs="0"/>
  <attribute name="ColorQuantization" minOccurs="0"/>
  <attribute name="DominantColor" minOccurs="0"/>
  <attribute name="GofGopColorHistogram" minOccurs="0"/>
  <attribute name="MotionTrajectory" minOccurs="0"/>
  <attribute name="ParametricMotion" minOccurs="0"/>
  <attribute name="MotionActivity" minOccurs="0"/>
</ DSType>

Relations of main elements

Description Scheme Example

Illustration of descriptions

Outline of contents

- Introduction
- Basic Components
- Content Description
- Audiovisual (AV) Descriptions
- Multimedia Description Schemes
- XM and Applications
- More Information

Type of descriptions

- Low level description (features, etc)
  - Generic and flexible
  - Intelligent / efficient search engine
- High level description (structures, concepts, etc)
  - Efficient and powerful
  - Lack of flexibility
Low-level Description

- Information in the creation and production processes
  - director, title, short feature movie
- Information related to the usage of the content
  - copyright pointers, usage history, broadcast schedule
- Information on the storage features of the content
  - storage format, encoding
- Information about low-level features in the content
  - colors, textures, sound timbres, melody

High-level Description

- Structural description
  - video segments, frames, still and moving regions, audio segments
  - Segment DS (representing the spatial, temporal or spatio-temporal structure)
- Conceptual (semantic) description
  - objects, events, and notions
  - links of the two descriptions

Basic description

- Elements
  - Information containers
  - containing data and other elements
  - <city> …… </city>
- Attributes
  - Attribute-value pairs used to characterize elements
  - <city population="10000"> …… </city>

Structured descriptions

- Structured descriptions are trees
- Trees are suitable for retrieval and search

Example: Audio description

<Mpeg7Main>
  <DescriptionMetadata>
    <Version>1.0</Version>
  </DescriptionMetadata>
  <ContentDescription>
    <AudioContent xs1:type="AudioType">
      <Audio>
        <CreationInformation>
          <Creation>
            <Title>The daily news</Title>
          </Creation>
        </CreationInformation>
      </Audio>
    </AudioContent>
  </ContentDescription>
</Mpeg7Main>
Outline of contents
- Introduction
- Basic Components
- Content Description
- Audiovisual (AV) Descriptions
- Multimedia Description Schemes
- XM and Applications
- More Information

Low level AV descriptors
- Video segments
  - Color
  - Camera motion
  - Motion activity
  - Music
- Still regions
  - Color
  - Shape
  - Position
  - Texture
- Moving regions
  - Color
  - Motion trajectory
  - Parametric motion
  - Spatio-temporal shape
- Audio segments
  - Spoken content
  - timber

Audio Framework

Audio description
- Low-level Description
  - spectrum, parametric, and temporal features
- High-level Description
  - Audio signature Description Scheme
  - Instrument timbre Description Schemes
  - The melody Description Tools
  - Sound recognition and indexing Description Tools
  - Spoken Content Description Tools

Audio Frame & Segment

Structural types in Audio Framework
- AudioSegmentType
- AudioType
- AudioDSType
- AudioLLDScalarType
- AudioLLDVectorType
- ScalableSeriesType
- SeriesOfVectorType
An illustration of the scalable series

<table>
<thead>
<tr>
<th>original series</th>
<th>***************</th>
</tr>
</thead>
<tbody>
<tr>
<td>scaled series</td>
<td>0 0 0 0 0 0 0 0</td>
</tr>
</tbody>
</table>

- k (index): 1 2 3 4 5 6 7 8 9 10 11 12 13
- ratio: 2 6 1 2
- numOfElements: 3 2 2 6
- totalNumOfSamples: 31

Audio descriptor: Basic

- Two basic audio Descriptors
  - AudioWaveform Descriptor
    - describes the audio waveform envelope (minimum and maximum)
  - AudioPower Descriptor
    - describes the temporally-smoothed instantaneous power

AudioWaveform Descriptor

```xml
<complexType name="AudioWaveformType">
  <extension base="mpeg7:AudioLDScalarType">
    <attribute name="minRange" type="float" use="optional"/>
    <attribute name="maxRange" type="float" use="optional"/>
  </extension>
</complexType>
```

AudioPower Descriptor

```xml
<complexType name="AudioPowerType">
  <extension base="mpeg7:AudioLDScalarType"/>
</complexType>
```

Audio descriptor: Basic Spectral

- AudioSpectrumEnvelope Descriptor
  - describes the short-term power spectrum
- AudioSpectrumCentroid Descriptor
  - describes the center of gravity of the log-frequency power spectrum
- AudioSpectrumSpread Descriptor
  - describing the second moment of the log-frequency power spectrum
- AudioSpectrumFlatness Descriptor
  - describes the flatness properties of the spectrum

Figure: AudioSpectrumEnvelope description of a pop song. The required data storage is NM values where N is the number of spectrum bins and M is the number of time points
Figure A 10-basis component reconstruction showing most of the detail of the original spectrogram including guitar, bass guitar, hi-hat and organ notes. The left vectors are an AudioSpectrumBasis Descriptor and the top vectors are the corresponding AudioSpectrumProjection descriptor. The required data storage is \(10(M+N)\) values.

Audio Signature Description

- AudioSignature Description Scheme provides a unique content identifier for the purpose of robust automatic identification of audio signals.
- Applications include:
  - audio fingerprinting
  - identification of audio
  - locating metadata for legacy audio content

Instrument Timbre Description

- Timbre is defined as the perceptual features that make two sounds having the same pitch and loudness sound different.
- Timbre Description describes the perceptual features with a reduced set of Descriptors:
  - HarmonicInstrumentTimbre Descriptor
  - LogAttackTime Descriptor
  - PercussiveInstrumentTimbre Descriptor
  - Combination with Basic Spectral Descriptors

Melody Description Tools

The melody Description Tools is to facilitate efficient, robust, and expressive melodic similarity matching.

- MelodyContour Description Scheme
  - 5-step contour representation
  - basic rhythmic information representation
- MelodySequence Description Scheme
  - supporting an expanded descriptor set and high precision of interval encoding
**Visual description**

- Color Descriptors
- Texture Descriptors
- Shape Descriptors
- Motion Descriptors for Video

**Basic Structures**

- Grid layout
- Time series
  - RegularTimeSeries
  - IrregularTimeSeries
- Multiple view
- Spatial 2D coordinates
- Temporal interpolation.

**Color Descriptors**

- Scalable Color Descriptor
  - A color histogram in HSV color space
  - Encoded by Haar Transform

**Dominant Color Descriptor**

- Clustering colors into a small number of representative colors
- It can be defined for each object, regions, or the whole image
- \[ F = \{ \{ c_i, p_i, v_i \}, s \} \]
  - \( c_i \): Representative colors
  - \( p_i \): Their percentages in the region
  - \( v_i \): Color variances
  - \( s \): Spatial coherency

**Dominant Color Descriptor**
**Color Layout Descriptor**
- Clustering the image into 64 (8x8) blocks
- Deriving the average color of each block (or using DCD)
- Applying DCT and encoding
- Efficient for
  - Sketch-based image retrieval
  - Content Filtering using image indexing

**Color Structure Descriptor**
- Scanning the image by an 8x8 pixel block
- Counting the number of blocks containing each color
- Generating a color histogram (HMMD)
- Main usages:
  - Still image retrieval
  - Natural images retrieval

**GoF/GoP Color Descriptor**
- Extends Scalable Color Descriptor
- Generates the color histogram for a video segment or a group of pictures
- Calculation methods:
  - Average
  - Median
  - Intersection

**Texture Descriptors**
- Homogenous Texture Descriptor
- Non-Homogenous Texture Descriptor (Edge Histogram)

**Homogenous Texture Descriptor**
- Partitioning the frequency domain into 30 channels (modeled by a 2D-Gabor function)
- Computing the energy and energy deviation for each channel
- Computing mean and standard variation of frequency coefficients
- $F = \{f_{DC}, f_{DP}, e_1, ..., e_{30}, d_1, ..., d_{30}\}$
- An efficient implementation:
  - Radon transform followed by Fourier transform

**2D-Gabor Function**
- It is a Gaussian weighted sinusoid
- It is used to model individual channels
- Each channel filters a specific type of texture
Radon Transform
- Transforms images with lines into a domain of possible line parameters
- Each line will be transformed to a peak point in the resulted image

Non-Homogenous Texture Descriptor
- Represents the spatial distribution of five types of edges
  - vertical, horizontal, 45°, 135°, and non-directional
- Dividing the image into 16 (4x4) blocks
- Generating a 5-bin histogram for each block
- It is scale invariant

Region-based Descriptor
- Expresses pixel distribution within a 2-D object region
- Employs a complex 2D-Angular Radial Transformation (ART)
- Advantages:
  - Describes complex shapes with disconnected regions
  - Robust to segmentation noise
  - Small size
  - Fast extraction and matching

Region-based Descriptor (2)
- Applicable to figures (a) – (e)
- Distinguishes (i) from (g) and (h)
- (j), (k), and (l) are similar

Shape Descriptors
- Region-based Descriptor
- Contour-based Shape Descriptor
- 2D/3D Shape Descriptor
- 3D Shape Descriptor
Contour-Based Descriptor
- It is based on Curvature Scale-Space representation

Curvature Scale-Space
- Finds curvature zero crossing points of the shape’s contour (key points)
- Reduces the number of key points step by step, by applying Gaussian smoothing
- The position of key points are expressed relative to the length of the contour curve

Curvature Scale Space (2)

Contour-Based Descriptor
- It is based on Curvature Scale-Space representation
- Advantages:
  - Captures the shape very well
  - Robust to the noise, scale, and orientation
  - It is fast and compact

Contour-Based Descriptor (2)
- Applicable to (a)
- Distinguishes differences in (b)
- Find similarities in (c) - (e)

Comparison
- Blue: Similar shapes by Region-Based
- Yellow: Similar shapes by Contour-Based
2D/3D Shape Descriptor
- A 3D object can be roughly described by snapshots from different angles
- Describes a 3D object by a number of 2D shape descriptors
- Similarity Matching: matching multiple pairs of 2D views

3D Shape Descriptor
- Based on Shape spectrum
- An extension of Shape Index (A local measure of 3D Shape to 3D meshes)
- Captures information about local convexity
- Computes the histogram of the shape index over the whole 3D surface

Motion Descriptors
- Motion Activity Descriptors
- Camera Motion Descriptors
- Motion Trajectory Descriptors
- Parametric Motion Descriptors

Motion Activity Descriptor
- Captures ‘intensity of action’ or ‘pace of action’
- Based on standard deviation of motion vector magnitudes
- Quantized into a 3-bit integer [1, 5]

Camera Motion Descriptor
- Describes the movement of a camera or a virtual view point
- Supports 7 camera operations

Motion Trajectory
- Describes the movement of one representative point of a specific region
- A set of key-points (x, y, z, t)
- A set of interpolation functions describing the path
Parametric Motion

- Characterizes the evolution of regions over time
- Uses 2D geometric transforms
- Example:
  - Rotation/Scaling:
    - $D_x(x,y) = a + bx + cy$
    - $D_y(x,y) = d - cx + by$

Outline of contents

- Introduction
- Basic Components
- Content Description
- Audiovisual (AV) Descriptions
- Multimedia Description Schemes
- XM and Applications
- More Information

Multimedia DSs

Multimedia Description Schemes are metadata structures for describing and annotating audio-visual (AV) content

- Basic Elements
- Content Management
- Content Description
- Content Organization
- Navigation and Access
- User Interaction

Organization of Multimedia DSs

Basic Element

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

Basic elements of DS

- Basic data types
  - a set of extended data types
  - vectors and matrices
- Constructs for linking media files
- Localizing pieces of content
- Describing
  - time, places, persons, individuals, groups, organizations, and textual annotation, etc
Schema tools

- Base types
- Root Element
- Top-level types
- Multimedia Content Entities
- Packages
- Description Metadata

Base Types

- Mpeg7
- Image Type
- Video Type
- Audio Type
- AudioVisual Type
- Multimedia Type
- MultimediaCollection type
- MultimediaProgramType
- Signal Type
- ElectronicInkType
- VideoEditing Type

Root Element

Complete Description
- via ContentDescription
- Description Unit
- via DescriptionUnit
- VideoSegment
- DominantColor

Top-level types

Multimedia Content Entities

Packages

- Associate Relations
  - Package descriptions
  - related description components into folders
  - Query Tool
  - Package descriptions often interactive queries
  - Signaling of what description elements are available for querying
  - A/V content description
  - A/V content description

- Browse
  - query
  - signal
  - Package description
- DB

- A/V content description
- A/V content description

- DB
  - Package description
  - A/V content description
  - A/V content description
Description Metadata

Basic datatypes

- defines datatypes that represent different kinds of constrained types.
  - Integer
  - Real
  - Matrix
  - String
  - countryCode

Links & media localization

- References datatype
  - refer to a part of the description
- Unique Identifier
  - allows the identification of the multimedia or other media content under description.
- Time description tools
  - YYYY-MM-DDThh:mm:ss:nnnFNNN+hh:mm
- Media localization tools

Time description tools

Content Management

- Creation and production information
  - Creation information
    - title, textual annotation, creators, and dates
  - Classification information
    - genre, subject, purpose, language
- Media coding, storage and file formats
  - format, compression, and coding
- Content usage
  - usage rights, usage record

Model for content, profile and instance
Content Description

• Structural aspects
• Semantics aspects

Structural aspects

• Segment entity description tools
• Segment attribute description tools
• Segment decomposition tools
• Segment relation description tools

Examples: T/S segments

Segment decomposition tools

Examples of Segment Decompositions: a) and b) Segment Decompositions without gap or overlap; c) and d) Segment Decompositions with gap or overlap.
Segment relation description tools

- Hierarchical Segment Tree
- Graph

Example: Segment trees

Example: Graph

Semantic aspects

- Semantic Entity
- Semantic Attribute
- Semantic Relation
Navigation and Access

- Summaries
  - hierarchical summaries
  - sequential summaries

- View, Partitions and Decompositions
  - decompositions in space, time and frequency
  - used in multi-resolution access and progressive retrieval

- Variations
  - selection of the most suitable of an AV program
  - adapt to the different capabilities of terminal devices, network conditions or user preferences

Hierarchical summary

Sequential summary

Partitions and Decompositions
**Views**

**Illustration of variations**

- **Collections**
  - group the contents into clusters
  - describes statistics and models of the attribute values
  - describe relationships among collection clusters

- **Models**
  - model the attributes and features of AV content
  - Probability Model
    - specify statistical functions and structures
  - Analytic Model
    - specify semantic labels
    - specify the confidence
    - build classifiers

**Content Organization**

**User Interaction**

- **User Preference**
  - context dependency in terms of time and place
  - relative importance of different preferences
  - privacy characteristics of the preferences
  - preferences update by agent or user

- **Usage History**
  - history of actions
  - used to determine the user's preferences
Outline of contents

- Introduction
- Basic Components
- Content Description
- Audiovisual (AV) Descriptions
- Multimedia Description Schemes
- XM and Applications
- More Information

**eXperimentation Model (XM)**

- **Simulation platform for:**
  - Ds, DSs, CSs, DDL
- **XM applications:**
  - the server (extraction) applications
  - the client (search, filtering and/or transcoding) applications

*CS: Coding Schemes*

**The XM applications**

- **Extraction from Media**
  - all low-level Ds or DSs should have an application class of this type
- **Search & Retrieval Application**
- **Media Transcoding Application**
- **Description Filtering Application**

**Extraction from Media**

**Search and retrieval application**
Media transcoding application

First, from a media database two features are extracted. Then, basing on the first feature, relevant media files are selected from the media database. The relevant media files are transcoded basing on the second extracted feature.

Real world application

 MDB = media database, DDB = description database.

First, from a media database two features are extracted. Then, basing on the first feature, relevant media files are selected from the media database. The relevant media files are transcoded basing on the second extracted feature.

MPEG-7 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 nBio-medical applications
Push and Pull applications

- **Push applications**
  - Example: Search engines for internet and DBs
  - Advantage: Many search engines work on standardized descriptions
- **Pull applications**
  - Example: Broadcast of video, Interactive TV
  - Advantage: Intelligent agents filter standardized descriptions

Example: Pull application

Example: Push application

Example: 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

Outline of contents

- Introduction
- Basic Components
- Content Description
- Audiovisual (AV) Descriptions
- Multimedia Description Schemes
- XM and Applications
- More Information

MPEG-7 and other Standards

- MPEG-1, -2, and -4 are designed to represent the information itself, while MPEG-7 is meant to represent information about the information.
- MPEG-1, -2, and -4 make content available, while MPEG-7 allows you to find the content you need.
Ultimate ambition of MPEG-7

• To make the web as searchable for multimedia content as it is searchable for text today

• To improve the use of computer systems as easy as possible

MPEG-7 beyond

• To mould computers around human requirements and not humans around computer requirements

• To enable content disclosure based on facts, rather than on human annotations

• To find information by rich spoken queries, hand-drawn images and address what most people expect computers to be able to do

More Information on WWW

• http://www.chiariglione.org/mpeg/

• http://www.mpegif.org/

Conclusions on MPEG-7

• MPEG-7:
  – AV content description for interoperable application

• Description Definition Language:
  – XML Schema (flexibility) + Binary version (efficiency)

• Description Schemes:
  – Library of description tools
  – Covers a wide range of generic needs

Conclusion