MPEG-7 Standard

Janko Ćalić
janko@cs.bris.ac.uk
http://www.cs.bris.ac.uk/home/janko/city/

MPEG Standards

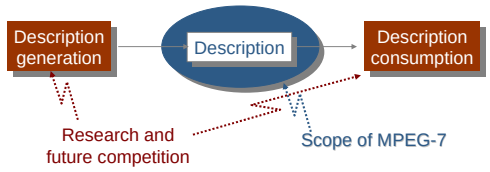
- **MPEG-1**:
  Storage of moving picture and audio on storage media (CD-ROM)
- **MPEG-2**:
  Digital television
- **MPEG-4**:
  Coding of natural and synthetic media objects for multimedia applications
- **MPEG-7**:
  Multimedia content description for AV material
- **MPEG-21**:
  Digital audiovisual framework: Integration of multimedia technologies (identification, copyright, protection, etc.)

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
- Descriptions independent of the data support

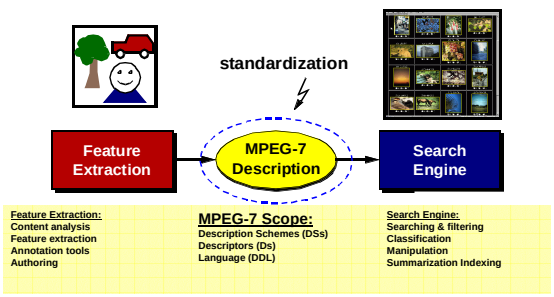
Scope of MPEG-7

- A standard for describing features of multimedia content



- 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 the description tools).

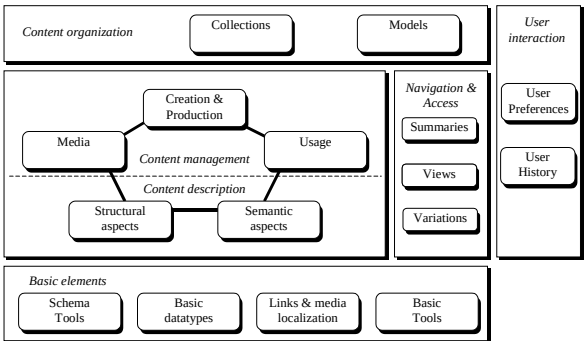
Scope of MPEG-7



Parts of MPEG-7 Standard

- ISO / IEC 15938 - 1: Systems
- ISO / IEC 15938 - 2: Description Definition Language
- ISO / IEC 15938 - 3: Visual
- ISO / IEC 15938 - 4: Audio
- ISO / IEC 15938 - 5: Multimedia Description Schemes
- ISO / IEC 15938 - 6: Reference Software

MPEG-7 Description Scheme



System tools & DDL

- System tools:
  - Streaming and delivery: Binarization of DDL and Error resilience
  - Access and synchronous consumption of description
  - Dynamic description
- Description Definition Language:
  - Definition of the Ds and DSs: XML Schema + MPEG-7 extensions
  - Instantiation (description): XML
  - Allow to define new entities

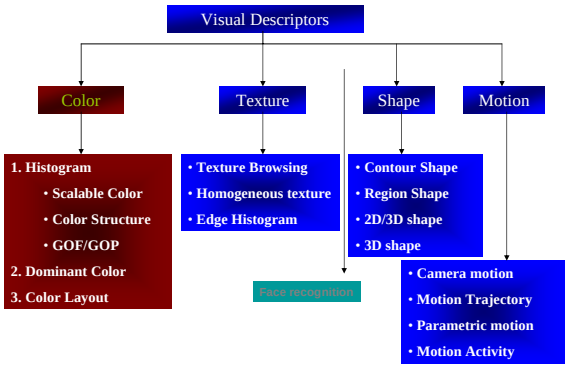
DDL: Schema definition

- XML Schema:
  - Datatypes
  - 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

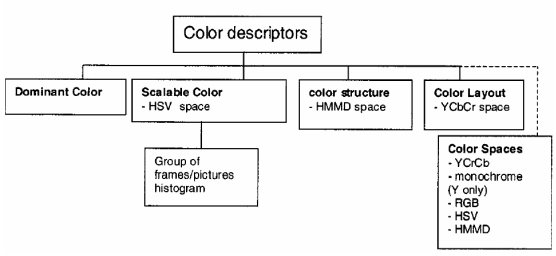
MPEG-7 Visual Descriptors

- Colour
  - Color Space, Color Quantization, Dominant Color, Scalable Color Color Layout,Color Structure, Group of Picture Color
- Texture
  - Homogeneous Texture, Texture Browsing, Edge histogram
- Shape
  - Region Shape, Contour Shape, Shape 3D
- Motion
  - Camera Motion, Motion Trajectory, Parametric Motion, Motion Activity
- Face Recognition, others

Visual Descriptors



Color Descriptors

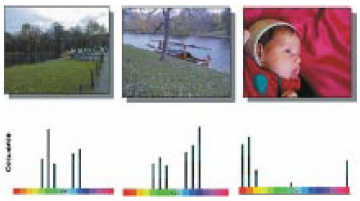


Color Spaces

- Constrained color spaces
  - Scalable Color Descriptor uses HSV
  - Color Structure Descriptor uses HMMD
- MPEG-7 color spaces:
  - Monochrome
  - RGB
  - HSV
  - YCrCb
  - HMMD

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

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

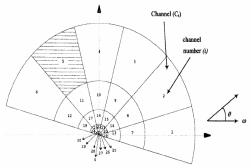
- Homogenous Texture Descriptor
- Edge Histogram

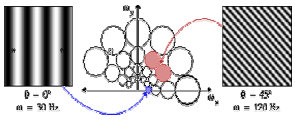
Homogenous Texture Descriptor

- 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
- $F = \{f_{DC}, f_{SD}, e_1,\ldots, e_{30}, d_1,\ldots, d_{30}\}$

2D-Gabor Function

- It is a Gaussian weighted sinusoid
- It is used to model individual channels
- Each channel filters a specific type of texture

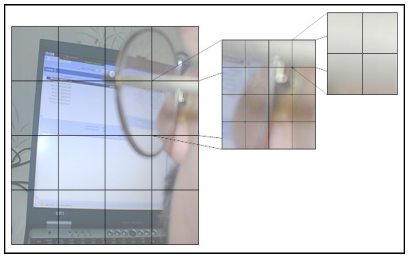


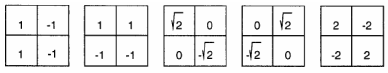


Edge Histogram

- 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

Edge Histogram





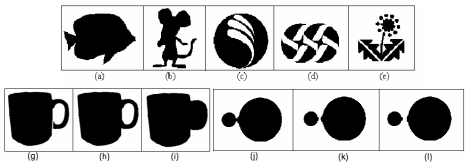
Shape Descriptors

- Region-based Descriptor
- Contour-based Shape Descriptor
- 2D/3D Shape Descriptor
- 3D Shape Descriptor

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

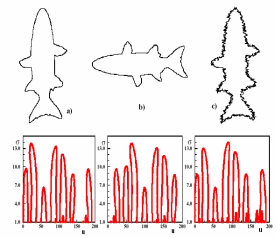
Region-based Descriptor (2)



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

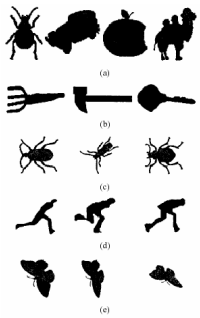
Contour-Based Descriptor

- Based on the 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
- Advantages:
  - Captures the shape very well
  - Robust to the noise, scale, and orientation
  - It is fast and compact

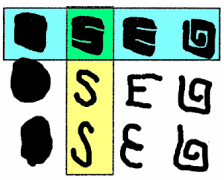


Contour-Based Descriptor (2)

- 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

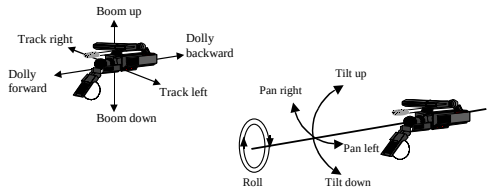
- Motion Activity Descriptors
- Camera Motion Descriptors
- Motion Trajectory Descriptors
- Parametric Motion Descriptors

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]

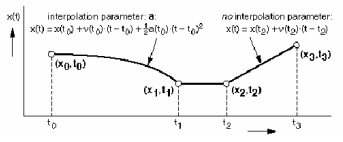
Camera Motion Descriptor

- 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:
    - $D_x(\mathbf{x},\mathbf{y}) = a + b\mathbf{x} + c\mathbf{y}$
    - $D_y(\mathbf{x},\mathbf{y}) = d - c\mathbf{x} + b\mathbf{y}$

References

1. T. Sikora, "The MPEG-7 Visual Standard for Content Description – An Overview", IEEE Trans. Circuits Syst. Video Technol., vol. 11, pp. 696-702, June 2001
2. S.-F. Chang, T.Sikora, and A. Puri, "Overview of MPEG-7 Standard", IEEE Trans. Circuits Syst. Video Technol., vol. 11, pp. 688-695, June 2001
3. J. M. Martinez, "Overview of the MPEG-7 Standard", ISO/IEC JTC1/SC29/WG1, 2001
4. B.S. Manjunath, J.-R. Ohm, V.V. Vasudevan, and A. Yamada, "MPEG-7 Color and Texture Descriptors", IEEE Trans. Circuits Syst. Video Technol., vol. 11, pp. 703-715, June 2001
5. M. Bober, "MPEG-7 Visual Shape Descriptors", IEEE Trans. Circuits Syst. Video Technol., vol. 11, pp. 716-719, June 2001
6. A. Divakaran, "An Overview of MPEG-7 Motion Descriptors and Their Applications", 9th Int. Conf. on Computer Analysis of Images and Patterns , CAIP 2001 Warsaw, Poland, 2001, Lecture Notes in Computer Science vol.2124, pp. 29-40
7. J. Hunter, "An overview of the MPEG-7 description definition language (DDL)", IEEE Trans. Circuits Syst. Video Technol., vol. 11, pp. 765-772, June 2001
8. F. Mokhtarian, S. Abbasi, and J. Kittler, "Robust and Efficient Shape Indexing through Curvature Scale Space", Proc. International Workshop on Image DataBases and MultiMedia Search, pp. 35-42, Amsterdam, The Netherlands, 1996

MPEG-7 Audio Descriptors

- Silence
  - SilenceType
- Spoken content (from speech recognition)
  - SpokenContentSpeakerType
- Timbre (perceptual features of instrument sounds)
  - InstrumentTimbreType, HarmonicInstrumentTimbreType, PercussiveInstrumentTimbreType
- Sound effects
  - AudioSpectrumBasisType, SoundEffectFeatureType
- Melody Contour
  - CountourType, MeterType, BeatType
- Description Schemes utilizing these Descriptors are also defined

Multimedia Description Schemes

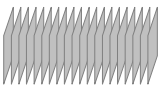
- **Content description**: representation of perceivable information
- **Content management**: information about the media features, the creation and the usage of the AV content;
- **Content organization**: representation the analysis and classification of several AV contents;
- **Navigation and access**: specification of summaries and variations of the AV content;
- **User interaction**: description of user preferences and usage history pertaining to the consumption of the multimedia material.

Reference Software: the eXperimentation Model

- XM software is the simulation platform for the MPEG-7 Descriptors (Ds), Description Schemes (DSs), Coding Schemes (CSs), and Description Definition Language (DDL)
- Besides the normative components, the simulation platform needs also some non-normative components, essentially to execute some procedural code to be executed on the data structures
- XM applications are divided in two types: the server (extraction) applications and the client (search, filtering and/or transcoding) applications

Low level Audio Visual descriptors

Video segments

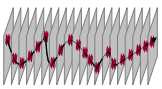


- Color
- Camera motion
- Motion activity
- Mosaic

Still regions

- Color
- Shape
- Position
- Texture

Moving regions

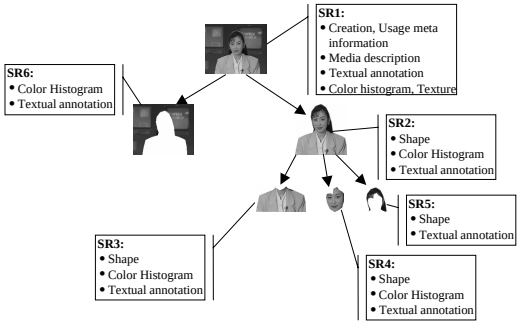


- Color
- Motion trajectory
- Parametric motion
- Spatio-temporal shape

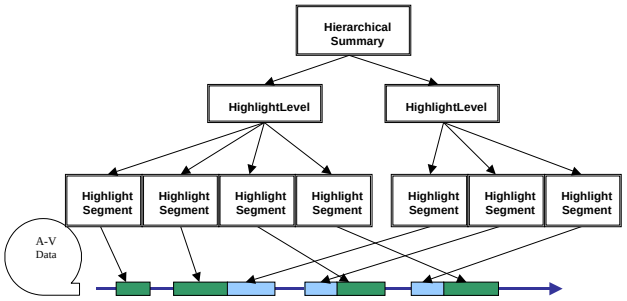
Audio segments

- Spoken content
- Spectral characterization
- Music: timbre, melody

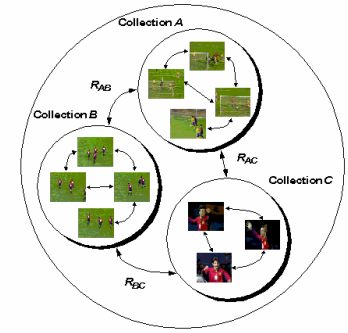
Example of Segment trees



Hierarchical summary



Collection Structure



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 Application areas

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