MPEG-7: Multimedia content description

Georgios Tziritas
Computer Science Department
http://www.csd.uoc.gr/~tziritas
Introduction

First release 2002 (ISO)

Audiovisual content description

Content-based multimedia retrieval

- images
- video
- 2D graphics
- 3D models
- audio
- speech

XML representation or binary coding
Definitions

- **Data** (image, audio, ...)
- **Feature** (color, timbre, motion, ...)
- **Descriptor** (color histogram, ...)
- **Descriptor value** (instantiation)
- **Description scheme** (structure and semantics)
- **Description** : (scheme and values)
- **Coded description**
- **Description Definition Language**

![Diagram showing the relationship between DDL, DS, and D nodes.](image)
Main components of the MPEG-7 Standard

Descriptive Definition Language

Instantiation

Tags

Encoding & Delivery

DS1

DS2

DS3

DS4

DS5

DS6

DS7

DS8

DS9

DS10

D1

D2

D3

D4

D5

D6

D7

D8

D9

D10

Descriptive Content

Structuring

Description Schemes

Tags

Encoding & Delivery

<scene id=1>
<time>....</time>
<camera>...</camera>
<annotation/>
</scene>

1 0 1 0 1

Descriptive Content

Definition

extension

Main components of the MPEG-7 Standard

Descriptors:
(Syntax & semantic of feature representation)
Complete description

- Multimedia Content (abstract)
  - Image
  - Video
  - Audio
  - Audio–Visual
  - Multimedia Collection
  - Signal
  - Ink Content
  - Analytic Edited Video

- Complete Description (abstract)

- Content Abstraction (abstract)
  - Semantic Description
  - Model Description
  - Summary Description
  - View Description
  - Variation Description

- Content Management (abstract)
  - User Description
  - Media Description
  - Creation Description
  - Usage Description
  - Classification Scheme Description
Description example

SR7:
- Color histogram
- Textual annotation

SR8:
- Color histogram
- Textual annotation

SR3:
- Shape
- Color histogram
- Textual annotation

SR6:
- Shape
- Color histogram
- Textual annotation

SR1:
- Creation, usage metainformation
- Media description
- Textual annotation
- Color histogram, texture

No gap, no overlap

SR2:
- Shape
- Color histogram
- Textual annotation

No gap, no overlap

SR5:
- Shape
- Textual annotation

SR4:
- Shape
- Color histogram
- Textual annotation
Color descriptors

- Dominant Color
  - HSV space
  - Group of frames/pictures histogram

- Scalable Color
  - HMMD space

- Color Structure
  - YCbCr space

- Color Layout
  - YCrCb
  - monochrome (Y only)
  - RGB
  - HSV
  - HMMD

Color similarity

![Color similarity diagram](image)
Color systems

RGB, YCbCr, Y (monochrome-intensity only)

\[ \text{Diff} = \text{Max} - \text{Min} \]
\[ \text{Sum} = (\text{max} + \text{min})/2 \]
Hue as defined for the HSV.

\[
\begin{align*}
\text{Max} &= \max(R, G, B); \quad \text{Min} = \min(R, G, B); \\
\text{Value} &= \max(R, G, B); \\
\text{if} (\text{Max} == 0) & \text{ then} \\
& \quad \text{Saturation} = 0; \quad \text{else} \\
& \quad \text{Saturation} = (\text{Max} - \text{Min})/\text{Max}; \\
\text{if} (\text{Max} == \text{Min}) & \text{ Hue is undefined (achromatic color);} \\
\text{otherwise:} \\
\text{if} (\text{Max} == R \, \&\& \, G > B) & \quad \text{Hue} = 60*(G-B)/(\text{Max}-\text{Min}) \\
\text{else if} (\text{Max} == R \, \&\& \, G < B) & \quad \text{Hue} = 360 + 60*(G-B)/(\text{Max}-\text{Min}) \\
\text{else} & \quad \text{Hue} = 60*(2.0 + (B-R)/(\text{Max}-\text{Min})) \\
\end{align*}
\]

Color histogram

256 Histogram values → Nonlinear Quantization → Haar Transform → Linear Quantization → Coefficients


**Color histogram**: HSV : 4+2+2 (11 bits/bin)

**Color histogram Haar transform**

**Extension to group of images**
Dominant colors

\[ F = \{\{c_i, p_i, v_i\}, s\}, \ i = 1, 2, ..., N \]

*k-means* algorithm

\( p_i \): 5 bits (percentage)

\( v_i \): 3 bits (variance)

Spatial coherency \((s)\):

average number of connecting pixels 3 x 3 (5 bits)
Color structure

Block 8 x 8
Local color histogram distribution HMMD

<table>
<thead>
<tr>
<th>Color</th>
<th>Bin Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0</td>
<td>h(0)+1</td>
</tr>
<tr>
<td>C1</td>
<td>h(1)</td>
</tr>
<tr>
<td>C2</td>
<td>h(2)+1</td>
</tr>
<tr>
<td>C3</td>
<td>h(3)+1</td>
</tr>
<tr>
<td>C4</td>
<td>h(4)</td>
</tr>
<tr>
<td>C5</td>
<td>h(5)</td>
</tr>
<tr>
<td>C6</td>
<td>h(6)+1</td>
</tr>
<tr>
<td>C7</td>
<td>h(7)</td>
</tr>
</tbody>
</table>
Color layout

Color system YCbCr
Any shape region

8 x 8 block after partition
Representative color for each block : average color
Discrete Cosine Transform
Zig-zag scanning
12 coefficients \((6+3+3)\)
Texture description

Texture browsing descriptor

- Regularity (2 bits)
- Directionality (2x3 bits)
- Coarseness (2x2 bits)

Homogeneous texture

Bank of orientation and scale sensitive filters

Gabor filters: 5 scales and 6 directions

\[ G_{P_s,r}(W, q) = \exp \left[ \frac{-(W - W_s)^2}{2S_{r_s}^2} \right] \cdot \exp \left[ \frac{-(q - q_r)^2}{2S_{q_r}^2} \right] \]

Texture (edge) description

Edges are broadly grouped into five categories: vertical, horizontal, 45 diagonal, 135 diagonal, and isotropic.

(a) 1 -1
    1 -1
(b) 1 1
    -1 -1

(c) $\sqrt{2}$ 0
    0 $-\sqrt{2}$
(d) 0 $\sqrt{2}$
    $-\sqrt{2}$ 0
(e) 2 -2
    -2 2
Shape descriptors

Bounding box

Region-based descriptor
  Angular Radial Transformation

Contour-based descriptor
  Curvature scale-space
  Curvature zero-crossings
  Prominent peaks of curvature

3D surface descriptor
  Shape index from principal curvatures
  Shape index histogram
  Planar regions
Motion descriptors / activity

Motion activity

<table>
<thead>
<tr>
<th>Activity Value</th>
<th>Range of $\sigma$ (Std. Dev. of motion vector magnitude)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$0 \leq \sigma &lt; 3.9$</td>
</tr>
<tr>
<td>2</td>
<td>$3.9 \leq \sigma &lt; 10.7$</td>
</tr>
<tr>
<td>3</td>
<td>$10.7 \leq \sigma &lt; 17.1$</td>
</tr>
<tr>
<td>4</td>
<td>$17.1 \leq \sigma &lt; 32$</td>
</tr>
<tr>
<td>5</td>
<td>$32 \leq \sigma$</td>
</tr>
</tbody>
</table>

Intensity of activity
Direction of activity
Spatial distribution
Temporal distribution
Motion descriptors

Camera motion

- Video Segment
  - Camera Motion
  - Motion Activity
- Mosaic
  - Warping Parameters
- Moving Region
  - Trajectory
  - Parametric Motion

- Boom up
- Dolly forward
- Track left
- Track right
- Boom down

- Roll
- Pan right
- Tilt up
- Pan left
- Tilt down
Motion descriptors / mosaic

Warping parameters

\[
x'_i = \frac{a_0 + a_1 x_i + a_2 y_i}{a_6 x_i + a_7 y_i + 1}
\]

\[
y'_i = \frac{a_3 + a_4 x_i + a_5 y_i}{a_6 x_i + a_7 y_i + 1}
\]
Motion trajectory describes the displacements of objects in time, objects being defined as spatio-temporal regions whose trajectories are important for the given application. The trajectory model is a first- or second-order piecewise approximation along time. The core of the description is a set of keypoints.
MPEG-7 sets a standard for describing multimedia content, in such a way that it can be managed, searched, filtered and identified in a quick and efficient way.

It does not address one application area in particular, but rather supports a wide range of applications.

There can be several descriptions for a single piece of multimedia content, all valid for a particular application or user.

To allow the necessary interworking in combination with the freedom of building competitive products, MPEG-7 only specifies the description tools themselves.

MPEG-7 descriptors are extracted from images or video sequences using suitable extraction methods and can be stored or transmitted entirely separate from the media content. The descriptors allow users or agents (or search engines) to evaluate similarity in images or video based on color, texture, object shape, global motion, or object motion features.