MPEG-7 : Multimedia content description

Visual content

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

Introduction

First release 2002 (ISO)

Audiovisual content description

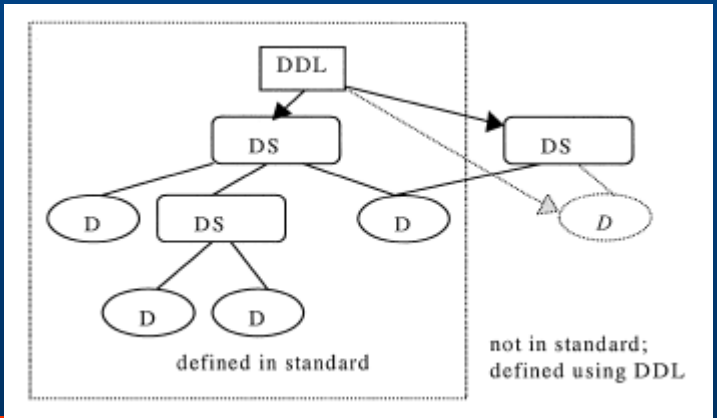
Content-based multimedia retrieval

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

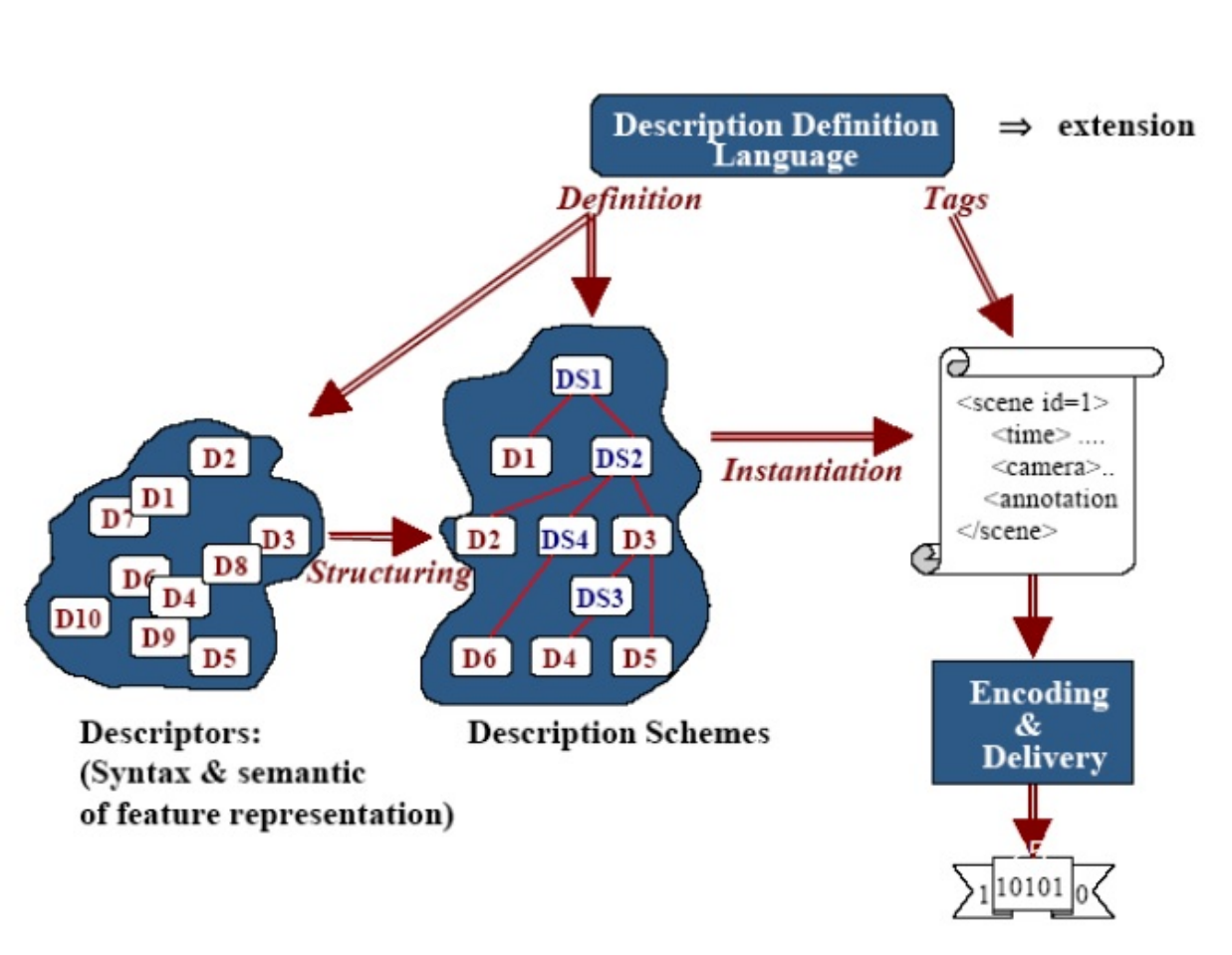
XML representation or binary coding

Definitions

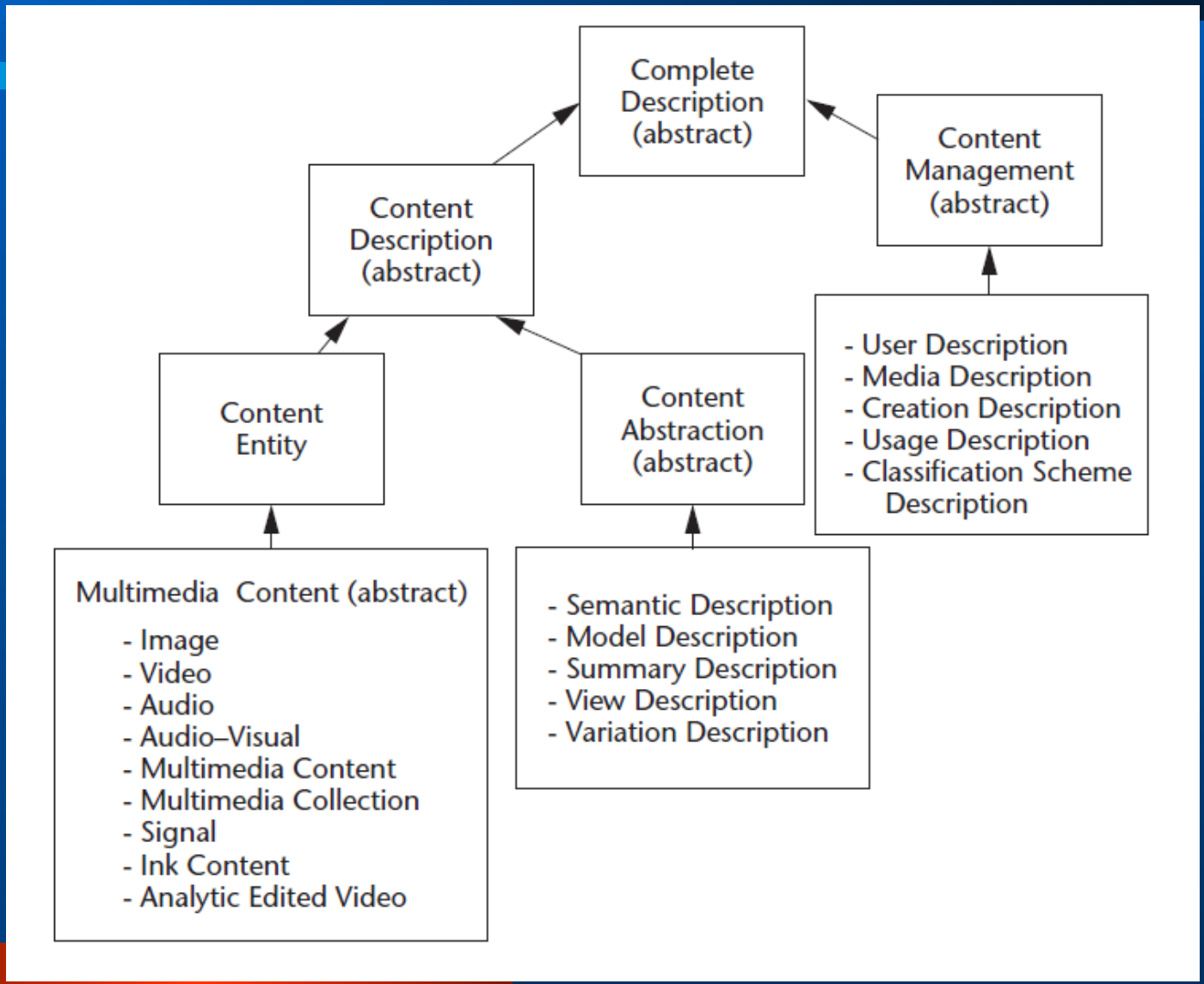
- **Data** (image, audio, ...)
- **Feature** (color, timbre, motion, ...)
- **Descriptor** (color histogram, ...)
- **Descriptor value** (instantiation)
- **Description scheme** (structure and semantics)
- **Description** : (scheme and values)
- **Coded description**
- **Description Definition Language**



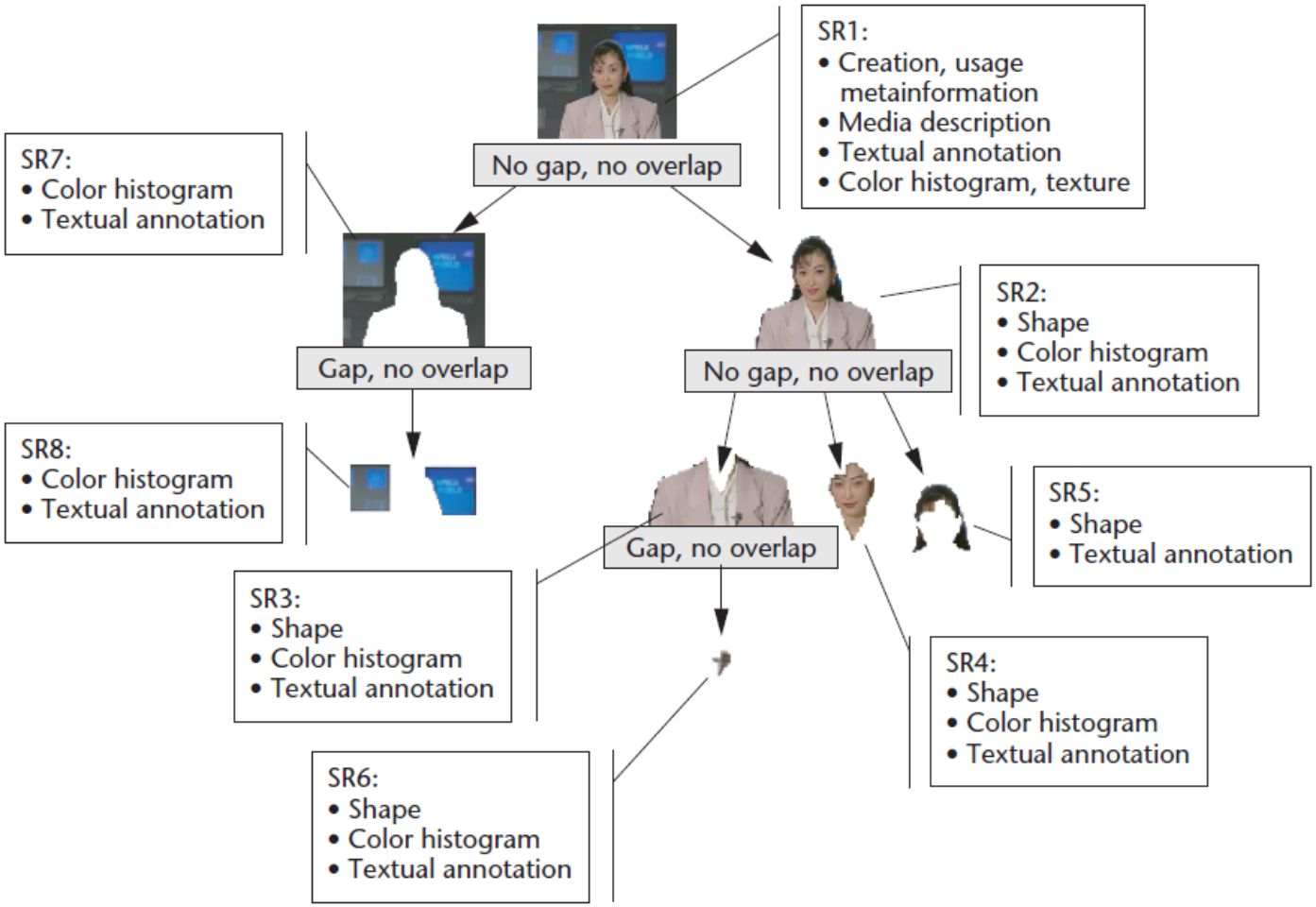
Main components of the MPEG-7 Standard



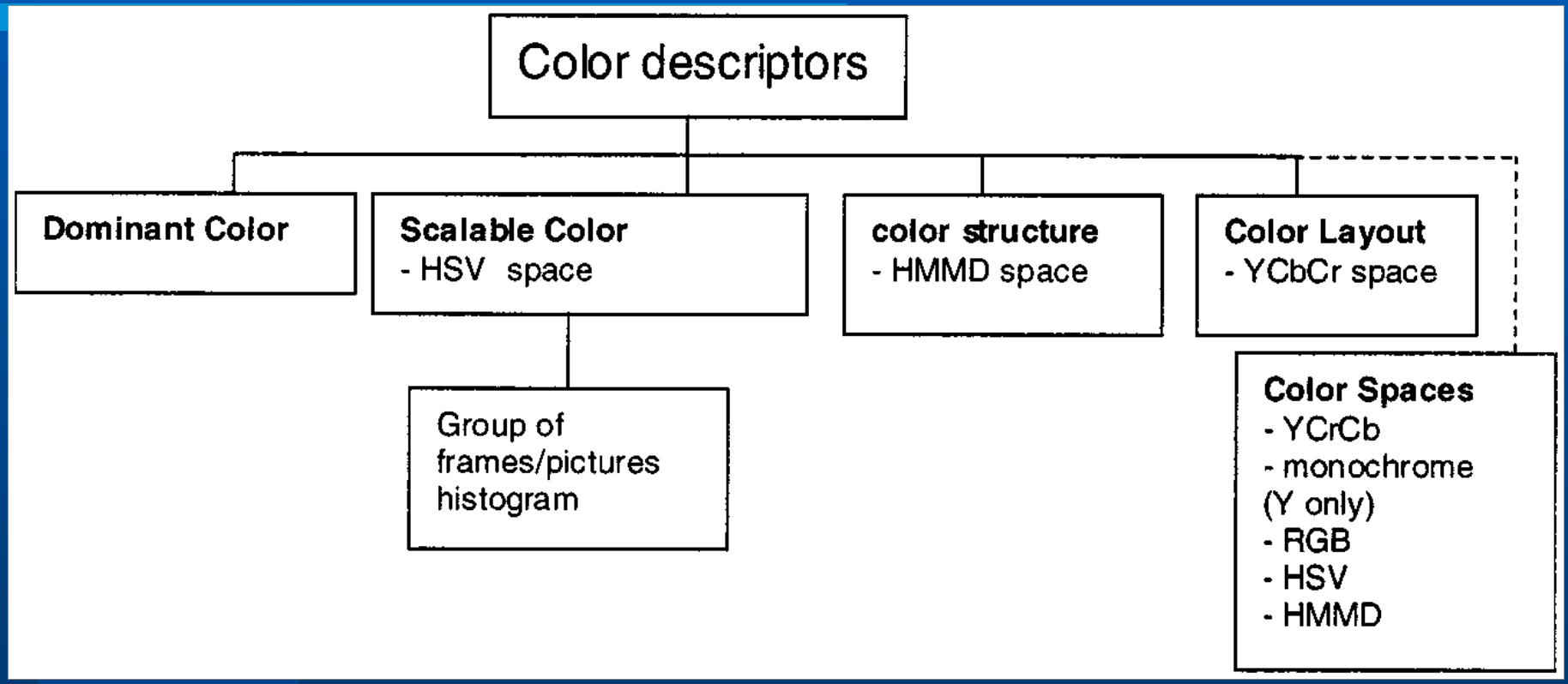
Complete description



Description example

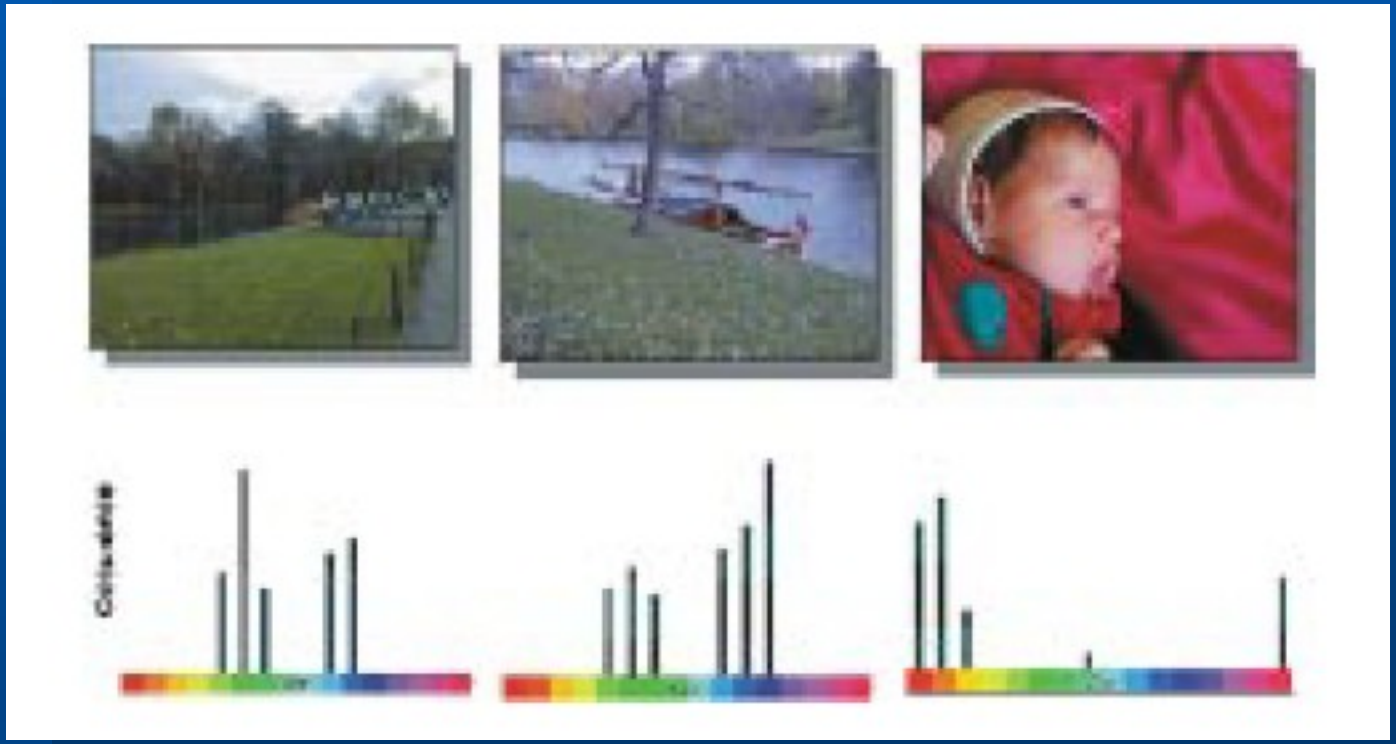


Color descriptors



Manjunah et al., **Color and texture descriptors**, IEEE Trans. on Circuits and Systems for Video Technology, June 2001

Color similarity

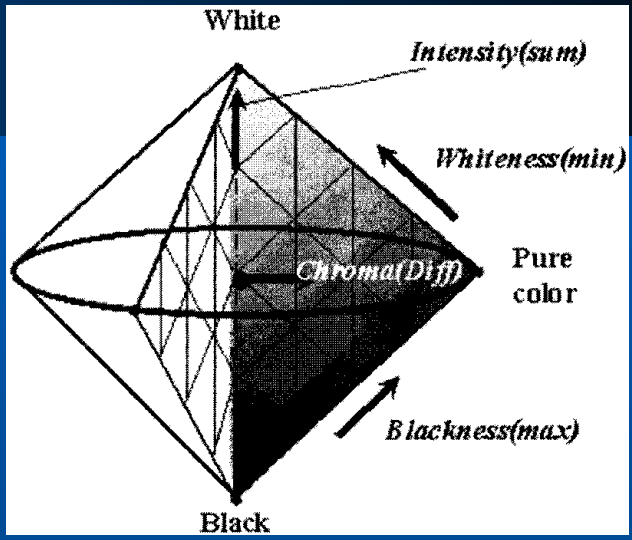


Color systems

RGB, YCbCr, Y (monochrome-intensity only)

HMMD

Diff=Max-Min
Sum=(max+min)/2
Hue as defined for the HSV.

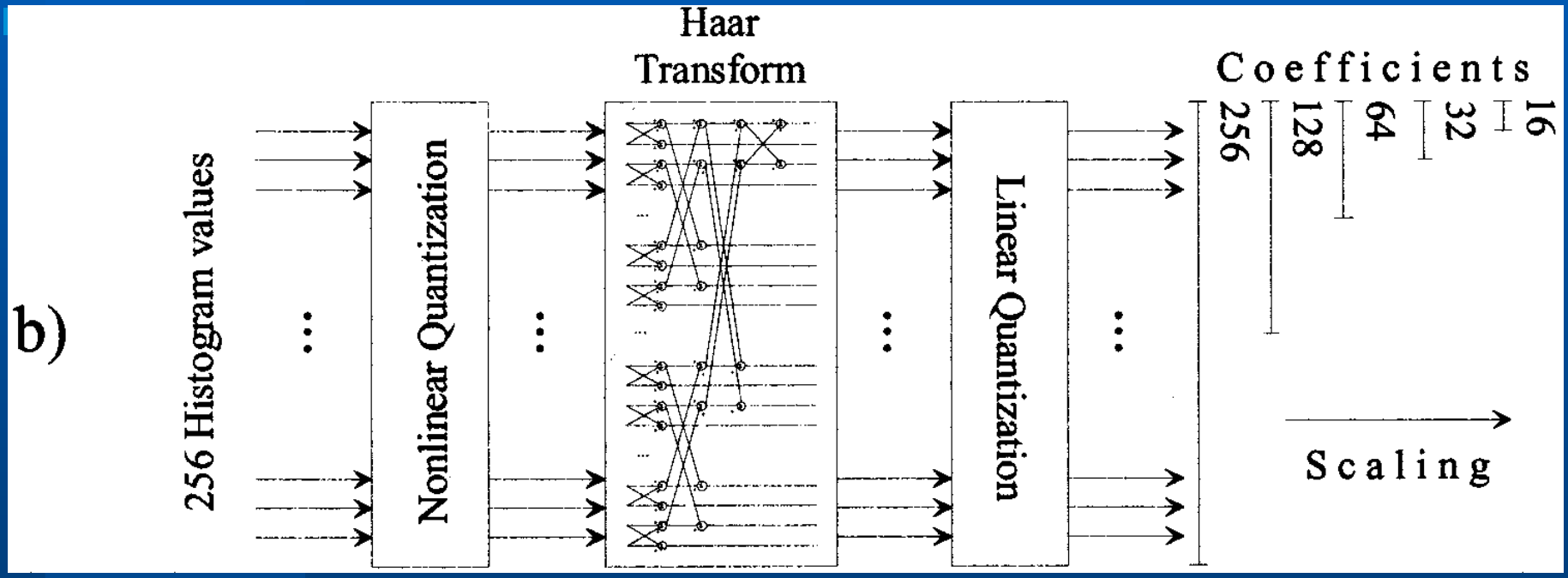


HSV

```
Max = max(R, G, B); Min = min( R, G, B);
Value = max(R, G, B);
if( Max == 0 ) then
  Saturation = 0; else
  Saturation = (Max-Min)/Max;
if( Max == Min ) Hue is undefined (achromatic color);
otherwise:
if( Max == R && G > B ) Hue = 60*(G-B)/(Max-Min)
else if( Max == R && G < B ) Hue = 360 + 60*(G-B)/(Max-Min)
else if( G == Max ) Hue = 60*(2.0 + (B-R)/(Max-Min))
else Hue = 60*(4.0 + (R-G)/(Max-Min))
```


Manjunah et al., **Color and texture descriptors**, IEEE Trans. on Circuits and Systems for Video Technology, June 2001

Color histogram



Manjunah et al., **Color and texture descriptors**, IEEE Trans. on Circuits and Systems for Video Technology, June 2001

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

Color histogram Haar transform

Extension to group of images

Dominant colors

$F = \{\{c_i, p_i, v_i\}, s\}, \quad i = 1, 2, ..., N$

k-means **algorithm**

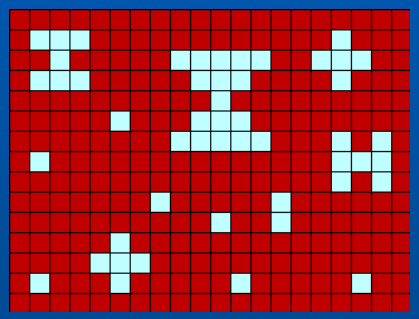
p_i **: 5 bits (percentage)**

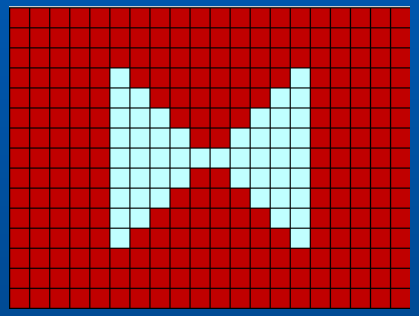
v_i **: 3 bits (variance)**

Spatial coherency (s) :

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

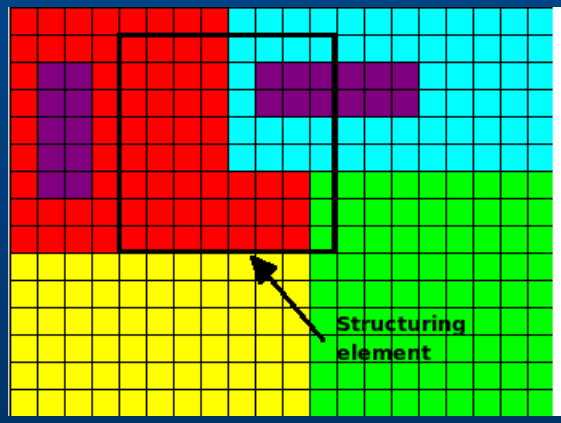
Color structure





Block 8 x 8

Local color histogram distribution HMMD



Color	Bin Value
C0	h(0)+1
C1	h(1)
C2	h(2)+1
C3	h(3)+1
C4	h(4)
C5	h(5)
C6	h(6)+1
C7	h(7)

Color layout

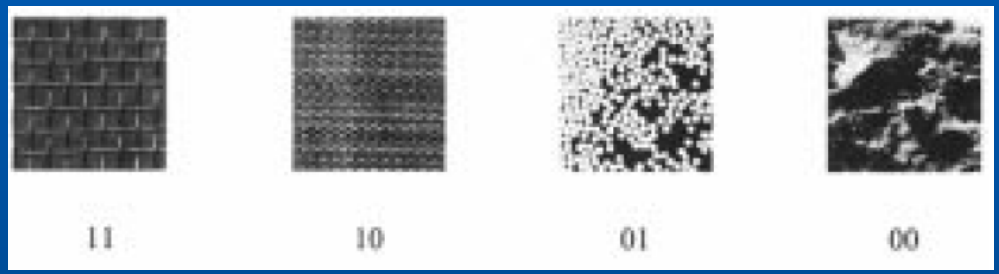
Color system YCbCr
Any shape region

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

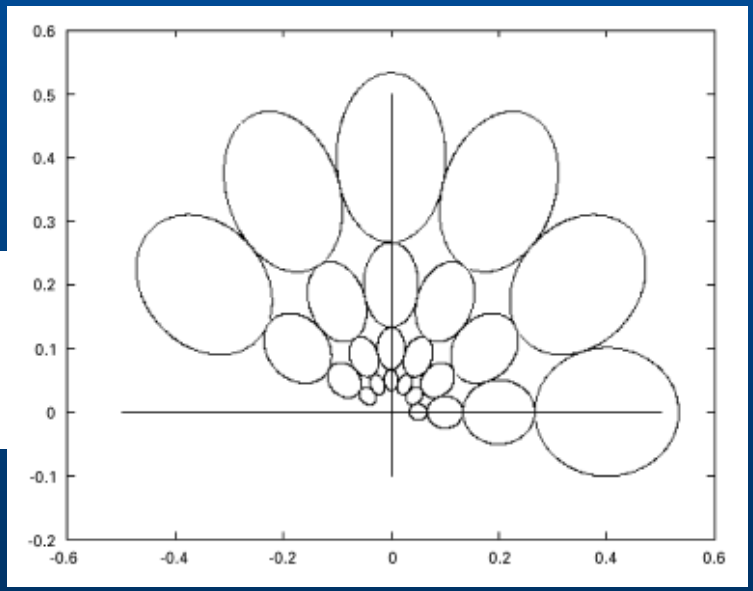
Regularity (2 bits)
Directionality (2x3 bits)
Coarseness (2x2 bits)



Homogeneous texture

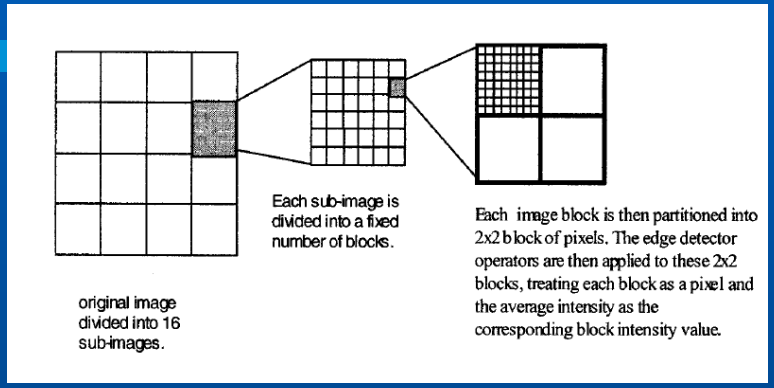
Bank of orientation and scale sensitive filters
Gabor filters : 5 scales and 6 directions

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



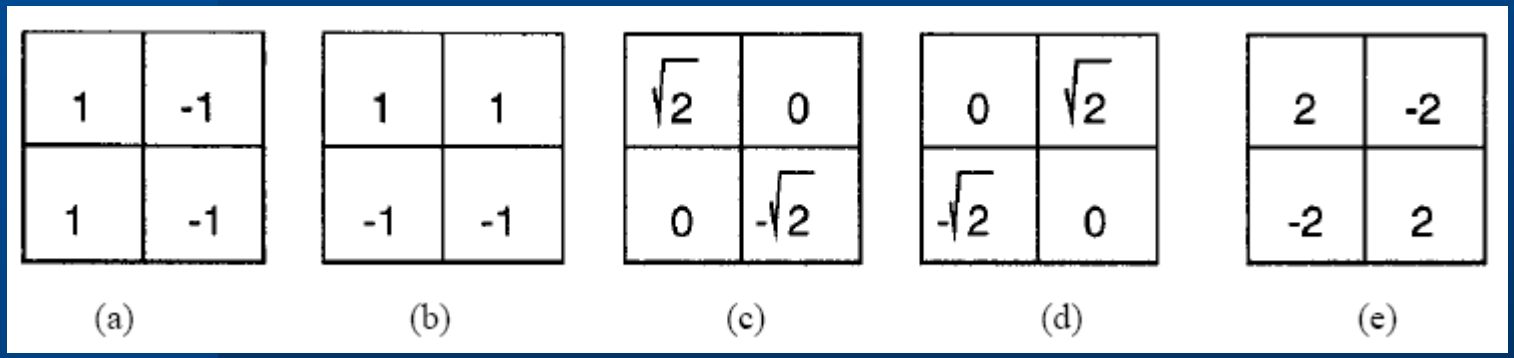
Manjunah et al., **Color and texture descriptors**,
IEEE Trans. on Circuits and Systems for Video Technology, June 2001

Texture (edge) description



Edge histogram description

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



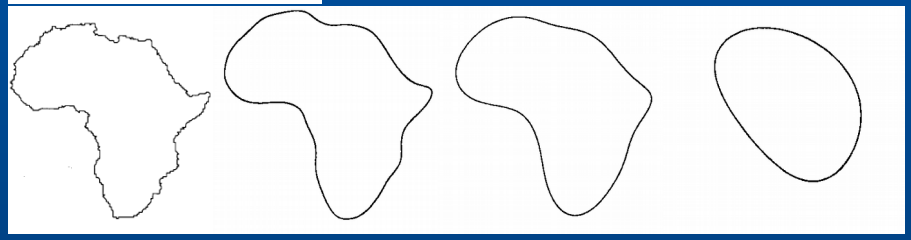
Shape descriptors

Bounding box

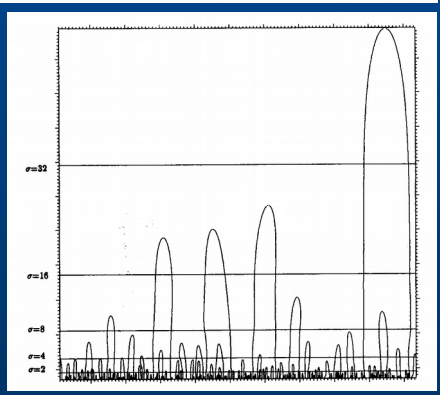
Region-based descriptor

- **Angular Radial Transformation**

Contour-based descriptor



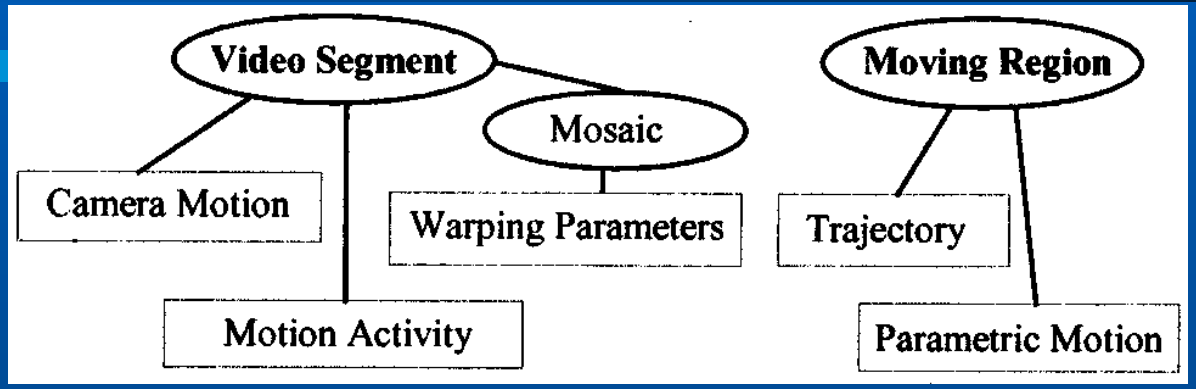
- **Curvature scale-space**
- **Curvature zero-crossings**
- **Prominent peaks of curvature**



3D surface descriptor

- **Shape index from principal curvatures**
- **Shape index histogram**
- **Planar regions**

Motion descriptors / activity

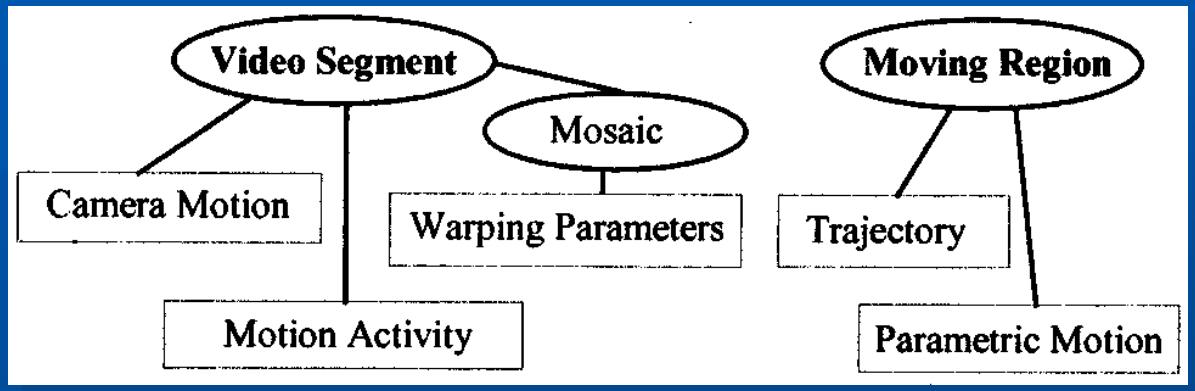


Motion activity

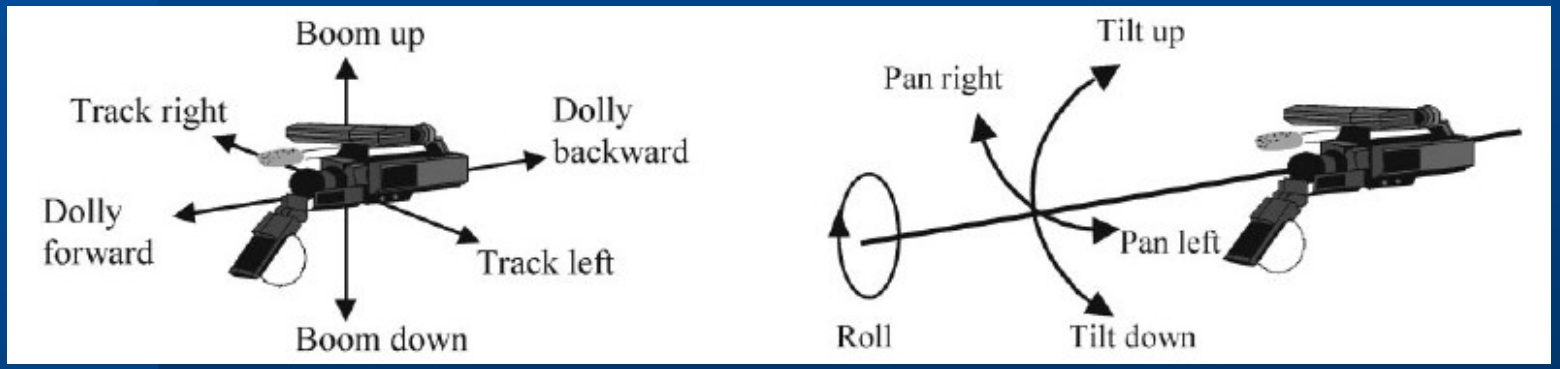
Activity Value	Range of σ (Std. Dev. of motion vector magnitude)
1	$0 \le \sigma < 3.9$
2	$3.9 \le \sigma < 10.7$
3	$10.7 \le \sigma < 17.1$
4	$17.1 \le \sigma < 32$
5	$32 \le \sigma$

Intensity of activity
Direction of activity
Spatial distribution
Temporal distribution

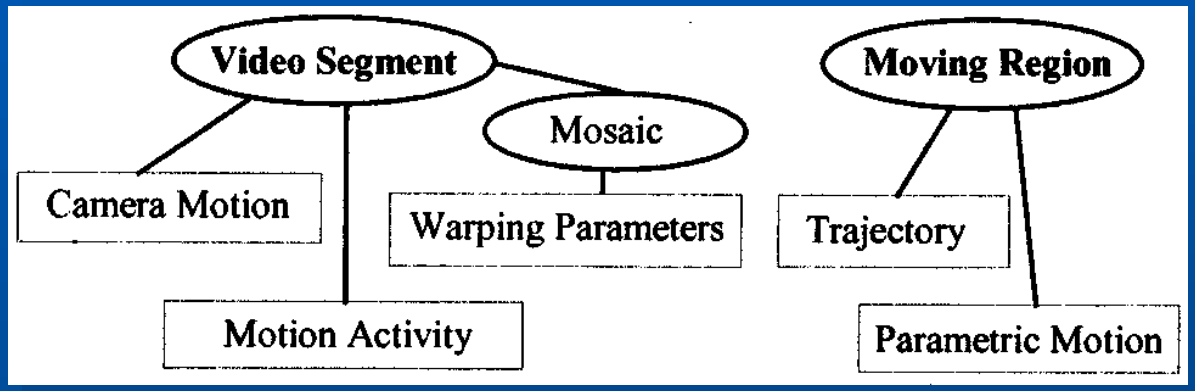
Motion descriptors



Camera motion



Motion descriptors / mosaic



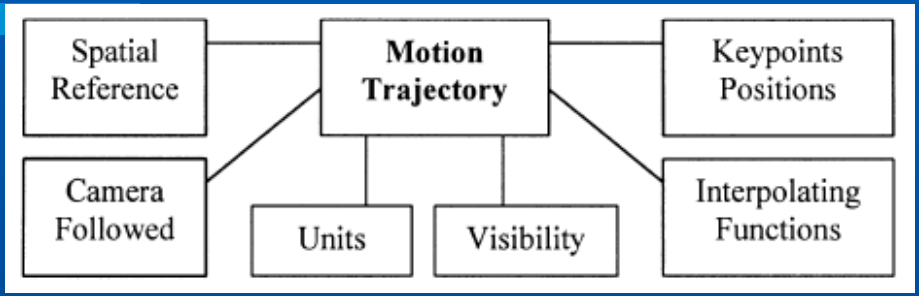
Warping parameters

$$x_i' = \frac{a_0 + a_1 x_i + a_2 y_i}{a_6 x_i + a_7 y_i + 1}$$

$$y_i' = \frac{a_3 + a_4 x_i + a_5 y_i}{a_6 x_i + a_7 y_i + 1}$$

Object motion descriptors

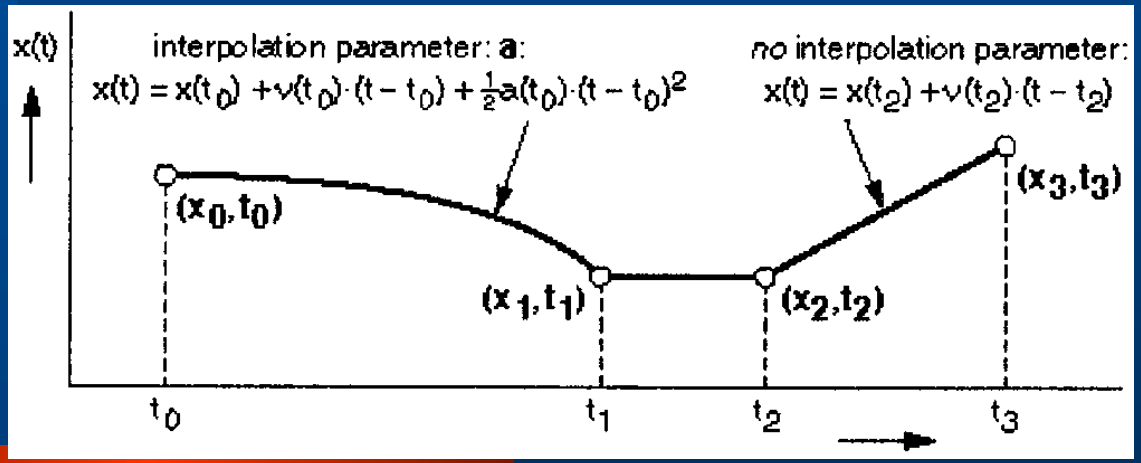
Motion trajectory



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*.



Summary

- 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 effcient 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.