ECE 417 Guest Lecture
Introduction to MPEG-7

Min-Hsuan Tsai

Apr 04, 2013

Background

- Multimedia Retrieval
  - Text-based retrieval → Content-based retrieval
  - Need:
    - Fast and accurate access
    - Personalized content production & consumption
    - Content management
    - Advanced query handling (visual, audio, sketch,...)
  - A solver:
    - Efficient processing tools to **create** description of AV material or to support the **identification** or **retrieval** of AV documents

What is MPEG-7

- Multimedia Content Description Interface
  - NOT focus so much on processing tools
    - NOT a standard for feature extraction
    - NOT a standard for matching/searching
    - NOT a standard for compression
  - Concentrate more on the **selection** of features to be described
  - Find a way to **structure** and **instantiate** the selected features with a common language
 - Efficient representation of audio-visual (AV) meta-data
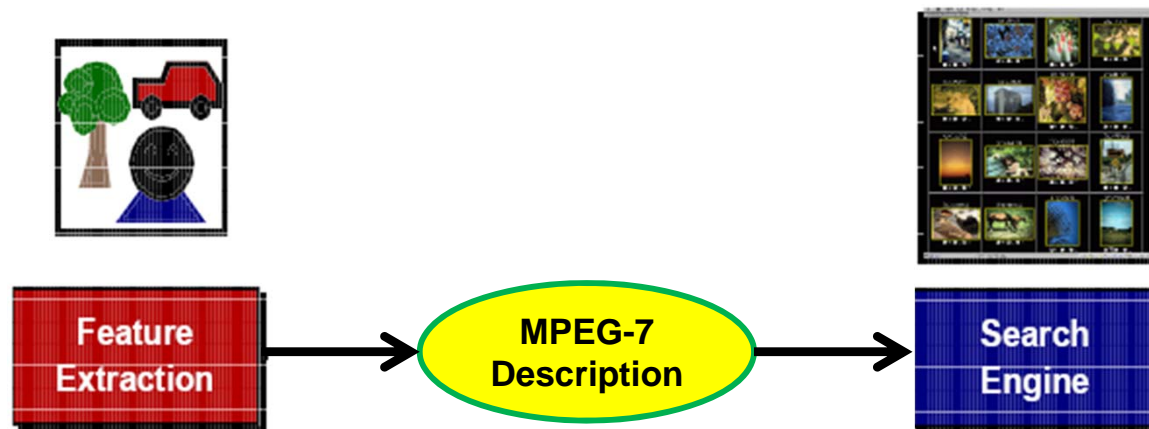
MPEG-7 Main Elements

- Descriptors (D)
  - syntax & semantics of feature representation
- Description Schemes (DS)
  - semantics & structure of the relationships between components (D's & DS's)
- Description Definition Language (DDL)
  - Standardized language to express D's and DS's
  - Allow creation, modification and extension of existing D's and DS's
- Systems tools
  - issues of synchronisation, transmission mechanisms...

MPEG-7 Main Elements
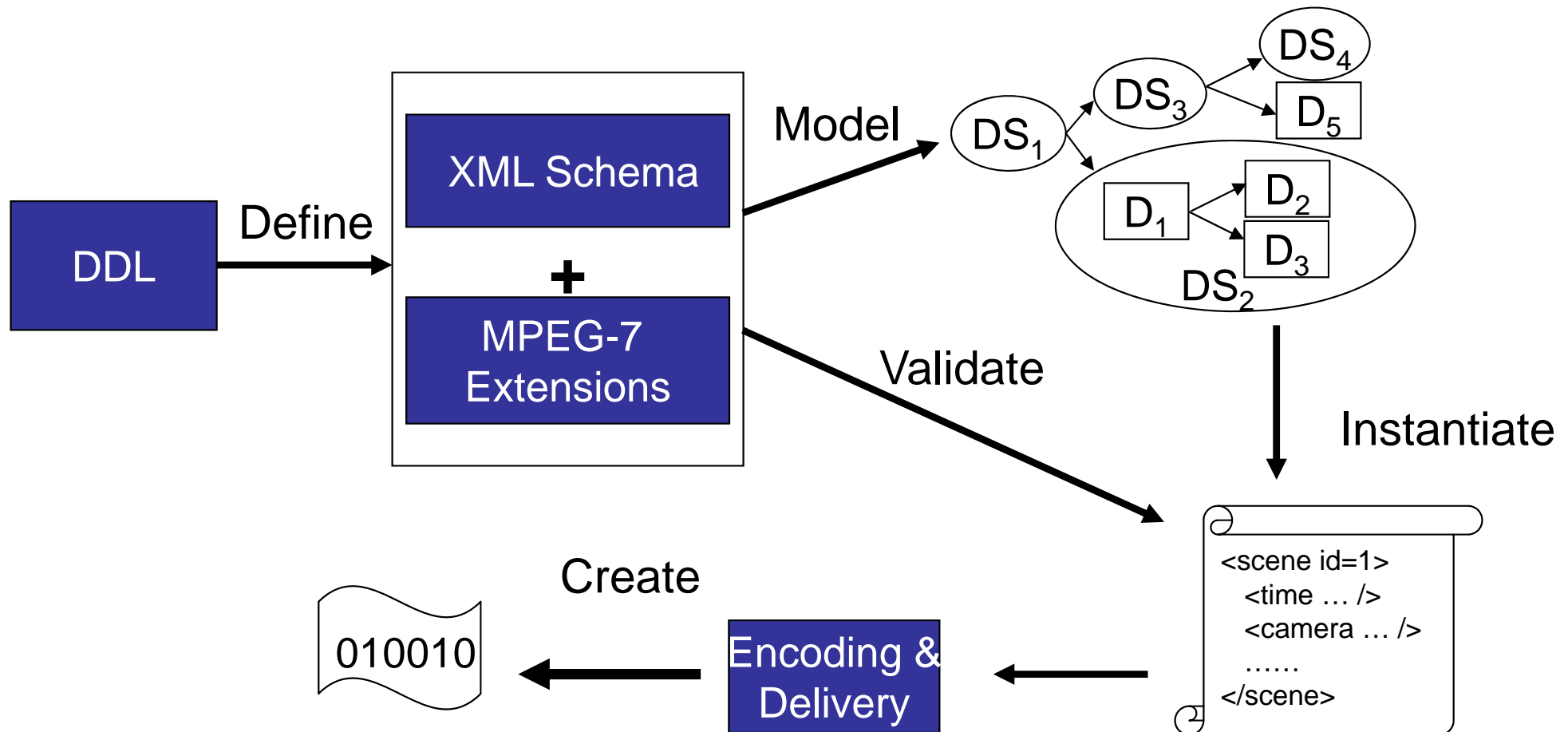
MPEG-7 DDL
(Description Definition Language)

- Foundations of MPEG-7 standard, provides the language for defining the structure and content of multimedia information
    - XML schema + MPEG-7 Extensions
    - With structural and value constraints
    - With reusability: rules provided by which user can combine, extend and refine existing D's and DS's

MPEG-7 DDL

- A simplified example

```
<VideoSegment id = "VS1" >
        <MediaTime>
                <MediaTimePoint> T0:0:0    </MediaTimePoint>
                <MediaDuration>    PT10M </MediaDuration>
        </MediaTime>

        <StructuredAnnotation>
                This is an example of video segment
        </StructuredAnnotation>

        <GofGopColorHistogram HistogramTypeInfo = "Average">
                ....
        </GofGopColorHistogram>

        <SegmentDecomposition Gap = "true" Overlap = "true" DecompositionType = "temporal">
                <VideoSegment id = "VS2" > .... </VideoSegment>
                <VideoSegment id = "VS3" > .... </VideoSegment>
                <VideoSegment id = "VS4" > .... </VideoSegment>
        </SegmentDecomposition>
</VideoSegment>
```
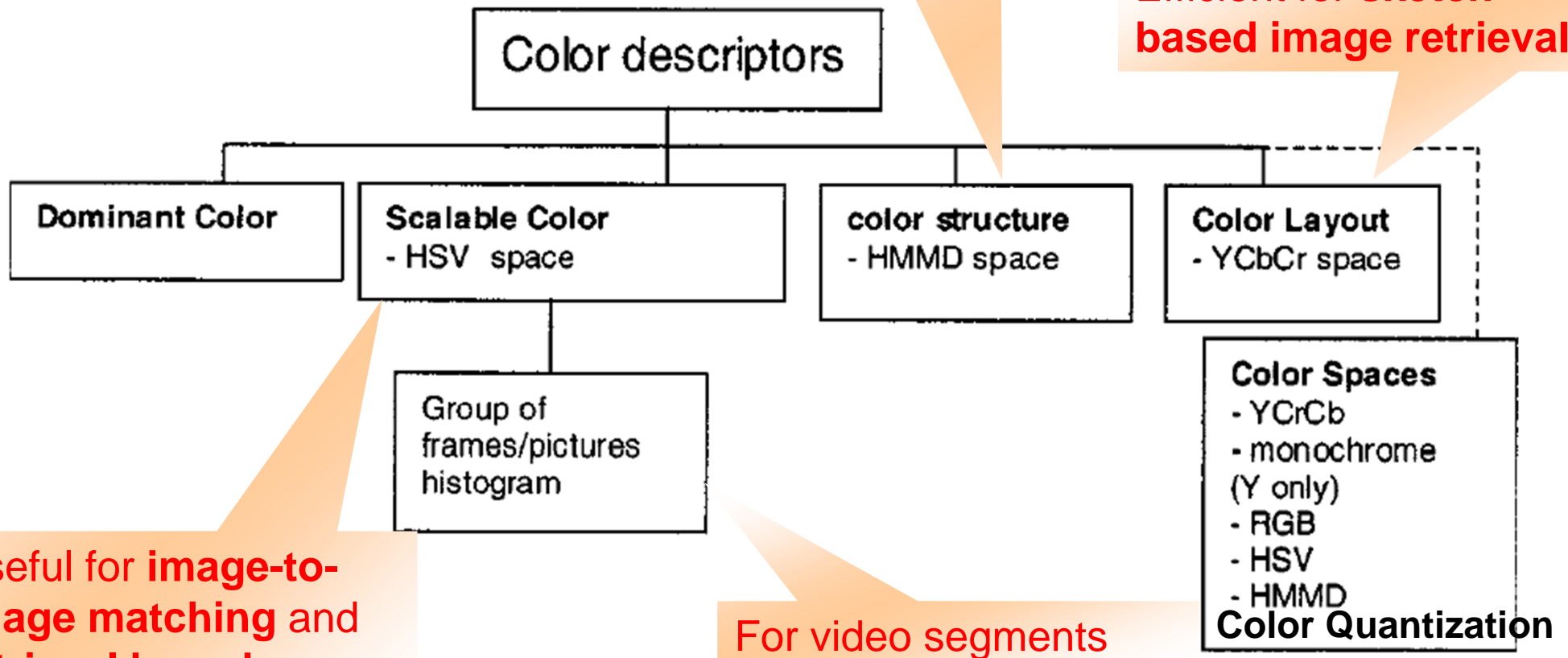
MPEG-7 Visual Descriptors

- Cover 6 basic visual features as
  - Color
  - Texture
  - Shape
  - Motion
  - Localization
  - Face Recognition

Color Descriptors

- Includes 7 D's

Useful for **natural image retrieval**

Efficient for **sketch-based image retrieval**

Color descriptors

| Dominant Color | Scalable Color - HSV space | color structure - HMMD space | Color Layout - YCbCr space |

Group of frames/pictures histogram

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

Color Quantization

Useful for **image-to-image matching** and **retrieval based on color feature**

For video segments

(Images courtesy of Vajihollahi and Farahboo

Color Descriptors (2)

- Color space descriptor
  - Allows a selection of a color space to be used (Monochrome, RGB, YCrCb, HSV, HMMD,…)

- Color quantization descriptor
  - Specifies the partitioning of the given color space into discrete bins

Color Descriptors (3)

- Dominant color descriptor (DCD)
  - Clustering colors into a small number of representative colors
 - Can be defined for each object, regions or the whole image
  - DCD is defined as

$$F = \{(c_i, p_i, v_i), s\}, \quad i=1,\ldots,N$$

- c_i: representative color, a vector in the color space
- p_i: the fraction of pixel in the image with c_i color
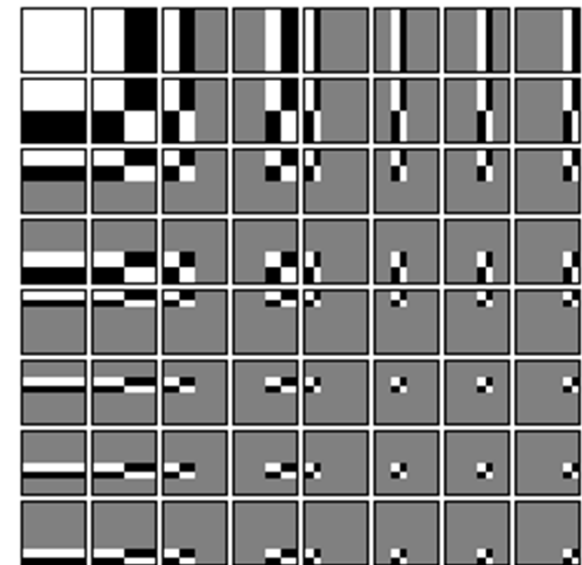- v_i: color variance of the pixels in the cluster
- s: the spatial coherency
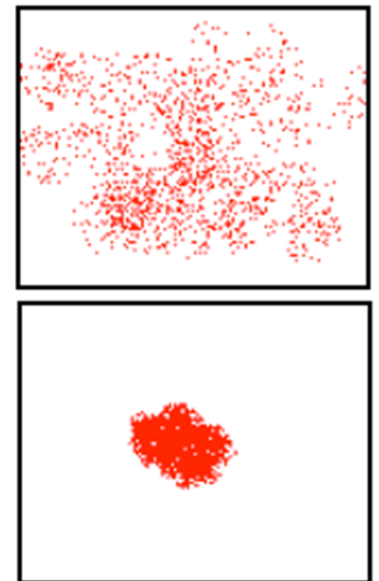
Color Descriptors (4)

- Scalable color descriptor (SCD)
  - A color histogram in HSV color space
  - Encoded by Haar Transform
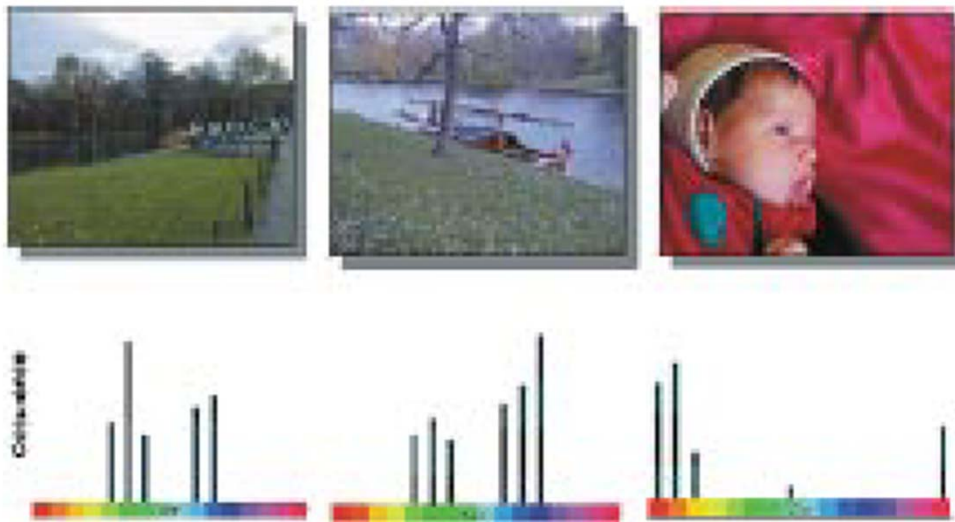
2D Haar basis (8x8)

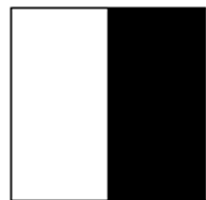 - Useful for image-to-image matching retrieval based on color feature

(Images courtesy of Sikora)

Texture Descriptors

- Includes 3 D's
  - Homogeneous texture descriptor (HTD)
  - Texture browsing descriptor (TBD)
  - Edge histogram descriptor (EHD)

Texture Descriptors

- Edge histogram descriptor
  - Represents the spatial distribution of five types of edges

a) vertical edge   b) horizontal edge   c) 45 degree edge   d) 135 degree edge   e) non-directional edge

  - Dividing the image into 16 (4x4) blocks
 - Generating a 5-bin histogram for each block
  - Scale invariant

Shape Descriptors

- Includes 3 D's
  - Region-based shape descriptor
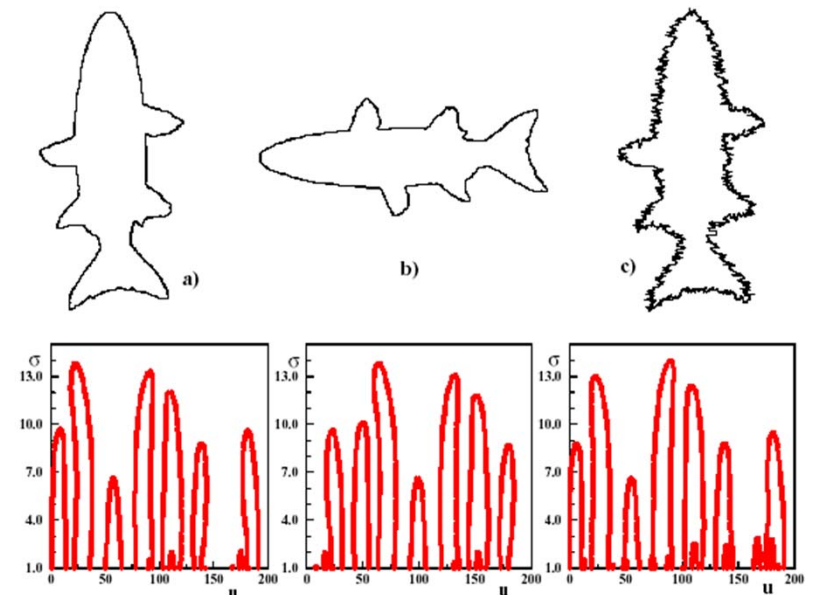 - Contour-based shape descriptor
 - 3-D spectrum descriptor

Shape Descriptors

- ## Region-based shape 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
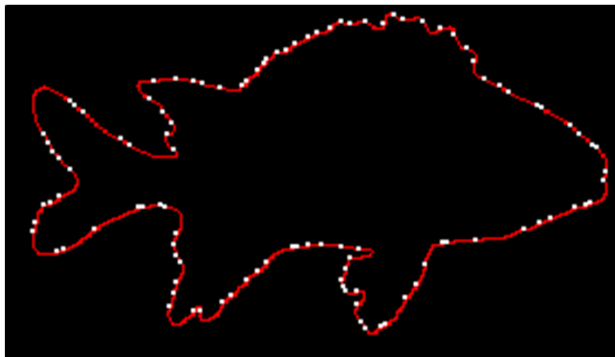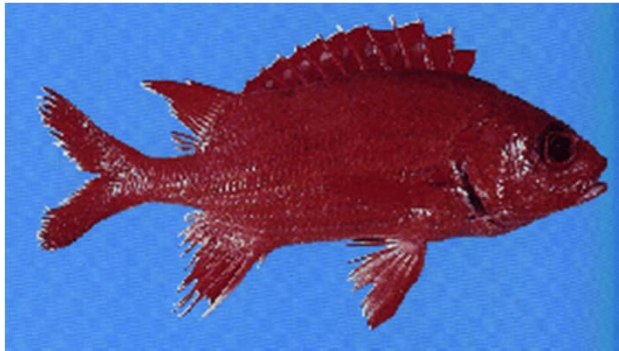

(a)   (b)   (c)   (d)   (e)
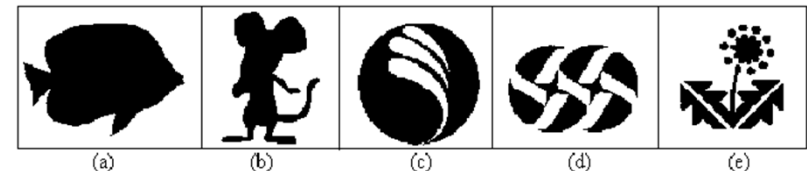
Shape Descriptors

- Contour-Based Descriptor
  - based on Curvature Scale-Space representation
    - 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
    - Position of key points are expressed relative to the length of the contour curve
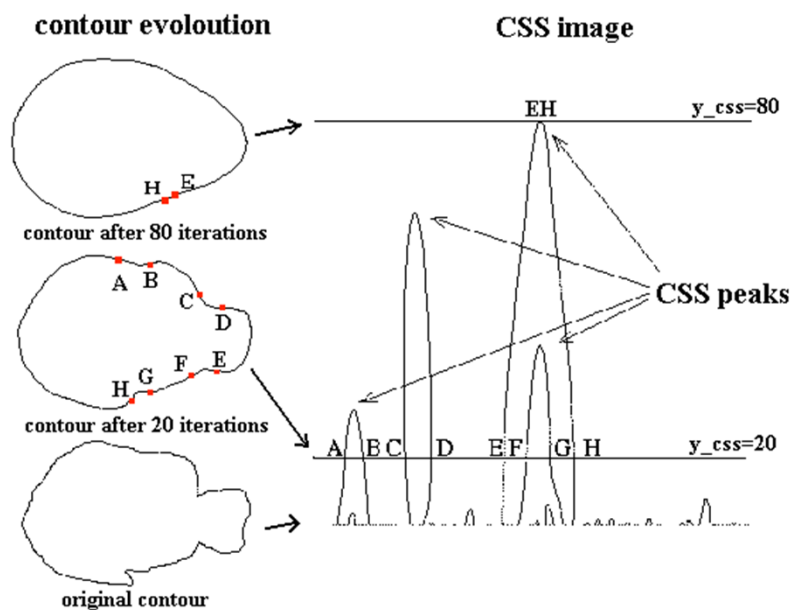
Curvature Scale-Space
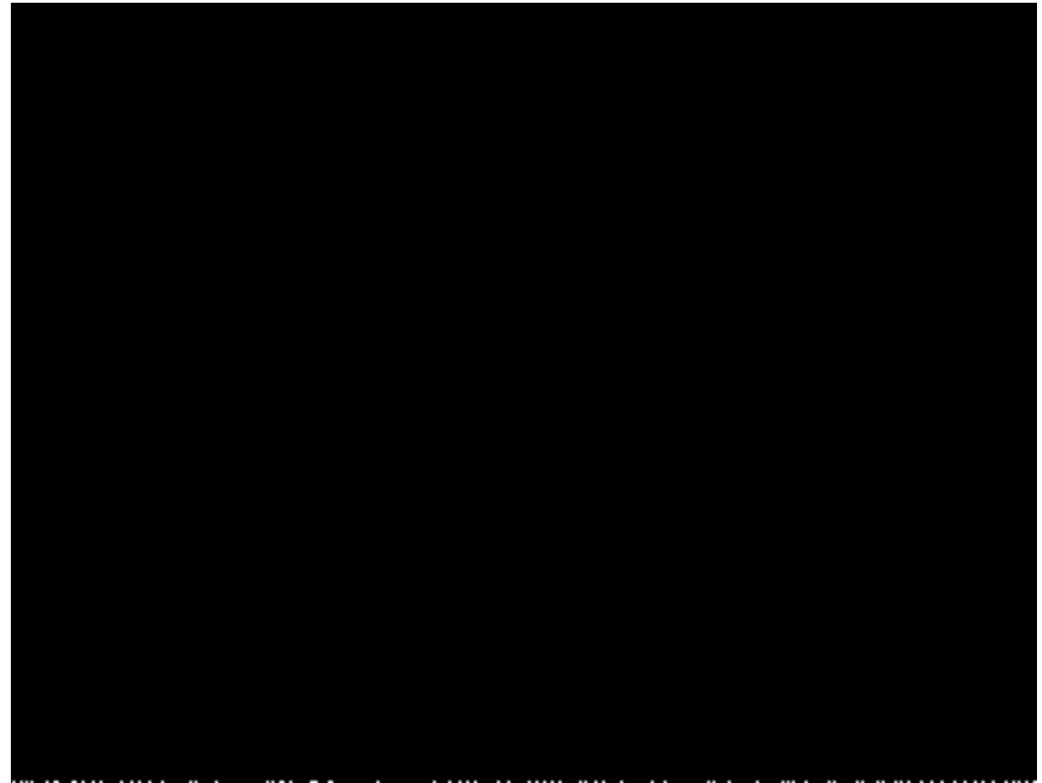
Shape Descriptors

- Contour-Based Descriptor
  - Robust to the noise, scale and orientation
  - Fast and compact
  - Examples:
    - Applicable to (a)
    - Distinguishes differences in (b)
    - Support semantically similar for human (c)
    - Robust to non-rigid deformations (d)
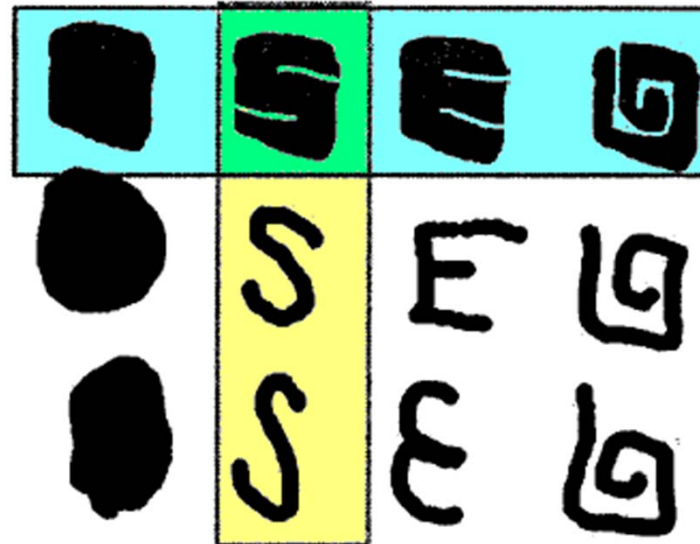   - Robust to distortion due to perspective transformation (e)
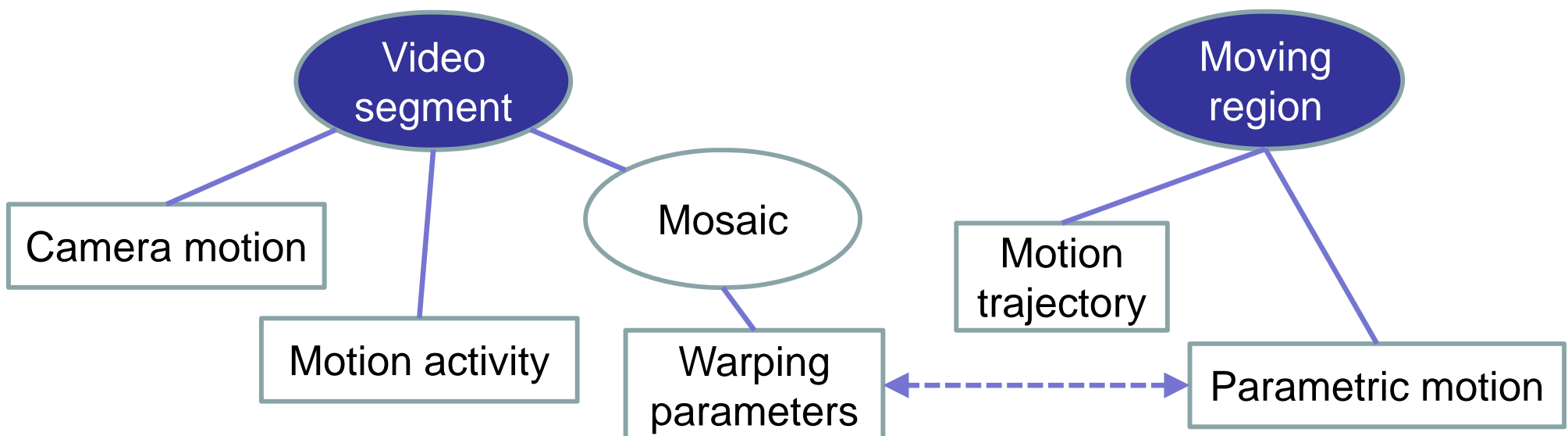
Shape Descriptors

- Comparison between region-based and contour-based

  – Blue: Similar shapes by Region-Based
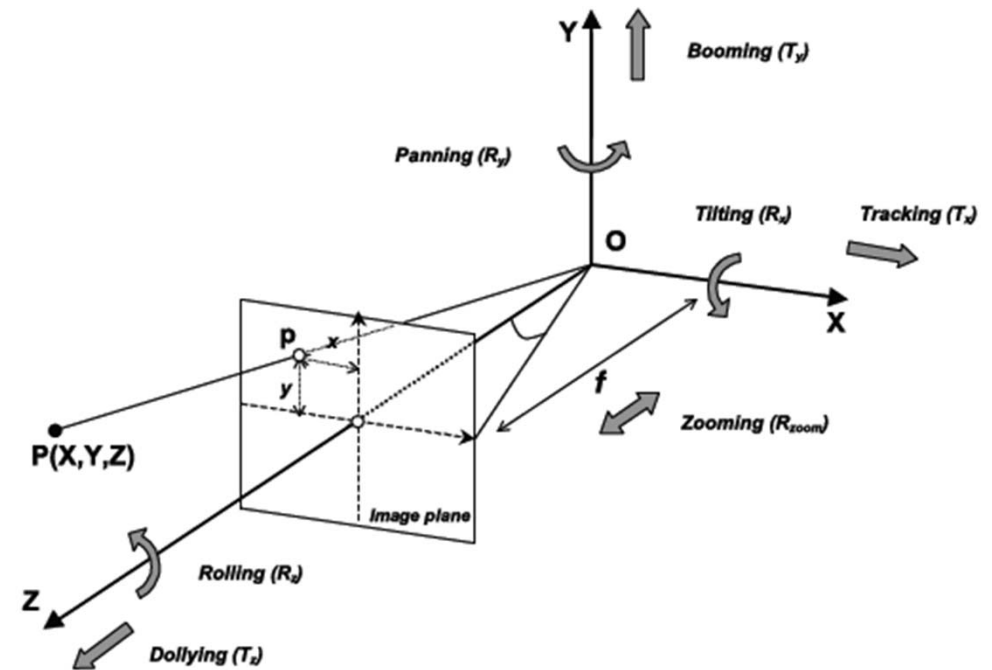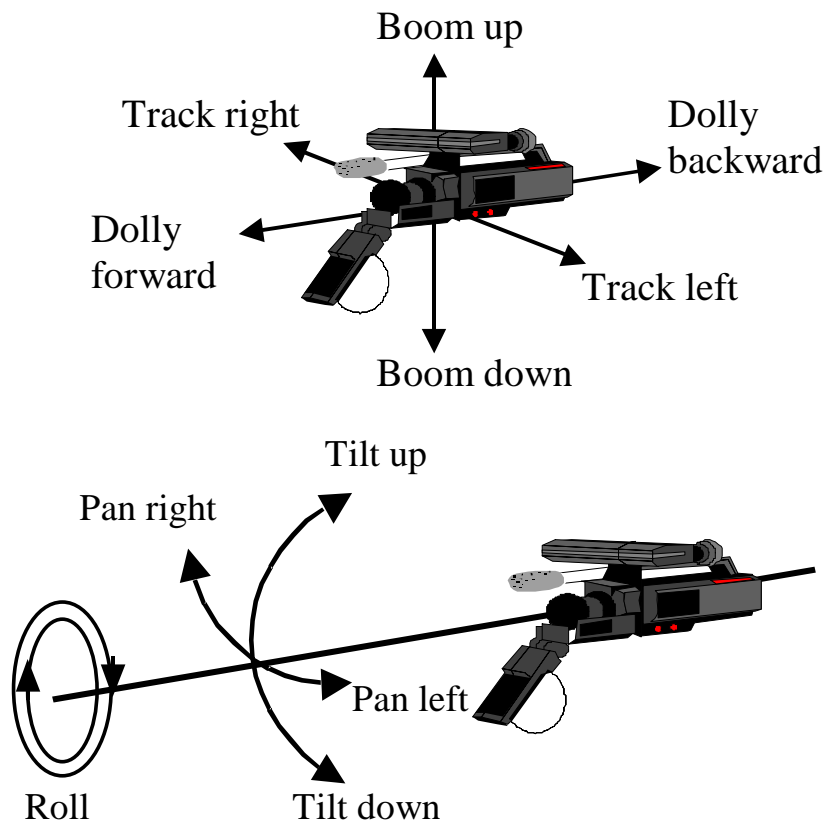 – Yellow: Similar shapes by Contour-Based

Motion Descriptors

- Includes 4 D's
  - Camera motion descriptor
  - Motion activity descriptor
  - Motion trajectory descriptor
  - Parametric motion descriptor

Motion Descriptors

- Camera motion descriptor
  - Describe the movement of a camera or a virtual viewpoint
  - Supports 7 camera operations

Localization Descriptors

- Includes 2 D's
  - Region locator
    - Specifies regions with a box or a polygon

  - Spatio-temporal locator
    - Describes spatial-temporal regions in a video sequence, such as moving object regions, and provides localization functionality

Face Recognition Descriptor

- Used to retrieve face images which match a query face image
  - Feature extraction
    - From a normalized face image with sized 56-by-46 pixels
    - Aligned such that the center of the two eyes are located on (24, 16)-th and (24, 31)-th pixel
  - Face recognition descriptor
    - The projection of a face vector onto a set of 48 basis eigenvectors ('eigenfaces') which span the space of possible face vectors

(a) mean face   (b) 1st vector   (c) 11th vector   (d) 21th vector   (e) 31th vector   (f) 41th vector