9th International Workshop on Content-Based Multimedia Indexing (CBMI-2011, Madrid)

Region-based Semantic Image Analysis: Application to Image and Video Indexing

J.R. Casas, F. Calderero, A. Gasull, X. Giró, M. León, F. Marqués, M. Pardàs, J. Pont, P. Salembier, D. Varas, C. Ventura, V. Vilaplana

Image and Video Processing Group
Signal Theory and Communications Department
http://gps-tsc.upc.es/imatge/
ETSETB — UPC

IMATGE

UPC Departament de Teoria del Senyal i Comunicacions
UNIVERSITAT POLITÈCNICA DE CATALUNYA

Current structure of the Group

Staff

Professors

- Josep R. Casas
- Antoni Gasull
- Xavier Giró
- Ferran Marqués
- Ramon Morros
- Albert Oliveras
- Montse Pardàs
- Javier Ruiz
- Philippe Salembier
- Elisa Sayrol
- Verónica Vilaplana

Research Assistants

- Albert Gil

PhD Students

- Marcel Alcoverro
- Felipe Calderero
- Jaime Gallego
- Miriam León
- Adolfo López
- Gerard Palau
- Jordi Pont
- Julio Rolón
- Jordi Salvador
- Xavier Suau
- Silvia Valero (UPC/INPG)
- David Varas
- Carles Ventura

Introduction (I)

Signal descriptors

- Color histogram
- Gabor filter bank
- Angular transform coefficients
- Affine model of the trajectory
- ...

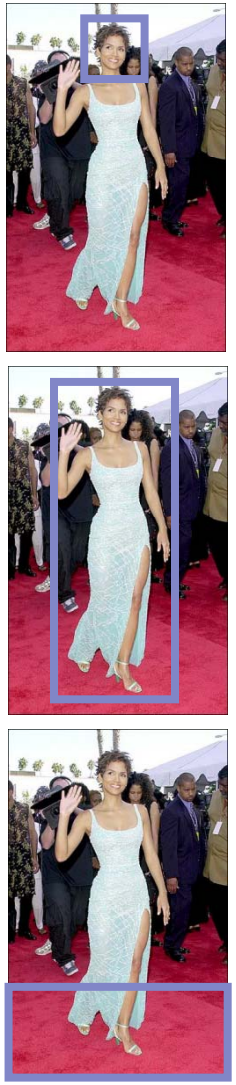
Semantic entities

- Face
- Person
- Groups
- Red carpet
- Smile
- Halle Berry
- Waving
- Oscar ceremony

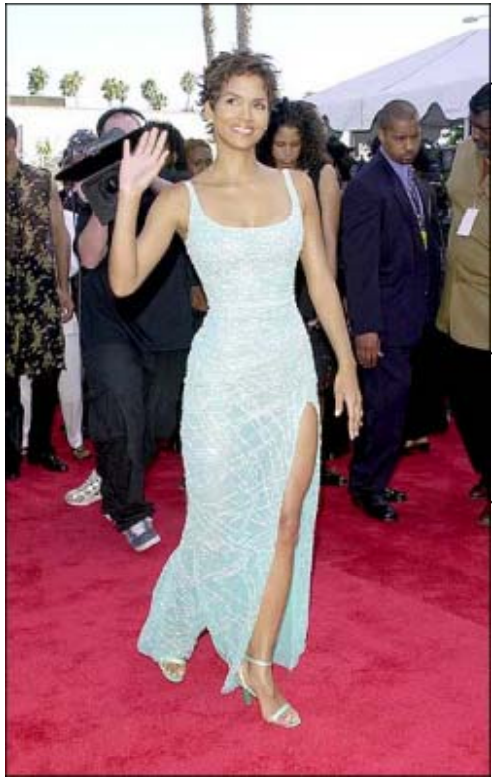
Introduction (II)

Semantic entities:

- ✓ Human face
- ✓ Human body
- ☐ Group of people
- ☐ Sky
- ☐ Tree
- ☐ Building
- ☐ Automobile
- ☐ Camera
- ✓ Carpet
- ☐ …

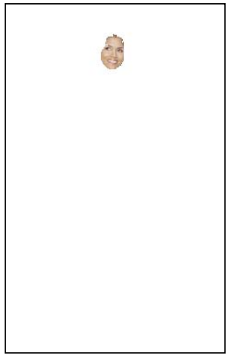


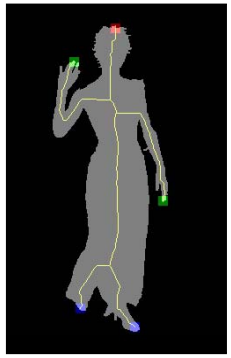
Introduction (III)



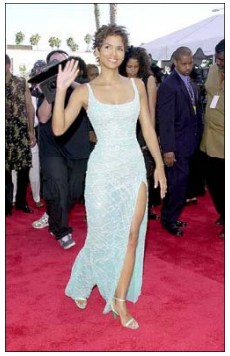
Semantic entities:

- ✓ Human face
- ✓ Human body
- □ Group of people
- □ Sky
- □ Tree
- □ Building
- □ Automobile
- □ Camera
- ✓ Carpet
- □ …





Introduction (IV)



Semantic entities:

- ☐ Human face
- ✓ Human body
- ☐ Group of people
- ☐ Sky
- ☐ Tree
- ☐ Building
- ☐ Car
- ☐ …

Segmentation and Feature extraction

High level **region-based** analysis

Human body

Low level descriptors: MPEG-7

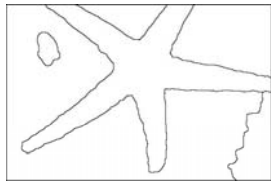
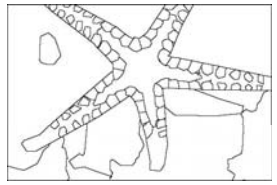
- **Image Partition**
- Color Histogram — ScalableColor
- Gabor filters — HomogeneousTexture
- Angular transform — RegionShape

Introduction (V)

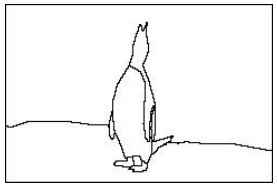
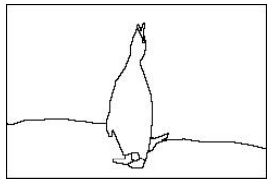
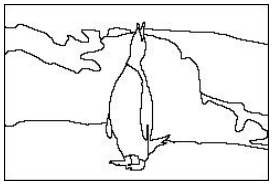
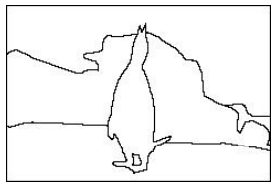
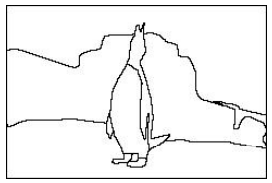
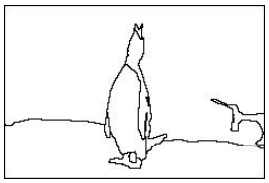
However: Segmentation is **an ill posed problem**:

- A unique solution does not exist and the partition definition depends on the final **application**.

Different levels of detail

Same level of detail

Introduction (V)

However: Segmentation is an ill posed problem:

- A unique solution does not exist and the partition definition depends on the final **application**.
- A given partition defines a **universe** of possible **regions**.
- To propose a **single algorithm** for generic object detection is **not feasible**.
- Nevertheless, a **region-based analysis** should help on detecting generic objects:
  - Better estimation of features, reduction of elements, …

We are going to apply these **region-based representation** concepts to the context of the i3media project.

Overall Presentation

- Introduction
  - Global problem introduction
  - Specific problem statement
- Contextualized scenarios
  - Specific approach
  - Generic approach
- Region-based image representation
  - Binary Partition Tree
  - Merging criteria
  - Assessing criteria
- Region-based object representation
  - Object definition
  - Perceptual object model
  - Examples
- Other topics
  - Structural object model
  - Graphic User Interface and Training
  - Query by Example
  - Extension to video analysis
  - Soccer analysis applications
- References and Conclusions

Problem statement (I)

Video (**massive** or **very large**) databases.

- Searching through them using:
  - Manually annotated semantic tags.
- Willing to:
  - Richer automatic/semi-automatic semantic annotations.
  - Query by Example search approaches.

Some data about **CCMA** content:

- Main channel (TV3) started 28 years ago; currently 7 channels.
- Historical: Storage system of 3PetaBytes (around 480.000 hours)
- Current: Broadcasted content: 41% News and 6,7% Sports
- Current: Every hour, 15 hours of new content are ingested (average).

Problem statement (II)

Access to the semantic content of the image requires **powerful image and video analysis** tools:

- Initially, based on **shot detection / key frame extraction**
  - **Already integrated algorithms** working since more than 10 years
- Annotate semantic objects during ingest time and **off-line**
  - **Different types of algorithms** may be used
    - Complexity versus Accuracy.
  - Algorithms with **learning capabilities**
    - Increase the number of semantic objects: Training algorithms
- Annotate semantic objects during **ingest time** and off-line
  - **Simpler** metadata
  - Use of **context**

Problem statement (III)

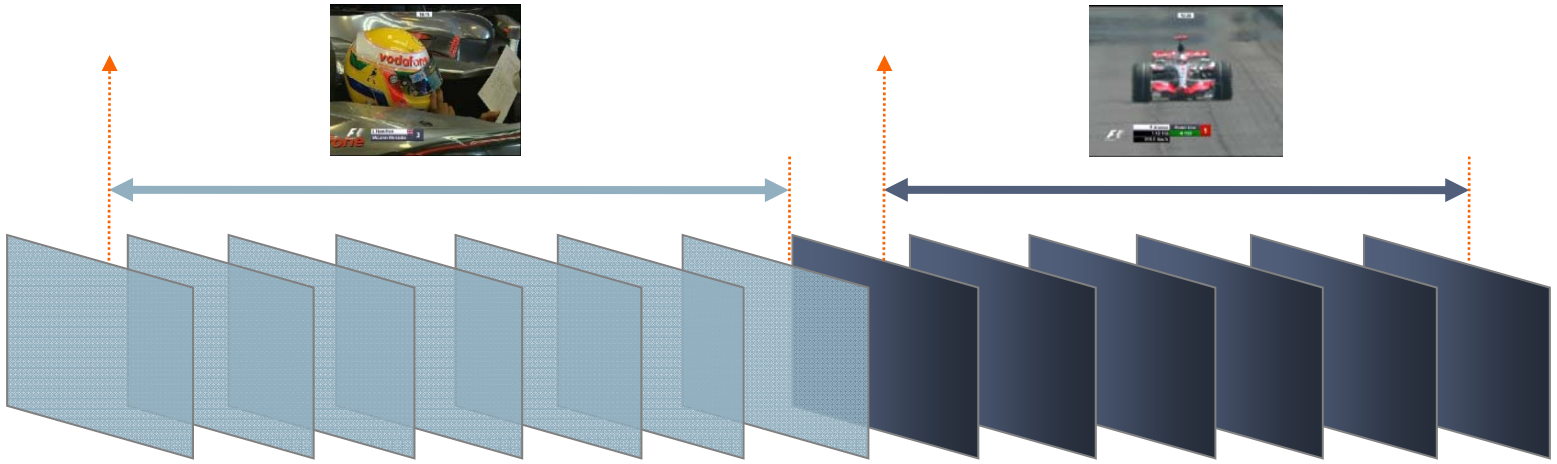
Semantic entities:

- Human Face
- Specific Person
- Text
- Specific Words
- Logos
- Specific Logo
- Specific Location
- …

- A new person

Videos similar to:

- An example image
- An example region
- A set of regions
- …



Problem statement (IV)

It is possible to define **three types of problems**:

- Detection of semantic objects in **strongly contextualized** scenarios:
  - Objects are **very specific** and no other objects are to be found
  - **News Agency content**: Table of content with a known format

Problem statement (V)

It is possible to define **three types of problems**:

- Detection of generic or specific semantic objects in **contextualized** scenarios:
  - The context **helps adapting** the object recognition algorithms.
  - **Sports**, **News** or specific channels (**Canal Parlament**)

Problem statement (VI)

It is possible to define **three types of problems**:

- Detection of semantic objects in generic scenarios: **no context** is available
  - **Generic algorithms** are necessary.
  - Preparing the data for **future searches**: new objects

Overall Presentation

- Introduction
  - Global problem introduction
  - Specific problem statement
- Contextualized scenarios
  - Specific approach
  - Generic approach
- Region-based image representation
  - Binary Partition Tree
  - Merging criteria
  - Assessing criteria
- Region-based object representation
  - Object definition
  - Perceptual object model
  - Examples
- Other topics
  - Structural object model
  - Graphic User Interface and Training
  - Query by Example
  - Extension to video analysis
  - Soccer analysis applications
- References and Conclusions

Contextualized scenarios (I)

A common problem is **shot or camera identification**:

- **Specific techniques** for a given scenario allow better performance
  - No unique definition of types of shots / cameras.

Beauty-shot

Wide-shot

Multiple general shot

Single general shot

Medium shot

Close Up

Detail shot

Soccer bench shot

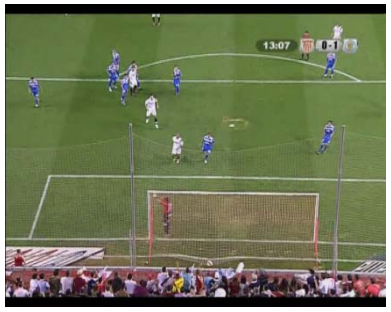
Ambient shot

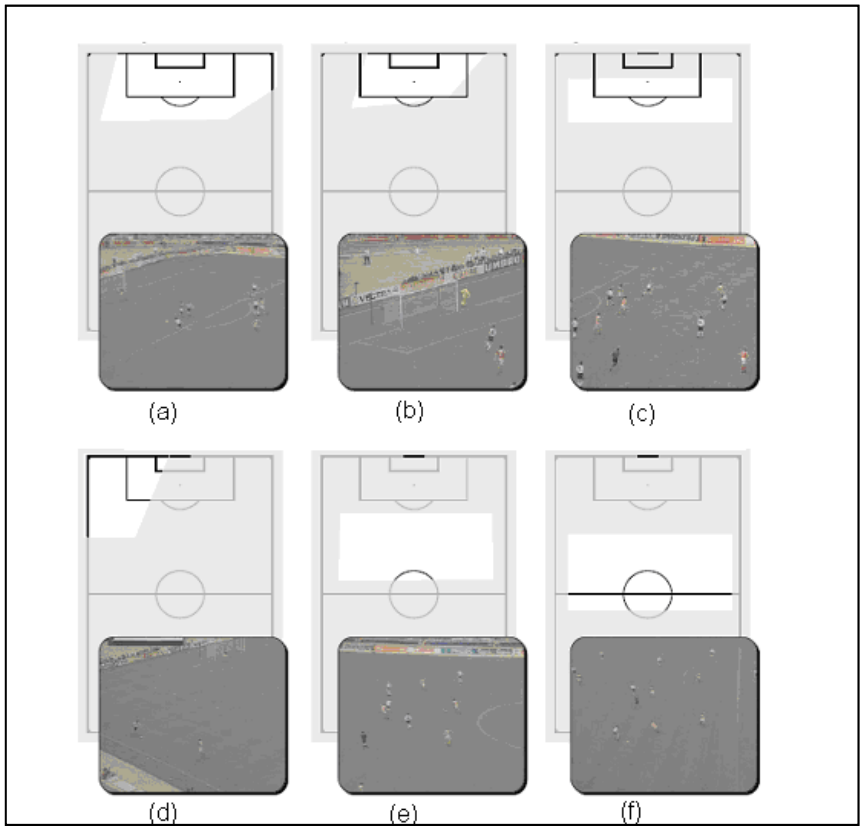
Aerial shot

Contextualized scenarios (II)

A common problem is **shot or camera identification**:

- **Specific techniques** for a given scenario allow better performance
  - No unique definition of types of shots / cameras: **Wide shot**





Contextualized scenarios (III)

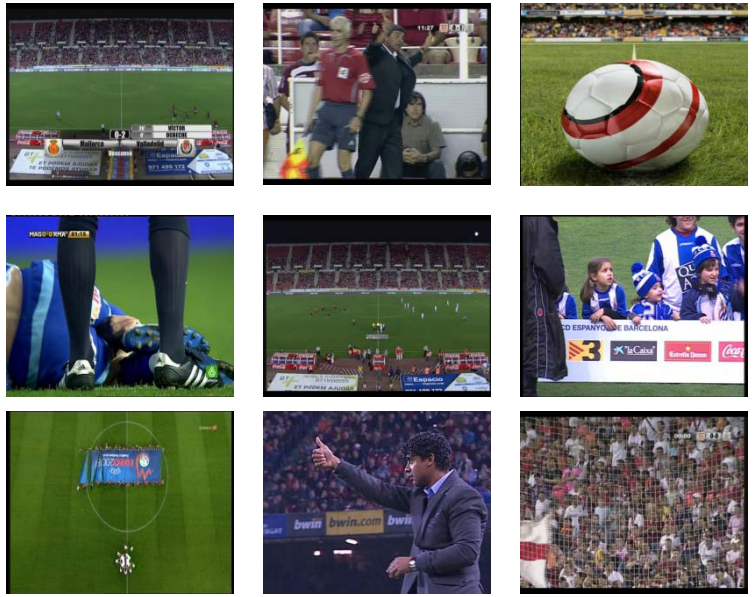
A common problem is **shot or camera identification**:

- **Specific techniques** for a given scenario allow better performance
  - Adopt a definition and select the **actually relevant** shots

Wide-shot

General shot

Medium shot

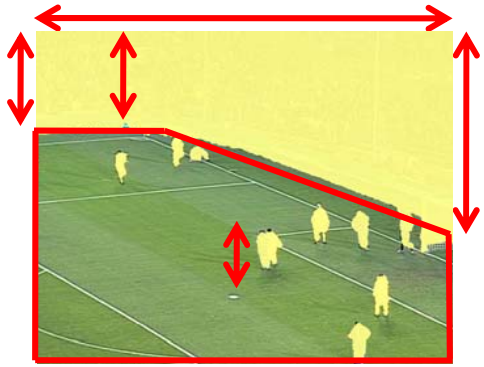


Other shots

Contextualized scenarios (IV)

A common problem is **shot or camera identification**:

- **Specific techniques** for a given scenario allow better performance
  - Adopt a definition and select the **actually relevant** shots
  - Define a **specific set of descriptors for that precise problem**



Match	Number of images	Accuracy
Malaga-Real Madrid	12056	96.7%
F.C.Barcelona-Malaga	30517	94.0%
Sevilla-Getafe	5111	92.9%
F.C.Barcelona-Recreativo	70000	92.0%
Sevilla-Real Madrid	32153	96.0%

Contextualized scenarios (V)

A common problem is **shot or camera identification**:

- **Specific techniques** for a given scenario allow better performance

Contextualized scenarios (VI)

A common problem is **shot or camera identification**:

- A generic **supervised classification approach** is also possible:
  - Define a **generic set of descriptors**
    - Use of **MPEG-7 descriptors**:
      - Dominant Color Descriptor
      - Color Structure Descriptor
      - Color Layout Descriptor
      - Texture Edge Histogram Descriptor
      - Contour Shape Descriptor
    - Analyze **similarity measures** and **combination strategies**
  - Design a **friendly Graphic User Interface**
    - **Extremely important** for the success of the application

Contextualized scenarios (VII)

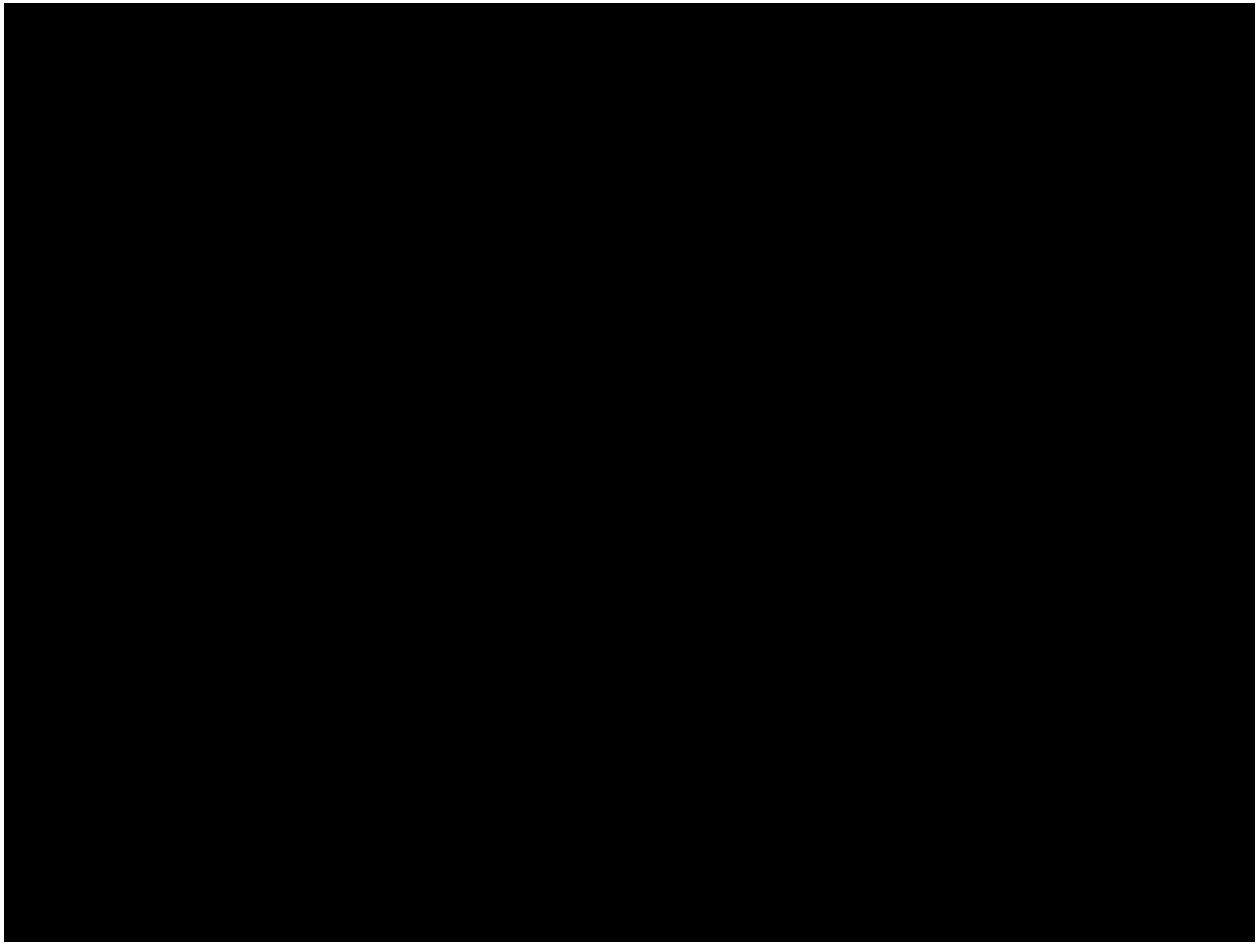
A common problem is **shot or camera identification**:

- A generic **supervised classification approach** is also possible:

Contextualized scenarios (VIII)

A common problem is **shot or camera identification**:

- A generic **supervised classification approach** is also possible:



Overall Presentation

- Introduction
  - Global problem introduction
  - Specific problem statement
- Contextualized scenarios
  - Specific approach
  - Generic approach
- Region-based image representation
  - Binary Partition Tree
  - Merging criteria
  - Assessing criteria
- Region-based object representation
  - Object definition
  - Perceptual object model
  - Examples
- Other topics
  - Structural object model
  - Graphic User Interface and Training
  - Query by Example
  - Extension to video analysis
  - Soccer analysis applications
- References and Conclusions

Image Representation (I)

Region-based representation:

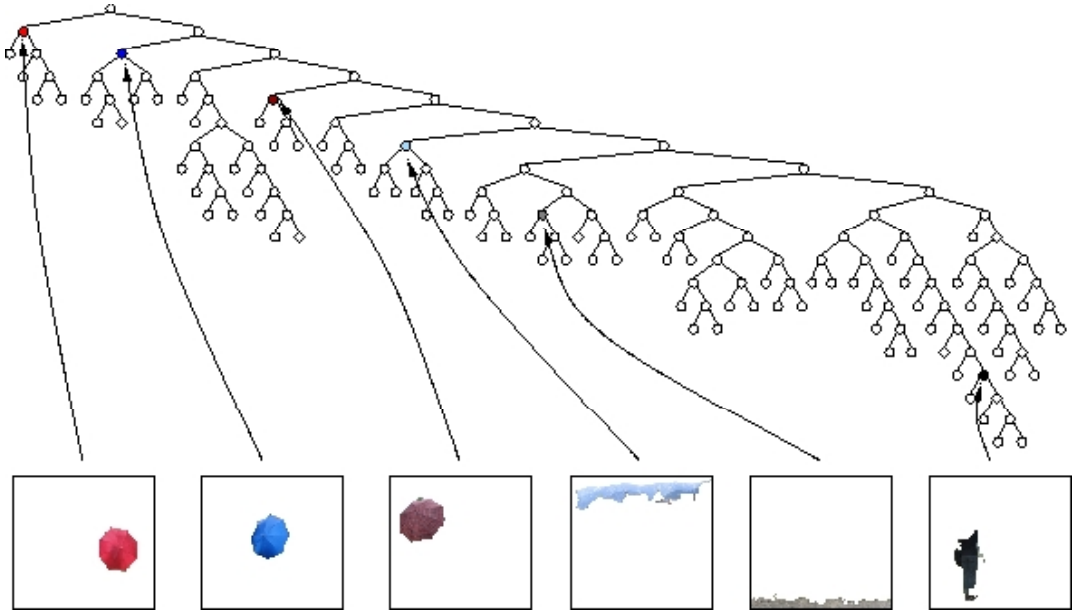
The image is divided into regions that **conform** (or are good anchor points) of **objects** in the scene:

- Less elements to analyze
- More robust estimations

Hierarchical representation:

A given partition and a set of possible regions formed by the **most likely mergings** of these initial regions.

- Further reduce the amount of elements

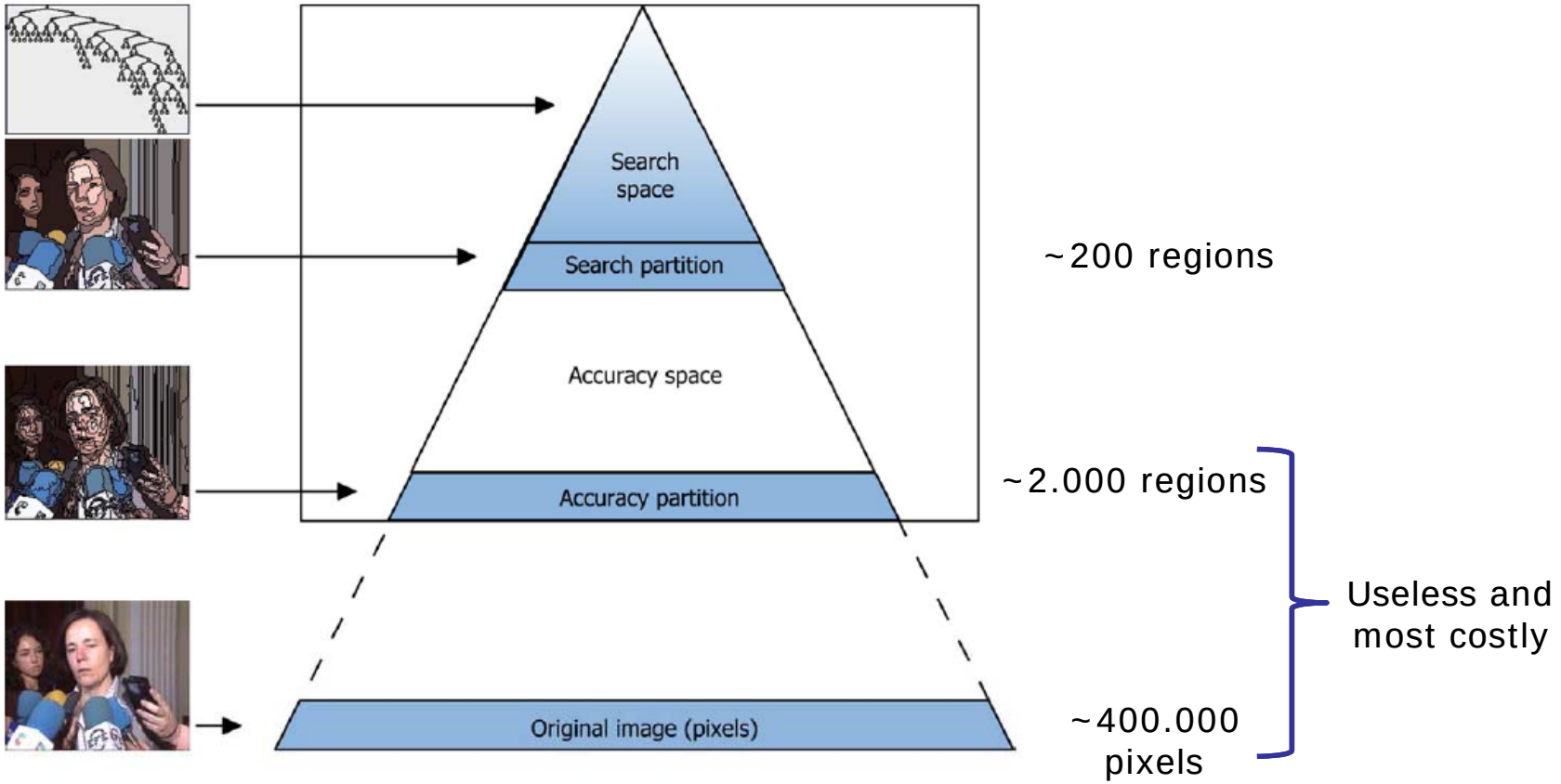


Region-based description:

A **set of descriptors** associated to each node in the tree.

- e.g.: MPEG-7 descriptors

Image Representation (II)



The best **merging criterion or combination of criteria** has to be found for the generic object representation application

Merging criteria (I)

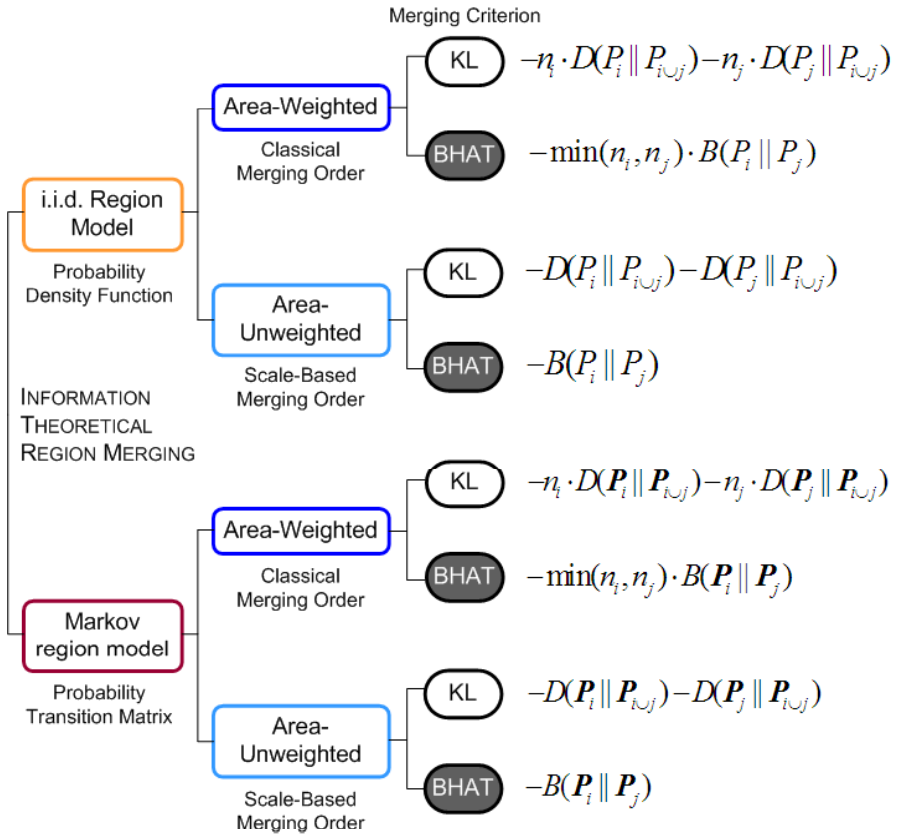
There exist two main families of region models with their associated merging criteria: **Non-statistical** and Statistical models.

- Modeling the region with its **mean value** in a 3D color space:
  - Typically, YUV or RGB color spaces
  - Merging criteria: comparison between means of regions to be merged
    - Weighting in a different way each channel depending on its dynamic range
    - Including size information in the merging cost
- Incorporating **contour complexity** to the model:
  - Prioritizing regions with smooth contours, no holes, …

Merging criteria (II)

There exist two main families of region models with their associated merging criteria: Non-statistical and **Statistical** models.

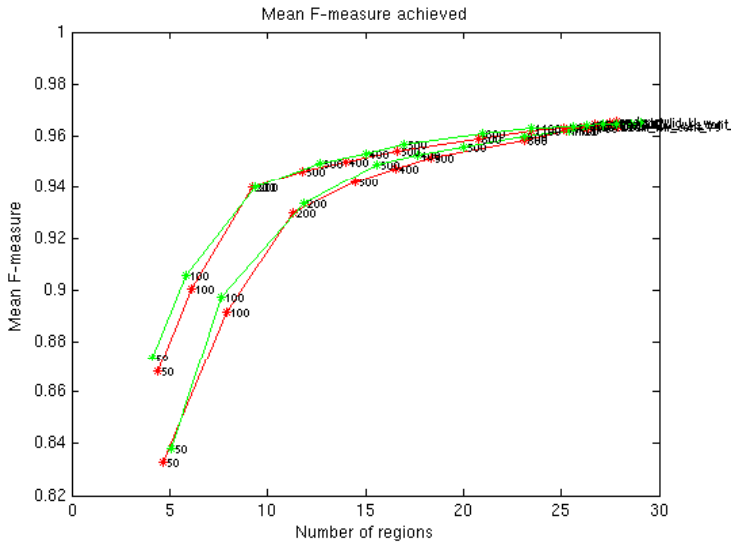
- Modeling the region with its **empirical distribution** in a 3D color space:
  - Typically, i.i.d or Markovian models
  - **Information Theory** based merging criteria
    - Kullback-Leibler or Bhattacharyya criteria
    - Include a scale factor



Merging criteria (III)

The **optimal sequence of criteria** is a combination of the previous ones.

- The complexity of the optimal model increases as we reach higher levels of the hierarchy.
  - Initial **flat zone labeling** is enough
    - Speeds up the process
  - **Search partition**: Color mean
    - Channel weighting
  - **Upper part**: Information Theory
    - Non-homogeneous quantization
    - Incorporating **contour complexity** to the model



Assessing criteria (I)

How to assess the quality of a partition?

- There exists a large number of **different measures** in the literature:

  - Precision / Recall: F-Measure
  - Jacard Index

  - Segmentation Covering
  - Global Consistency Error
  - Bipartite Graph Matching
  - Variation of Information
  - Hoover Measures

  - Precision / Recall for regions
  - Rand Index

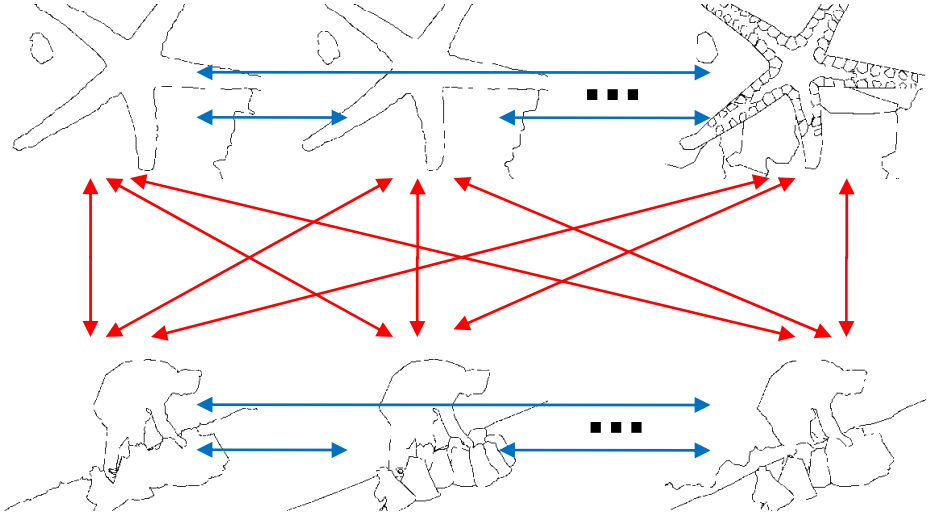
Detailed analysis reveals that they can be interpreted and structured into a **few classes:**

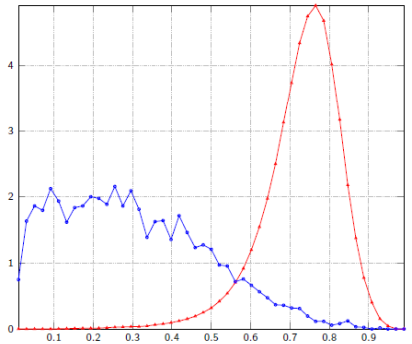
1. Single object interpretation
2. Pixel set partition
3. Pairs of pixels classification

Assessing criteria (II)

How to assess the quality of a partition?

- There is an agreement on the way to assess a given quality measure:
  - **Berkeley data base** of human segmented images
  - **Same versus Different** image discrimination:
    - Takes into account **mutual refinement**

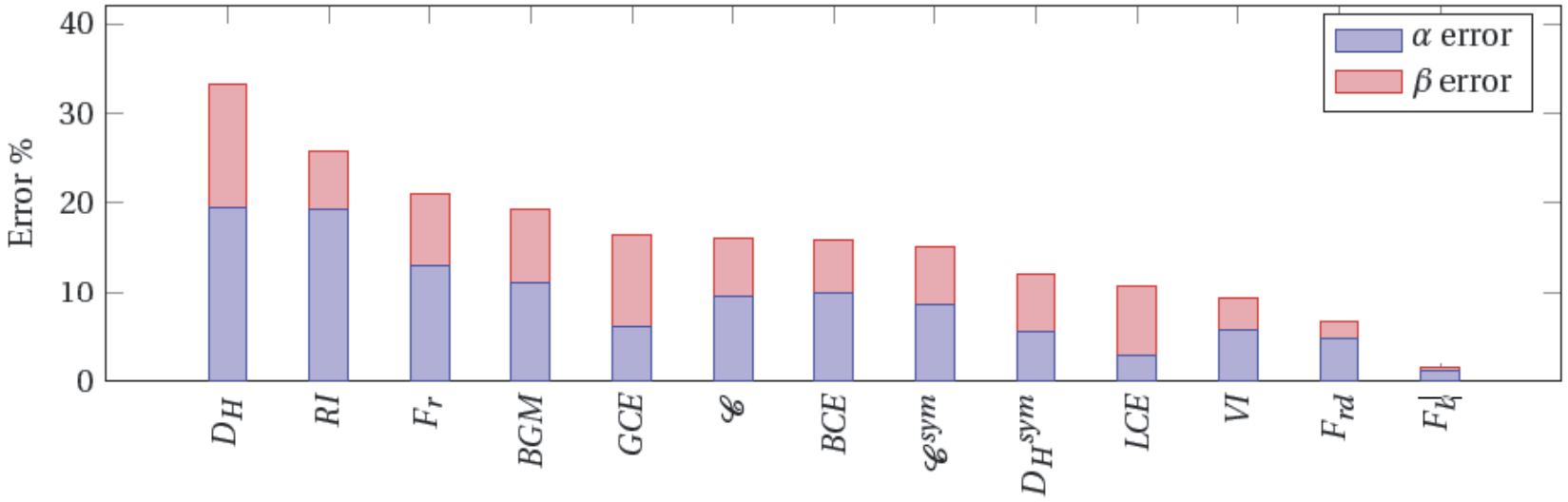




Assessing criteria (III)

How to assess the quality of a partition?

- From the whole set of assessed quality measures, the so-called **region detection F-measure** (F_{rd}), and **boundary F-measure** (F_b) outperform all the other measures in terms of Same versus Different image discrimination.
  - Both measures are a trade-off between **Precision and Recall**



Overall Presentation

- Introduction
  - Global problem introduction
  - Specific problem statement
- Contextualized scenarios
  - Specific approach
  - Generic approach
- Region-based image representation
  - Binary Partition Tree
  - Merging criteria
  - Assessing criteria
- Region-based object representation
  - Object definition
  - Perceptual object model
  - Examples
- Other topics
  - Structural object model
  - Graphic User Interface and Training
  - Query by Example
  - Extension to video analysis
  - Soccer analysis applications
- References and Conclusions

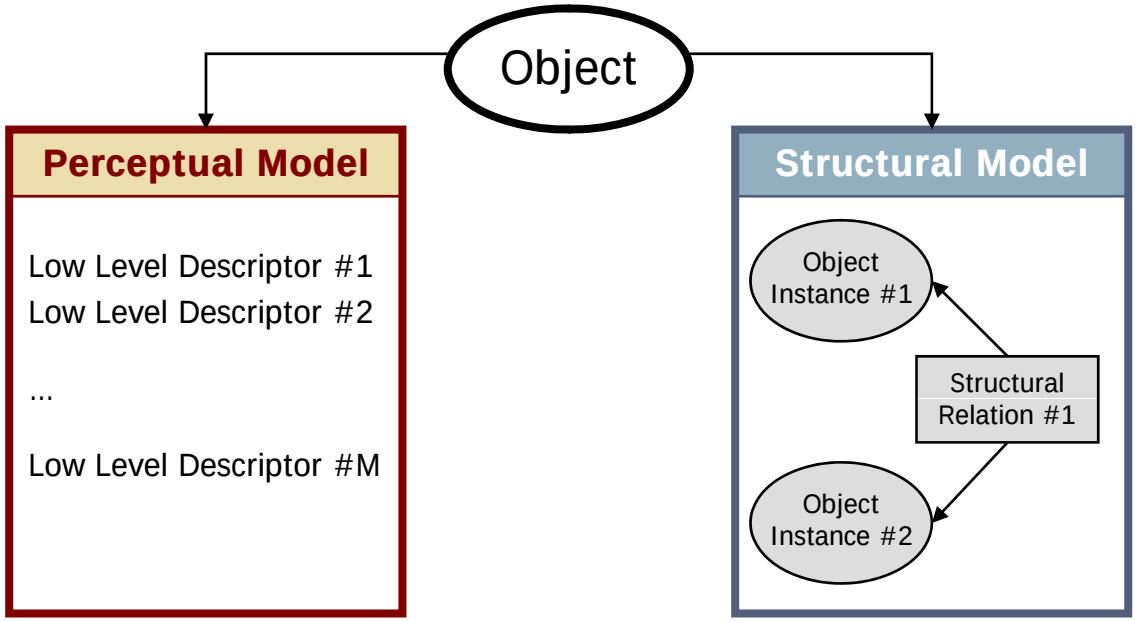
Global Object Model (I)

Objective:

- To define a model for Semantic Entities (objects) that can cope with the various **levels of variability** of objects.

Approach:

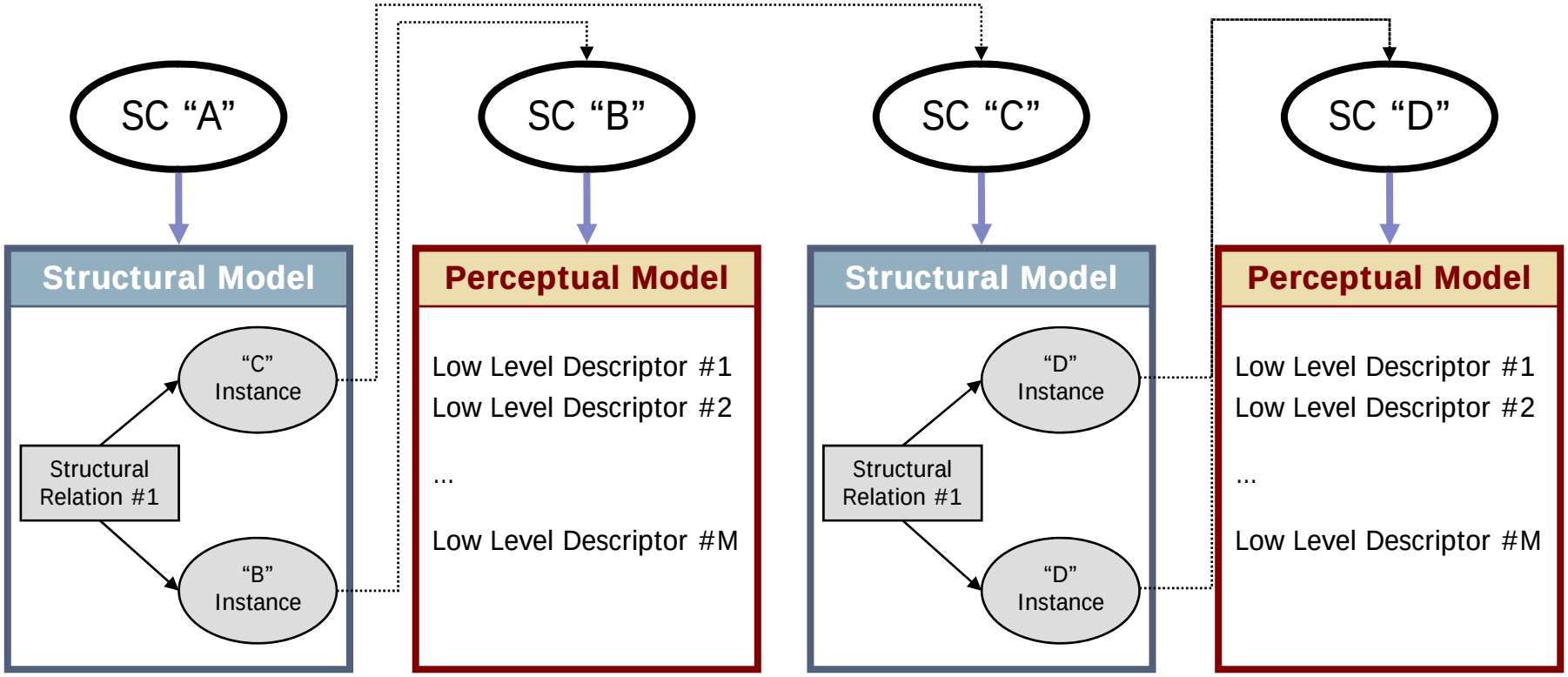
- A **hierarchical** representation of objects:
  - Low variability: Perceptual model
  - High variability: Structural model



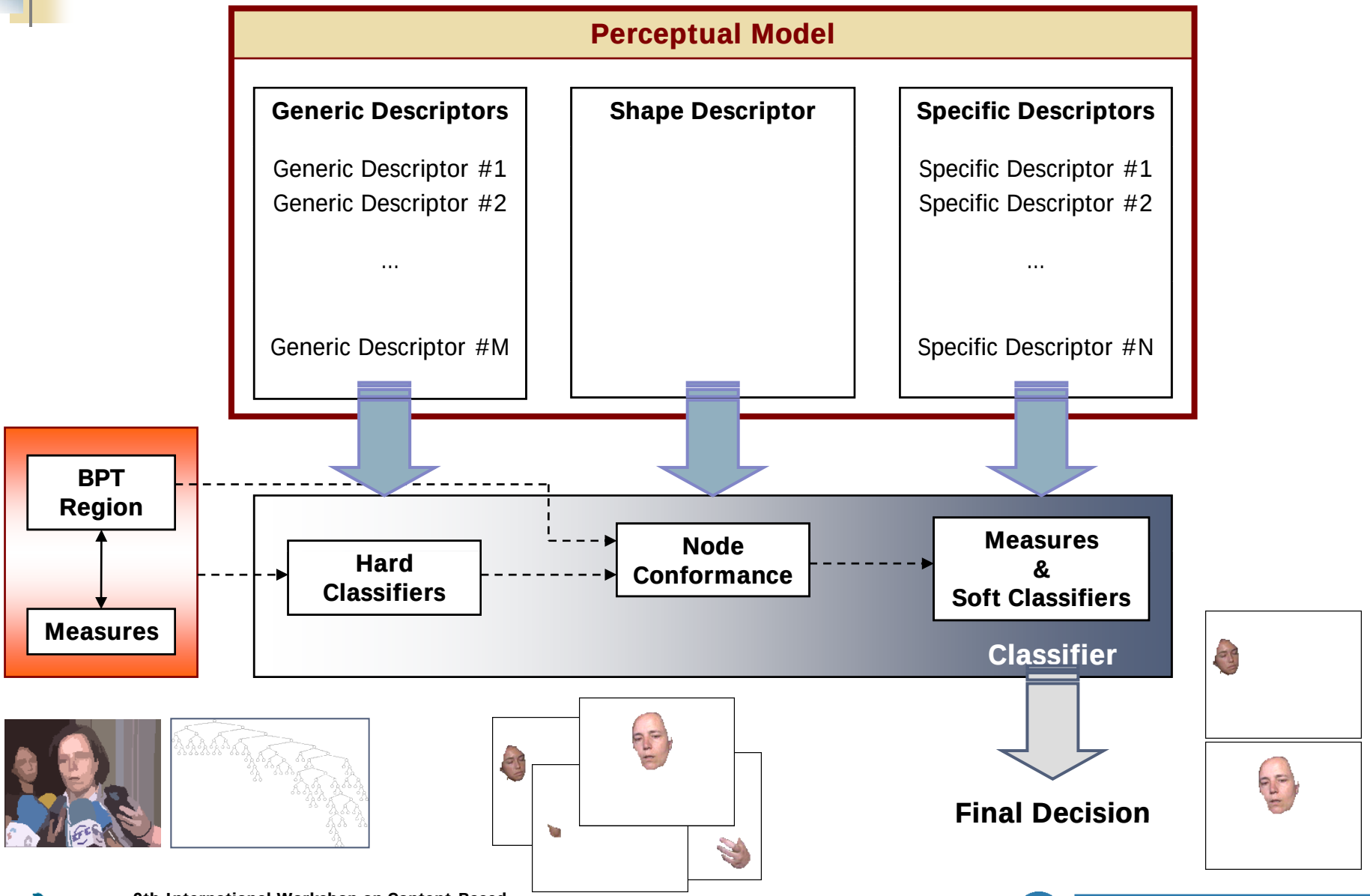
Global Object Model (II)

The model is **hierarchical** in the sense that the object instances of simpler objects that define a structural model can be modeled by:

- Instances of other structural models, or
- Perceptual models.



Perceptual Model (I)



Perceptual Model (II)

Generic Descriptors: Associated with low-level visual attributes common to any object. They are simple and easy to measure. They can be computed and stored at ingesting time. Each descriptor is associated with a simple threshold classifier.

- **Geometry and shape**: *Position*, *Size*, *Orientation*, *Aspect ratio*, *Oriented aspect ratio*, *Compactness*, *Circularity*
- **Color**: *Color mean*
- **Texture**: *Homogeneity*

Shape Information: Information about the object shape is used to modify the area of support of the node to conform to the shape model.

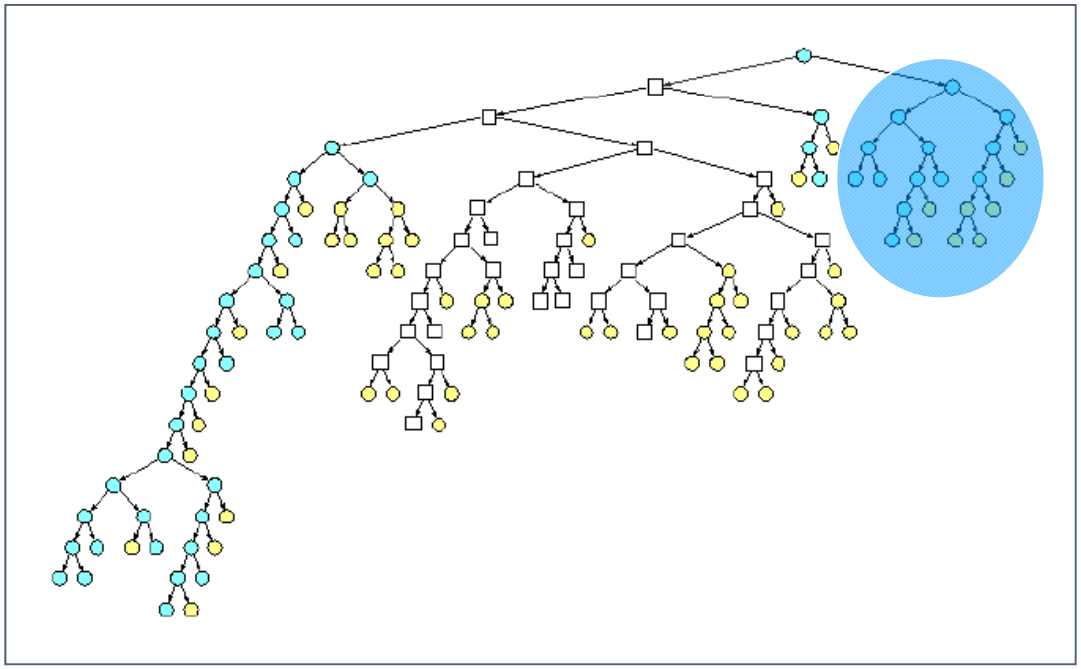
Specific Descriptors: More complex and costly, but they are computed on a few object-conformed nodes. Specific for each type of object.

- **Geometry and shape**: *Symmetry, Hausdorff distance*
- **Color**: *Dominant Colors, Color Histogram*
- **Texture**: *PCA coefficients, Haar coefficients*

The classifiers are trained with different techniques (density methods, SVDD, reconstruction error, Real-Adaboost, etc.)

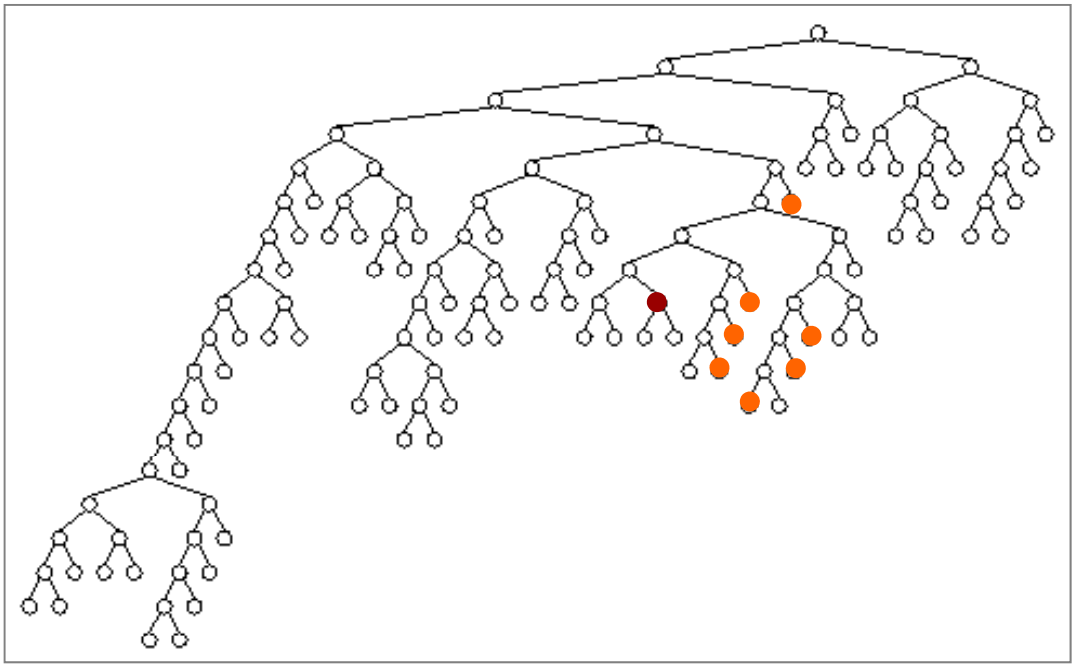
Perceptual Model: Frontal Face (I)

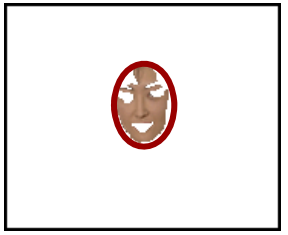
- Color descriptor (in blue):
  - Regions with mean color very different from usual skin colors.
- Geometrical descriptors (in yellow):
  - Regions with unlike aspect ratio or too small.

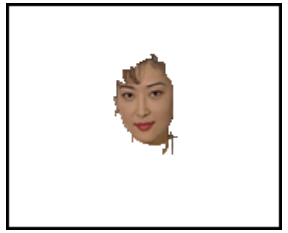


Perceptual Model: Frontal Face (II)

- Candidate selection (in maroon):
  - Non-complete representation of the object.
- Shape descriptor (in orange):
  - Union of regions that may not be linked in the BPT.







Perceptual Model: Frontal Face (III)

Attributes: Frontal or nearly frontal views of human faces can be described in terms of skin-like color regions with a close to elliptical shape. Faces present a specific texture pattern due to the relative positions of eyes, nose and mouth.

Generic Descriptors:

- *Color mean*
- *Size, Aspect ratio, Compactness, Circularity*
- *Homogeneity*

Shape Information:

- Fitting with an ellipse

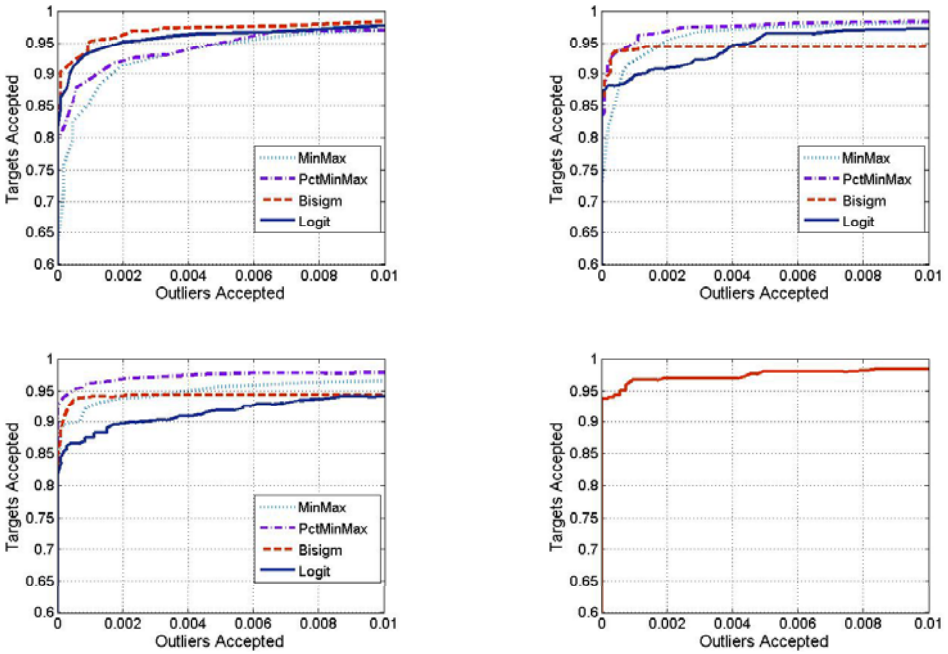
Specific Descriptors:

- *Dominant Colors*
- *Hausdorff distance, Symmetry*
- *PCA coefficients, Haar coefficients*

Perceptual Model: Frontal Face (IV)

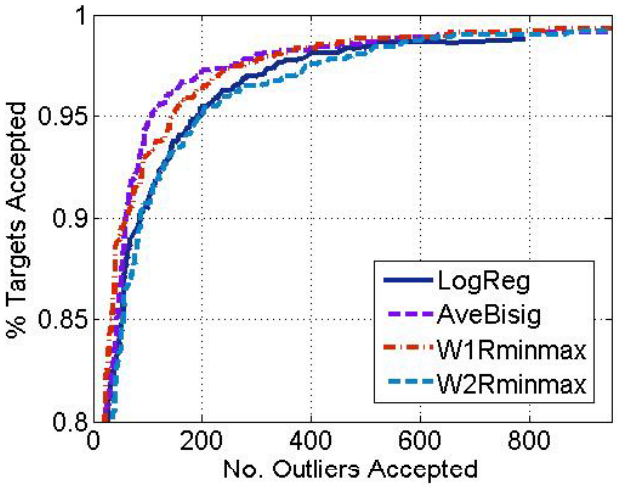
To fuse the different classifiers, several aspects have to be further analyzed:

- **Diversity of classifiers**: To select a set of diverse classifiers such that the their combination improves the performance of the best single classifier
- **Normalization**: To transform outputs into a common domain before combining them
  - Min-Max
  - Robust Min-Max
  - Double sigmoid
  - Logistic regression
- **Combination rules**:
  - Simple average (1)
  - Max, min, median
  - Trimmed mean
  - Product
  - Weighted average (2, 3)
  - Logistic regression (4)



Perceptual Model: Frontal Face (V)

Data Base: 1020 faces in 980 images, large variability, complex background



Viola and Jones:
93% 98 false positives (max)

Perceptual Model: Caption text (I)

Attributes: Caption text can be broadly described as text typed inside a rectangular box, horizontally aligned, highly contrasted with respect to the background and with a characteristic textured aspect.

Generic Descriptors:

- *Aspect ratio, Compactness*
- *Homogeneity*

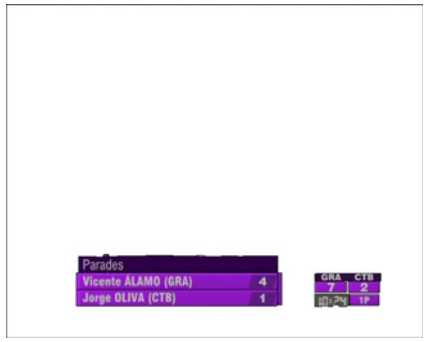
Shape Information:

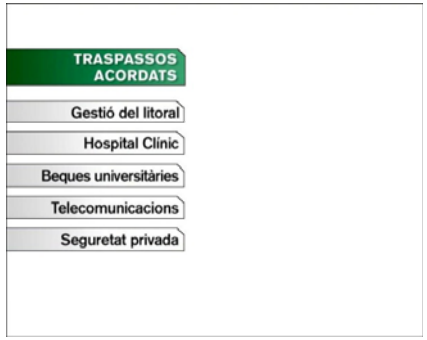
- Fitting to a rectangle

Specific Descriptors:

- *Haar coefficients*
- *Homogeneous Texture Descriptor*

Perceptual Model: Caption text (II)





Perceptual Model: Sky region (I)

Attributes: Sky colors may cover a broad range from saturated blue to gray. Sky may appear as one or several connected components which are more likely to be found on the top of the image, and to present a smooth texture.

Generic Descriptors:

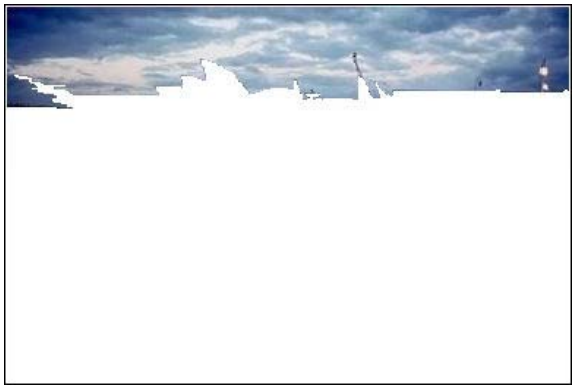
- *Position*
- *Color mean*
- *Homogeneity*

Shape Information:

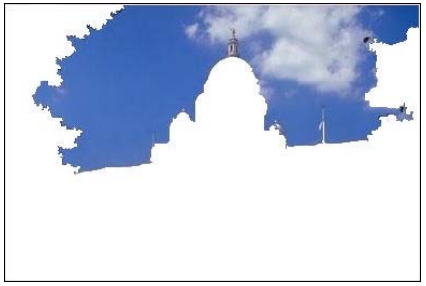
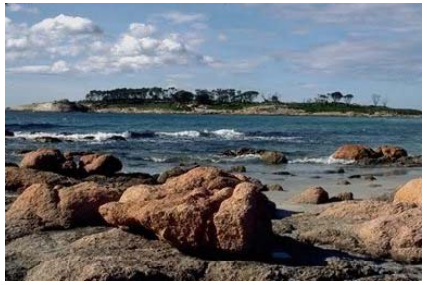
- Connection among components
- Position within the image

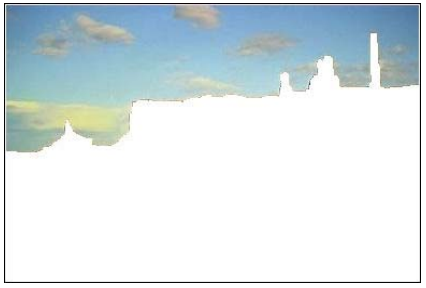
Specific Descriptors:

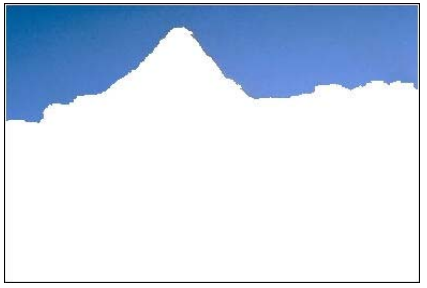
- *Dominant Colors*



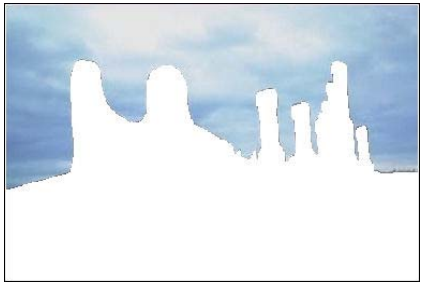
Perceptual Model: Sky region (II)







Perceptual Model: Traffic Signals

Attributes: European warning traffic signs are defined by their triangular shape and their color distribution: a white triangle, that may contain some black structures, surrounded by a red frame.

Generic Descriptors:

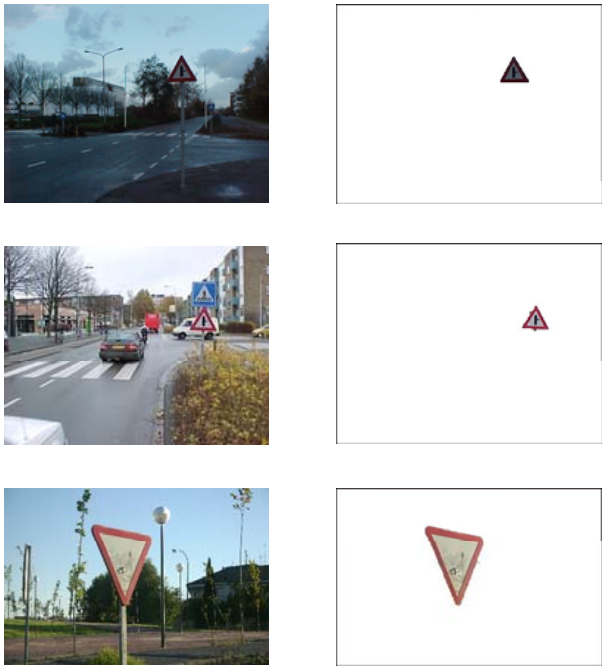
- *Size, Aspect ratio, Compactness*
- *Color mean*
- *Homogeneity*

Shape Information:

- Fitting to a triangle shape (rotations)

Specific Descriptors:

- *Color Histogram*
- *Hausdorff distance*



Perceptual Model: Summary

	SKY	TEXT	SIGN	FACE
Objects for training	50	100	30	2000
Img. with+without obj.	300+100	100+50	70+50	2550+100
Objects for detection	300	249	70	2590
RECALL	0.97	0.86	0.94	0.97
PRECISION	0.93	0.89	0.99	0.95
Obj. with ground truth	200	130	70	200
Segmentation accuracy	0.92	0.89	0.98	0.88

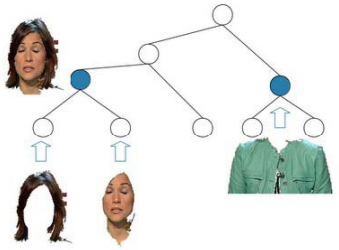
Overall Presentation

- Introduction
  - Global problem introduction
  - Specific problem statement
- Contextualized scenarios
  - Specific approach
  - Generic approach
- Region-based image representation
  - Binary Partition Tree
  - Merging criteria
  - Assessing criteria
- Region-based object representation
  - Object definition
  - Perceptual object model
  - Examples
- Other topics
  - Structural object model
  - Graphic User Interface and Training
  - Query by Example
  - Extension to video analysis
  - Soccer analysis applications
- References and Conclusions

Other topics (I)

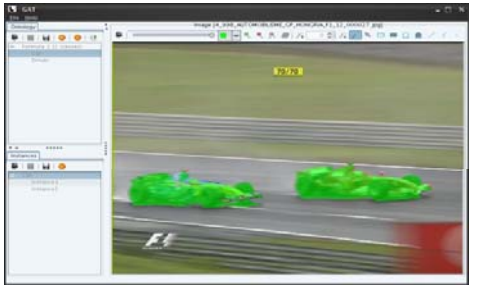
New work in the **structural object model**.

- Automatic decomposition **its main parts**
- **Unsupervised clustering** of similar parts
- Handles **multiple view detection**



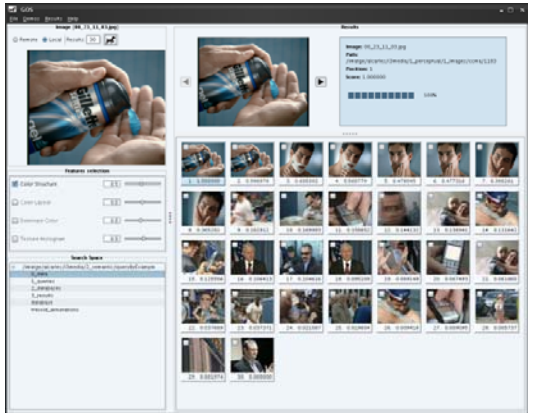
Specific work in **GUI for annotation**:

- Including **new interactions**
- Allowing **region-based** image annotation.
- Towards **video annotation**.



Specific work in **Query by Example**:

- Having **complete images** as example
- Towards **region based** searches
- Including **feed-back** and **clustering**



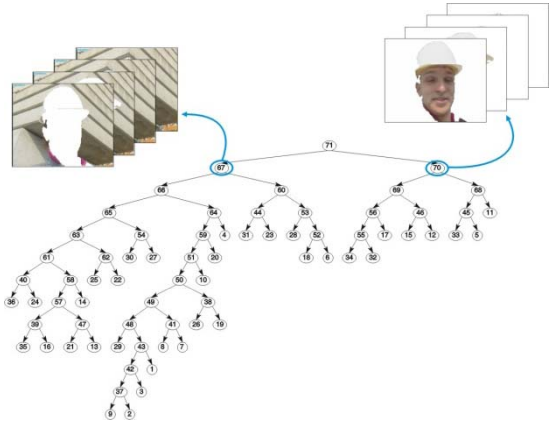
Other topics (II)

Not everything can be easily described in terms of regions.

- Combination with **salient points** approach:

Extensions to video:

- **Trajectory tree** for global video analysis:

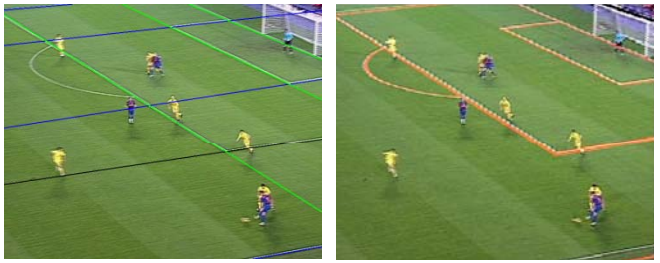


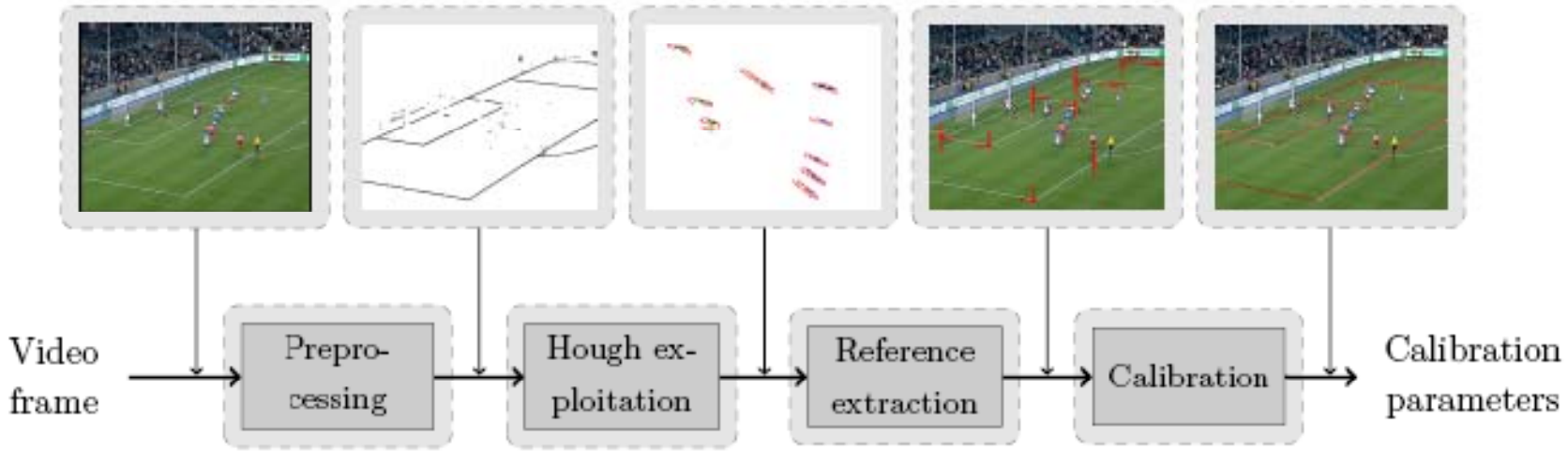
- Region-based **object tracking**.

Other topics: Football analysis (I)

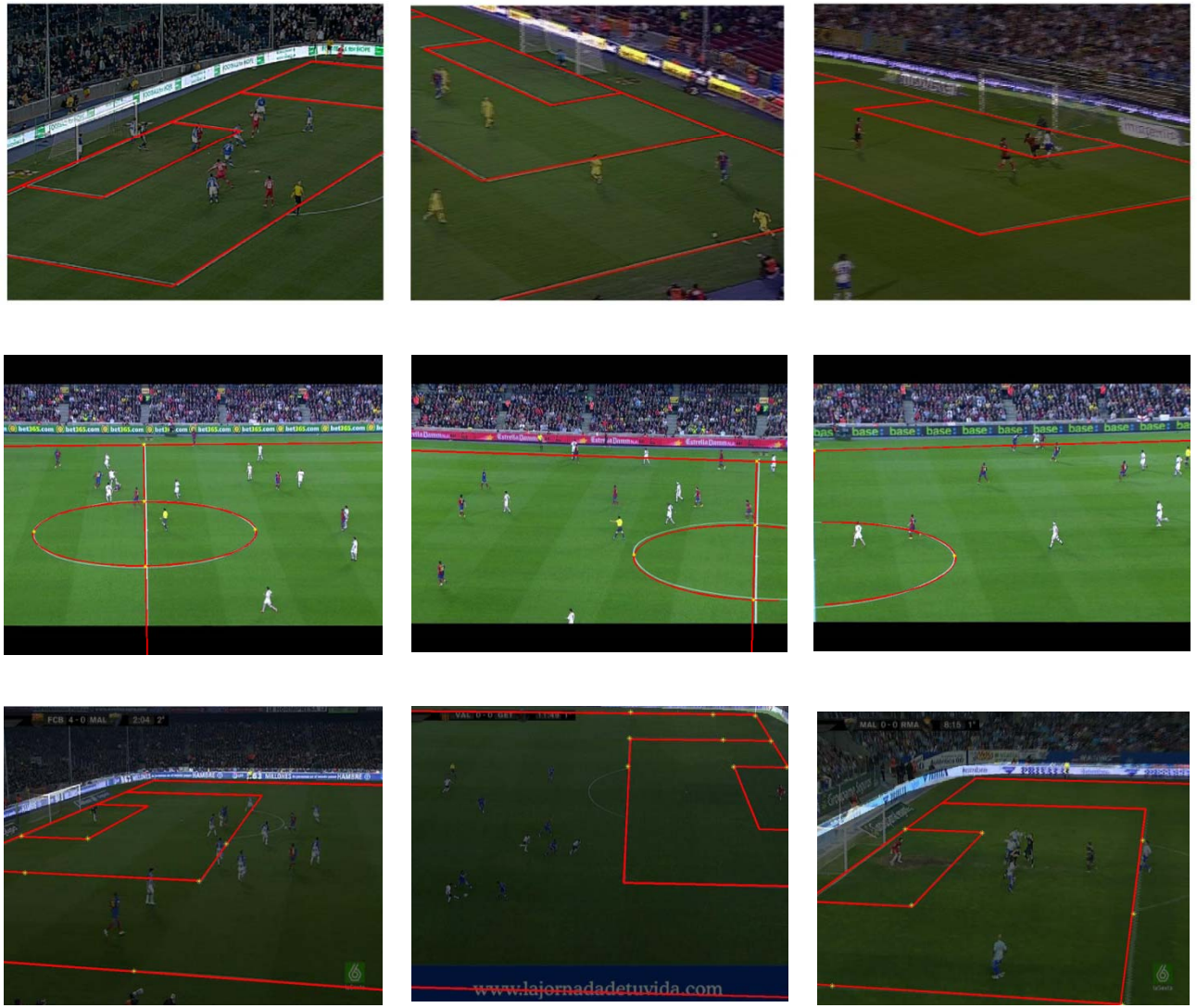
Specific work in the context of **football games**.

- Extraction of the **camera parameters**
- Camera **view** classification
- Towards **event classification**



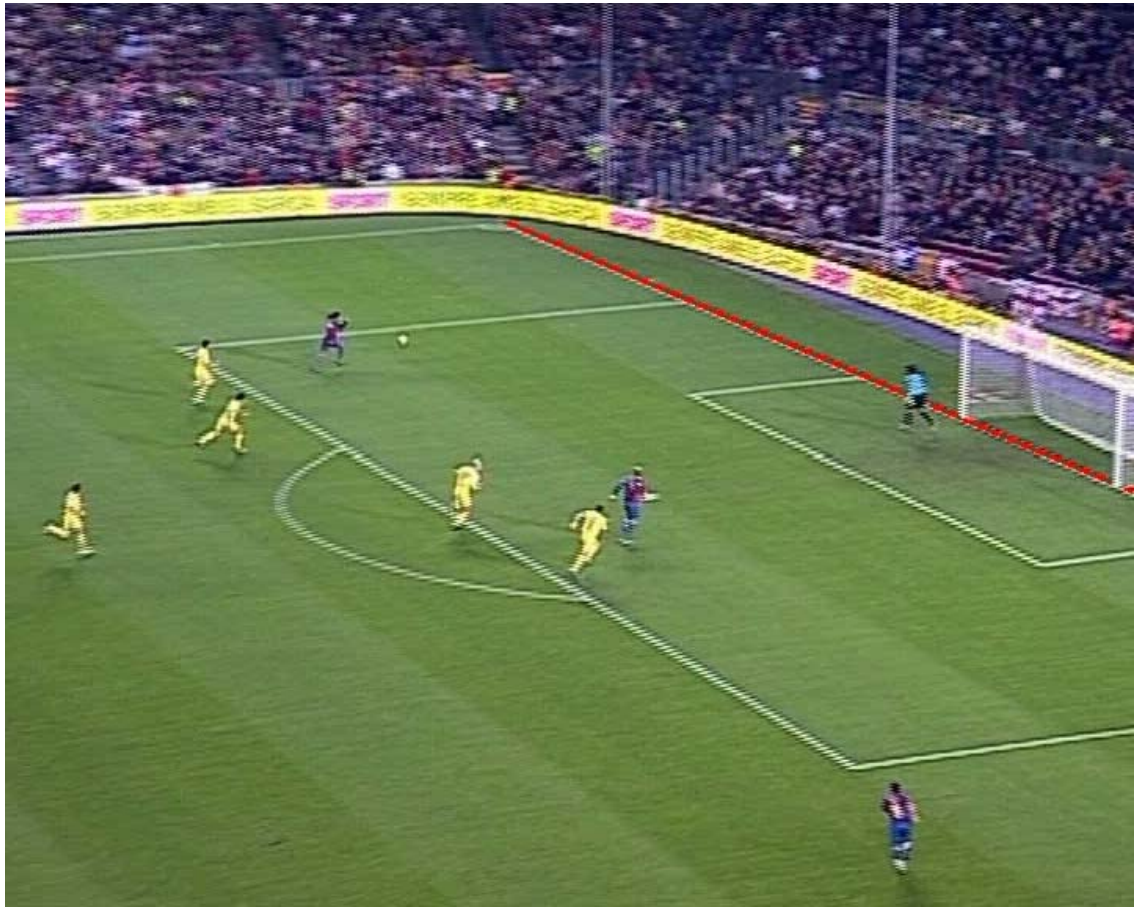


Other topics: Football analysis (II)



Other topics: Football analysis (III)

Once the camera parameters have been identified as well as the various lines in the field layout, several applications can be tackled: **Off-side analysis**



Other topics: Football analysis (IV)

Once the camera parameters have been identified as well as the various lines in the field layout, several applications can be tackled: **Ball position**

References (I)

- Region-based analysis
  - P. Salembier and F. Marqués. ***Region-based representations of image and video: Segmentation tools for multimedia services***. IEEE Trans. on Circuits and Systems for Video Technology, 9(8):1147-1169, Dec. 1999.
- Binary Partition Tree
  - P. Salembier, L. Garrido. ***Binary Partition Tree as an Efficient Representation for Image Processing, Segmentation, and Information Retrieval***, IEEE Trans. on Image Processing, 9(4):561-576, April 2000.
  - V. Vilaplana, F. Marqués, P. Salembier, ***Binary Partition Trees for Object Detection***, IEEE Transactions on Image Processing, vol. 17, nº 11, pp. 2201-2216, November 2008.
  - F. Calderero, F. Marqués, ***Region Merging Techniques Using Information Theory Statistical Measures***, IEEE Transactions on Image Processing, vol. 19, nº 6, pp. 1567 - 1586, June 2010
- Perceptual/Structural model
  - X. Giro, V. Vilaplana, F. Marqués, P. Salembier. ***Chapter 7: Automatic extraction of visual objects information***, in Multimedia content and Semantic Web edited by G. Stamou and S. Kollias, Wiley. May 2005.
- Graphic User Interfaces
  - X. Giró, N. Camps, F. Marqués, ***GAT: A graphical annotation tool for semantic regions***, on Multimedia Tools and Applications, Springer, Volume 46, Issue 2, pp. 155-174, January 2010.

References (II)

- Perceptual analysis
  - F. Marqués, V. Vilaplana. ***Face segmentation and tracking based on connected operators and partition projection***. Pattern Recognition, 35(3):601-614, 2002.
  - M. Leon, V. Vilaplana, A. Gasull, F. Marqués, ***Caption text extraction for indexing purposes using a hierarchical region-based image model***, Proceedings of the ICIP-09, IEEE International Conference on Image Processing, pp. 1869 - 1872, El Cairo, Egypt, November 2009
  - V. Vilaplana, F. Marqués, M. Leon, A. Gasull, ***Object detection and segmentation on a hierarchical region-based image representation***, Proceedings of the ICIP-10, IEEE International Conference on Image Processing, pp. 3393-3396, Hong Kong, China, September 2010
- Structural analysis
  - X. Giró, F. Marqués ***Detection of Semantic Objects using Description Graphs***, IEEE International Conference on Image Processing, pp. 1201 – 1204, Geneva, Italy. September 2005.
  - X. Giro, F. Marques, ***From Partition Trees to Semantic Trees***, Proceeding of International Workshop Multimedia Content Representation Classification and Security (MRCS), p.306-313. Lectures Notes in Computer Science (LNCS 4105), Springer. Istanbul, Turkey, setiembre de 2006
- Video extensions
  - C. Dorea, M. Pardàs, F. Marqués, ***Trajectory Tree as an Object-oriented Hierarchical Representation for Video***, IEEE Trans. on Circuits and Systems for Video Technology, vol. 19, nº 4, pp. 547-560, April 2009.
  - V. Vilaplana, F. Marqués, ***Region-based mean shift tracking: Application to face tracking***, IEEE International Conference on Image Processing, pp. 2712-2715, San Diego, USA, October 2008
- Content-based image retrieval
  - X. Giro, C. Ventura, J. Pont, S. Cortes, F. Marqués, ***System architecture of a web service for Content-Based Image Retrieval,*** Proceedings of the CIVR-10, ACM, pp. 358-365, Xian, China, julio de 2010

9th International Workshop on Content-Based Multimedia Indexing (CBMI-2011, Madrid)

Region-based Semantic Image Analysis: Application to Image and Video Indexing

J.R. Casas, F. Calderero, A. Gasull, X. Giró, M. León, F. Marqués, M. Pardàs, J. Pont, P. Salembier, D. Varas, C. Ventura, V. Vilaplana

Image and Video Processing Group
Signal Theory and Communications Department
http://gps-tsc.upc.es/imatge/
ETSETB — UPC

IMATGE

UPC Departament de Teoria del Senyal i Comunicacions
UNIVERSITAT POLITÈCNICA DE CATALUNYA