Region-based Segmentation Utilizing Multiple Sources of Information, Its Applications and Evaluation

Tomasz Adamek

<tomasz@tid.es>

CENTRE for DIGITAL VIDEO PROCESSING
Dublin City University, Ireland

Telefonica I+D (R&D)
Barcelona, Spain

Madrid
Date: 2008-06-09

Overview

- Introduction
- Review of the main categories of image segmentation approaches
- Evaluation of image segmentation
- RSST with syntactic visual features
- Semi-automatic segmentation
- Segmentation in K-Space participation in TRECVid'07

Introduction
Definition

- **Image segmentation** – a broad class of processes of partitioning an image into disjoint connected regions which are homogeneous with respect to a certain criteria such as
  - low-level features
  - general prior knowledge about the world
  - high-level knowledge
  - or even user interactions

Introduction
Benefits to CBIR

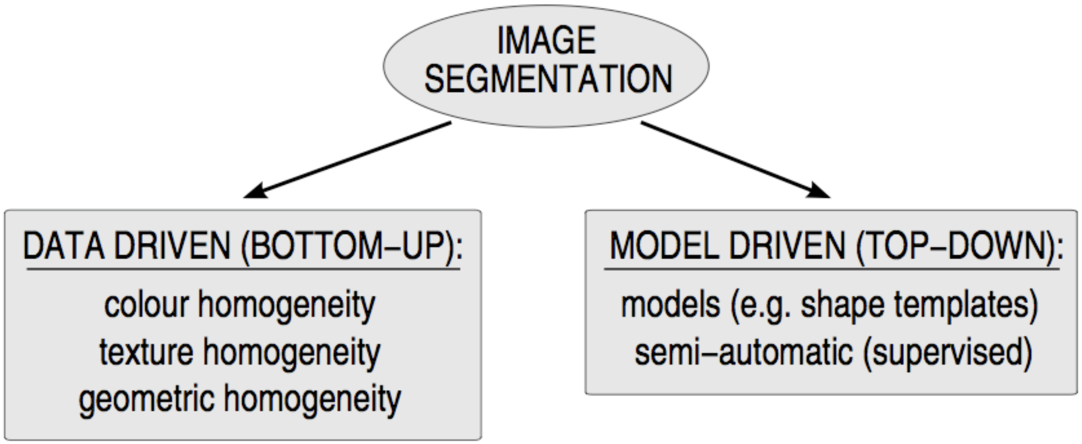
- Spatial image & video segmentation is a fundamental step for many computer vision applications
- Enabling technology for understanding scene structure and identifying relevant objects
- Benefit: representation of scenes at region or object granularity

Overview

- Introduction
- **Review of the main categories of image segmentation approaches**
- Evaluation of image segmentation
- RSST with syntactic visual features
- Semi-automatic segmentation
- Segmentation in K-Space participation in TRECVid'07

Review of the main categories of approaches
Classification of segmentation techniques

- Region-based vs. boundary based
- According to mathematical methodology:
  - variational, statistical, graph-based, and morphological
- Visual features-driven (bottom-up) vs. model-driven (top-down)
- Object-based vs. region-based



Review of the main categories of approaches
Classification of segmentation techniques, cont.

- Visual features-driven and model-driven approaches are not mutually exclusive
  - They could/should be used both for segmentation and object extraction

Review of the main categories of approaches
Grouping cues: low-level features

- **Colour** – primary low-level feature in practically all image segmentation techniques:
  - Choice of an appropriate colour space
  - Perceptually uniform, e.g. CIE LUV or CIE LAB
- **Texture** – arguably the second most important low-level feature:
  - Multi-orientation filter banks
  - The second-moment matrix
  - **NOT A THOUGHTLESS EXTENSION** of the colour feature
    - Adaptive scale selection
    - Filters responses along the direction of the edge
    - Colour and texture often provide conflicting evidence

Review of the main categories of approaches
Grouping cues: “middle level” features

- **General Geometric Cues** – general geometric prior knowledge about the world allowing meaningful segmentation without object specific models
  - Active Contours
  - Perceptual Grouping
    - Motivated by theories from the field of Cognitive Psychology regarding human perception and object recognition
    - Principles of Gestalt theory and the principle of common-cause
  - Integration of such information with other grouping cues is a non-trivial task
  - In many approaches performed in an ad-hoc manner
  - Very little work has been done to quantitatively demonstrate their usefulness

Review of the main categories of approaches
Grouping cues: models

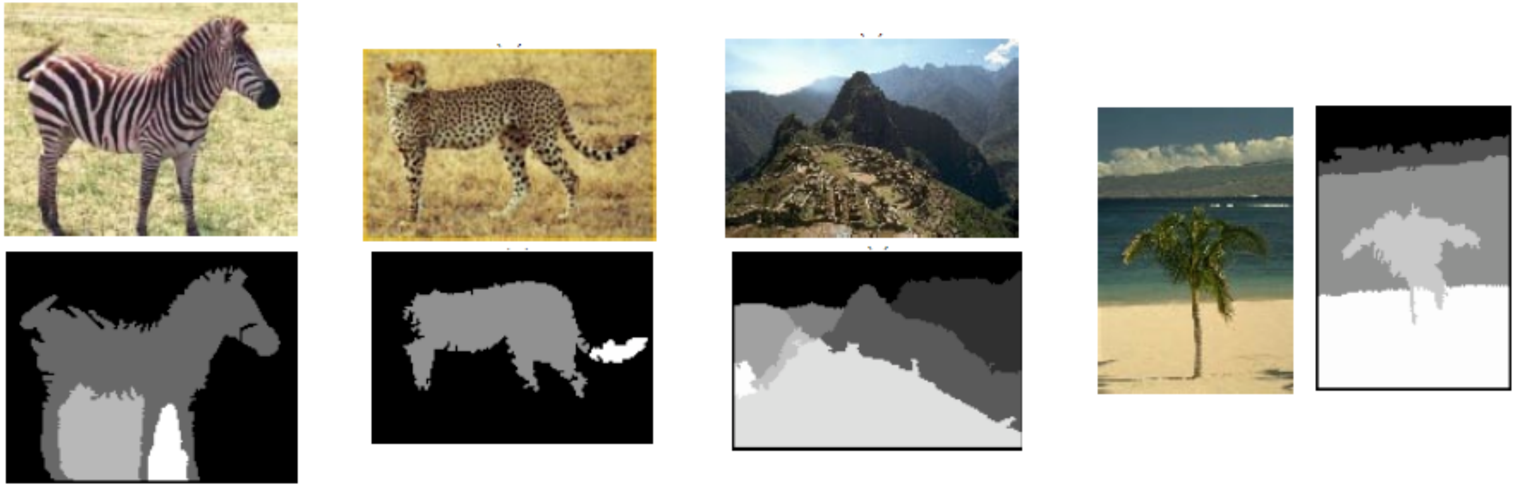
- **Application Specific Models** – partitioning guided by models of specific objects to be segmented
- Object-based segmentation, detection and recognition
- Examples:
  - Shape models
  - Models of colour (or intensity) variations
  - Interest points

Review of the main categories of approaches
Grouping cues: user interactions

- User Interactions (supervised approaches) – a user defines what objects are to be segmented by his/her interactions
- Interactions provide clues as to the nature of the user 's requirements which are then used by an automatic process to segment the required object
- The strategies to user interaction can be grouped in three classes:
  - Feature – based, e.g. markers such as single point seeds or scribbles
  - Contour-based, e.g. contour model placed close to the object's borders
  - Region-based, e.g. merging/splitting regions

Review of the main categories of approaches
Bottom-up: clustering

- K-Means algorithm
  - E.g. **K-Means with Connectivity Constraint (KMCC) [Mezaris et al. 2004]**



- Parametric models of feature distributions
  - E.g. Blobworld [Carson et al. 2002]
- Non-parametric methods based on density estimation
  - E.g. Mean-Shift [Comaniciu & Meer 2002]

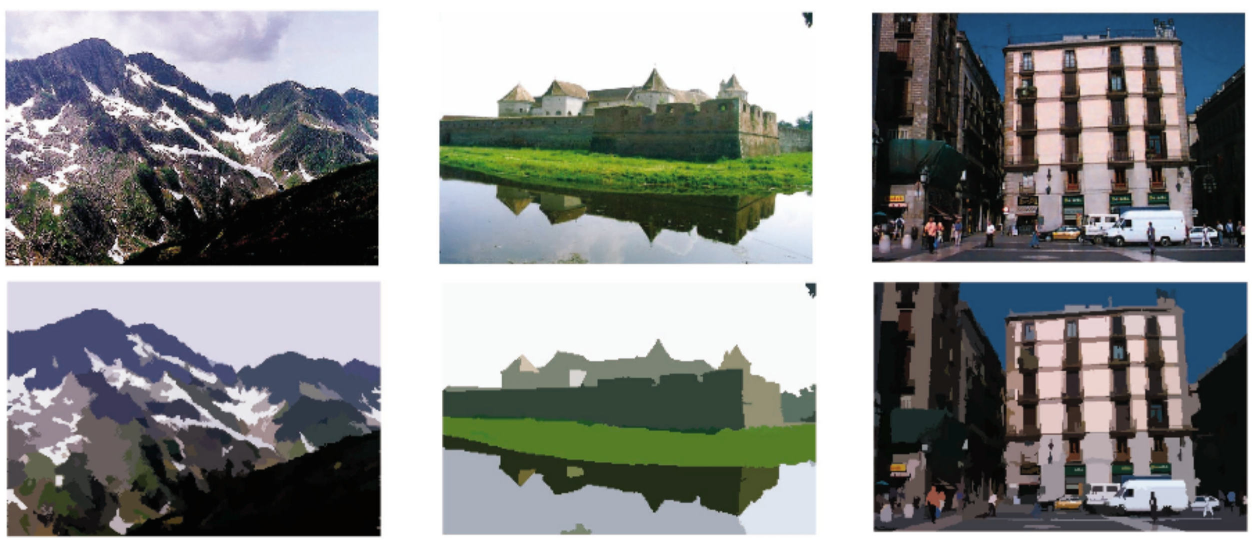
Review of the main categories of approaches
Bottom-up: clustering

- K-Means algorithm
  - E.g. K-Means with Connectivity Constraint (KMCC) [Mezaris et al. 2004]
- Parametric models of feature distributions
  - E.g. **Blobworld [Carson et al. 2002]**

- Non-parametric methods based on density estimation
  - E.g. Mean-Shift [Comaniciu & Meer 2002]

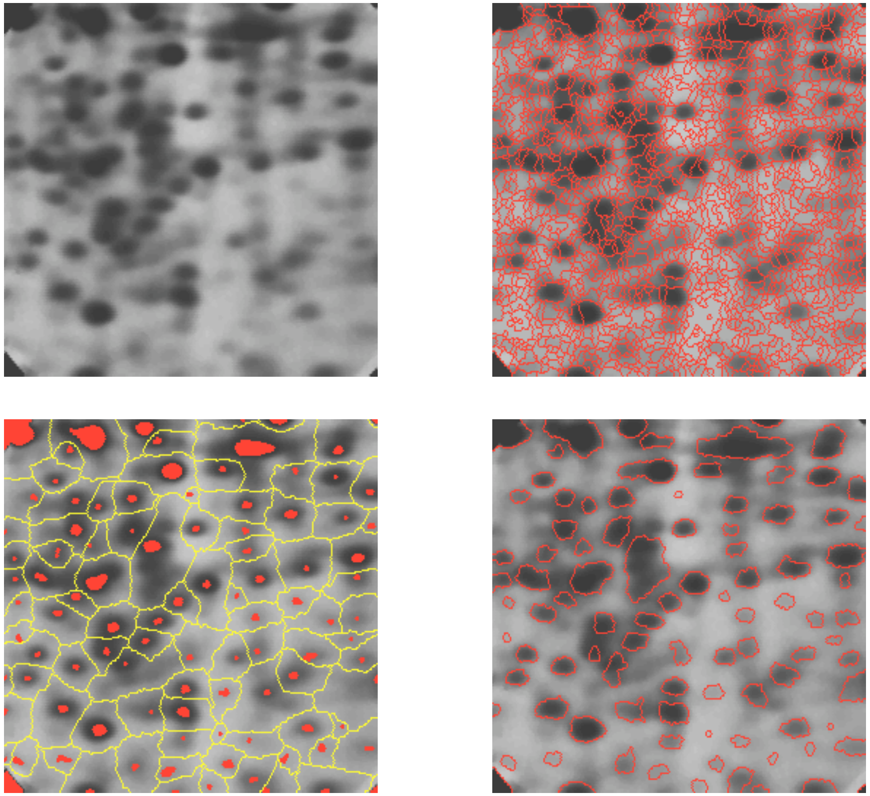
Review of the main categories of approaches
Bottom-up: clustering

- K-Means algorithm
  - E.g. K-Means with Connectivity Constraint (KMCC) [Mezaris et al. 2004]
- Parametric models of feature distributions
  - E.g. Blobworld [Carson et al. 2002]
- Non-parametric methods based on density estimation
  - E.g. **Mean-Shift [Comaniciu & Meer 2002]**



Review of the main categories of approaches
Bottom-up: mathematical morphology

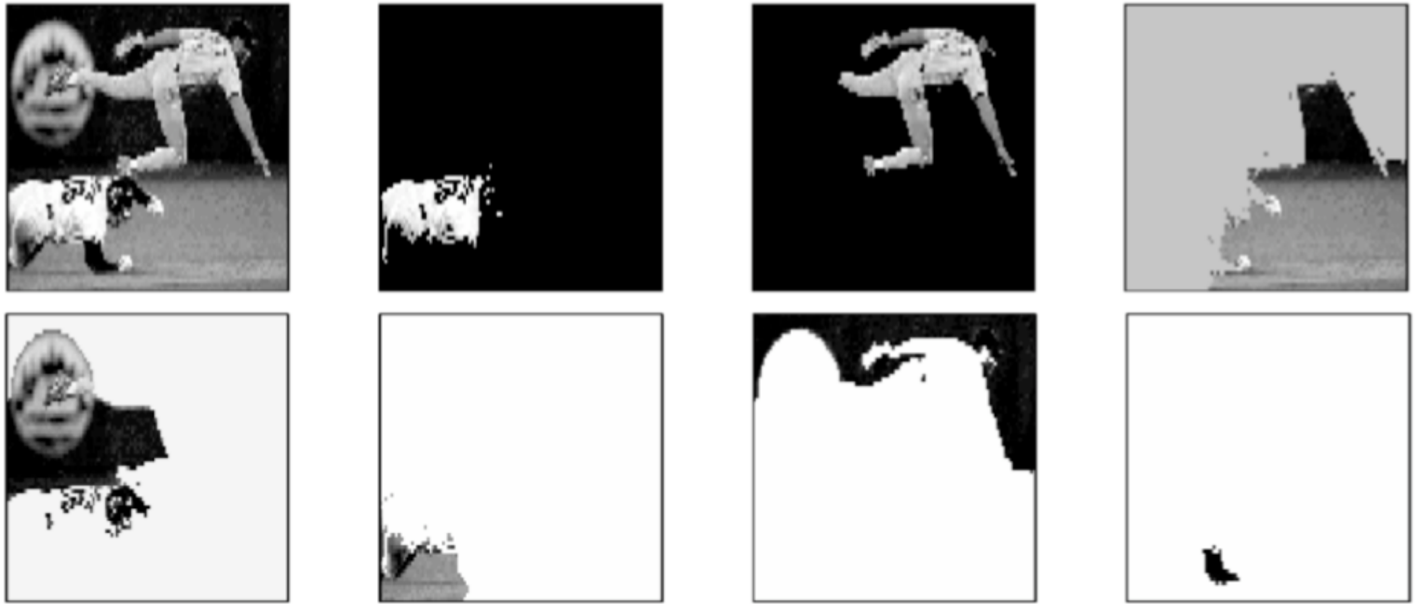
- Mathematical Morphology
  - E.g. Watershed transform [Vincent & Soille 1991]



Images from Serge Beucher

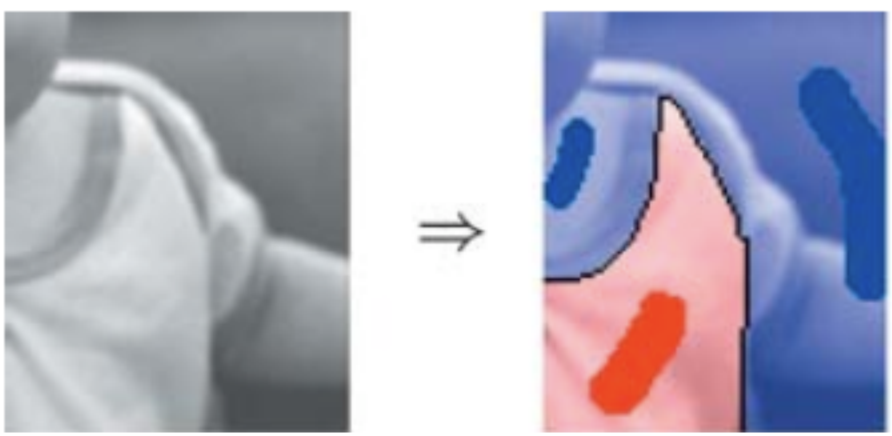
Review of the main categories of approaches
Bottom-up: graph theoretic algorithms

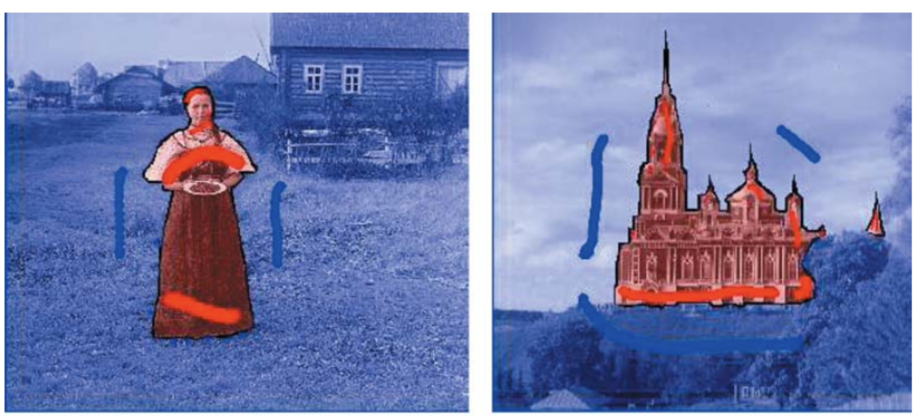
- Region merging
  - E.g. Recursive Shortest Spanning Tree (RSST) [Alatan et al. 1998]
- Region splitting
  - Normalized Cut [Shi & Malik 2000]



Review of the main categories of approaches
Top-down: supervised

- Supervised segmentation based on markers
  - Seeded Region Growing (SRG) [Adams & Bischof 1994]
  - Utilizing Binary Partition Tree (BPT) [Salembier & Garrido 2000]
  - Graph Cuts [Boykov & Jolly 2001]



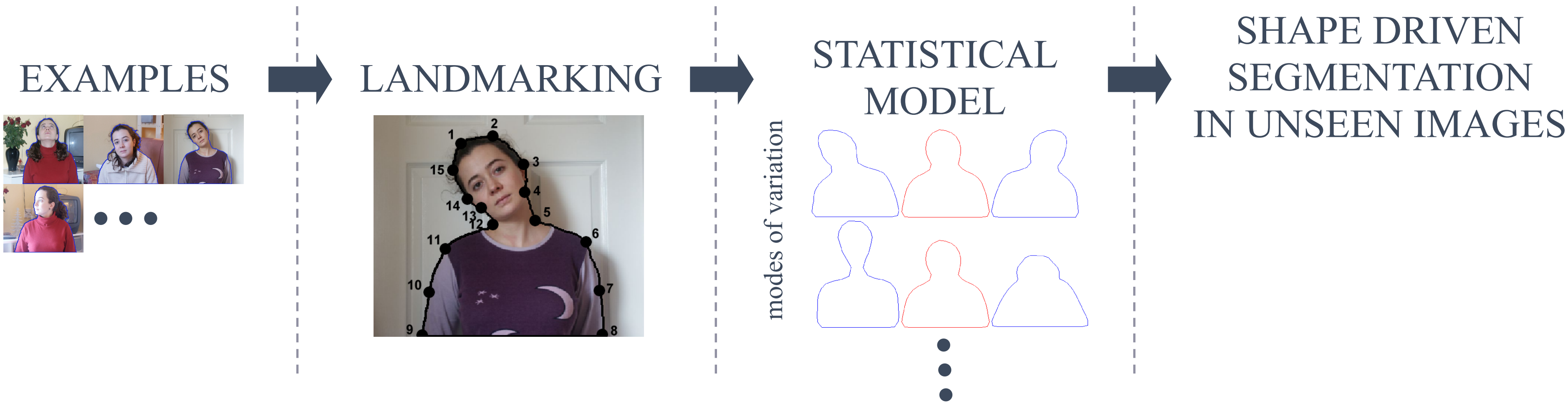


Review of the main categories of approaches
Top-down: supervised

- Supervised segmentation based on contours
  - Active Contours (Snakes) [Kass & Witkin 1988]
  - Geodesic Active Contours [Caselles et al. 1997]
  - Active Regions [Zhu & Yuille 1996]

Review of the main categories of approaches
Top-down: shape driven

- Templates carefully constructed ("hand-crafted") for a particular problem
- Templates derived semi-automatically via statistical analysis of training
  - E.g. Point Distribution Model (PDM) (also called Active Shape Models(ASMs) or "Smart Snakes") [Cootes & Taylor 1992]
  - E.g. Appearance Models (AAMs) [Cootes & Taylor 2004]



Overview

- Introduction
- Review of the main categories of image segmentation approaches
- **Evaluation of image segmentation**
- RSST with syntactic visual features
- Semi-automatic segmentation
- Segmentation in K-Space participation in TRECVid'07

Evaluation of Image Segmentation
Importance of objective assessment

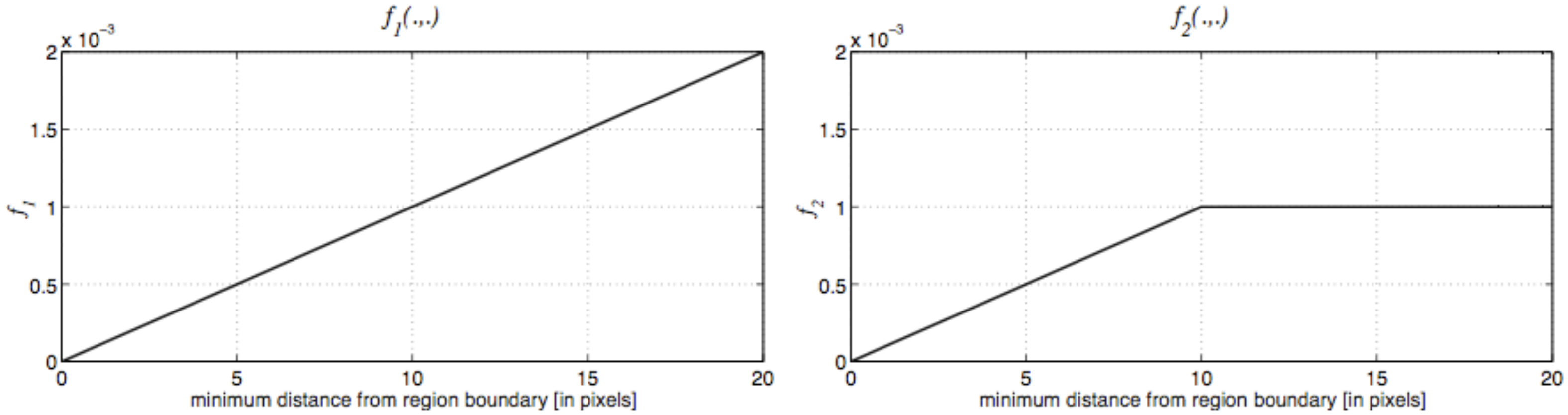
- Meaningful assessment of performance is crucial for the development of any algorithm
- Objective evaluation of segmentation quality has received relatively little attention in the literature compared to the large number of publications on the topic of image segmentation itself
- Crucial for comparison of algorithms
- Could help answering fundamental questions regarding the feasibility of incorporating segmentation in the context of a particular application
- None of the methods proposed in the literature are currently well-established in the research community

Evaluation of Image Segmentation
Evaluation strategies

- Set of qualitative results
- Subjective ad-hoc assessment by a representative group of human viewers, e.g. following recommendations for video quality evaluation developed by ITU [ITU-R, BT. 500-11]
- Measuring the overall performance of the system in which it is integrated
- Objective automatic evaluation
  - **Standalone** (no reference) – exploits available knowledge about expected properties of desirable segmentations for a particular application
  - **Relative** (with ground-truth) – quantify differences between the evaluated and reference segmentation

Evaluation of Image Segmentation
Relative evaluation methods

- Finding a measure assessing segmentation quality aligned with human perception of segmentation quality
- Early approaches:
  - Spatial accuracy measured in terms of misclassified pixels [Wollborn'98]
  - Weighting of misclassified pixels according to their distance to the reference object [Villegas et al. 99]



Evaluation of Image Segmentation
Relative evaluation methods, cont.

- Combination of a comprehensive number of relevant features [Correia et al. 2000]
  - shape fidelity, geometrical similarity, edge content, statistical data similarity, e.g. similarity of the brightness and "redness"
- [Gelasca et al. 2004] demonstrate impact of topology changes
  - Various types of topological errors contribute differently to the overall annoyance, e.g. holes contribute more annoyance compared to added regions of the same size
  - Annoyance values corresponding to added regions quickly reaches saturation
  - Proportions in which they contribute to the annoyance are content dependent

Evaluation of Image Segmentation
Eval. in the context of CBIR systems

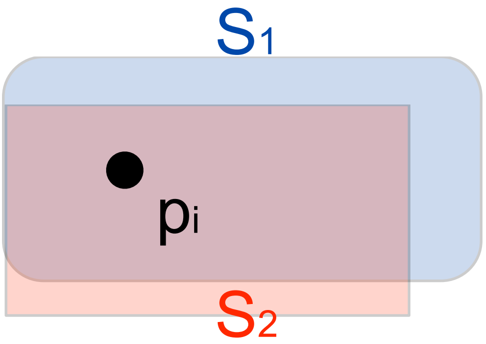
- Assessment integrating **Spatial Accuracy Error** and over- and under-segmentation [Mezaris et al. 2003]
- Both the evaluated segmentation and the ground-truth mask might contain several regions
- Establishing exclusive correspondence between regions in the reference and evaluated masks
- Three different types of errors are taken into account:
  1. **Errors** for the **associated pairs of regions** from both reference and evaluated masks
  2. **Errors** due to **under-segmentation** computed based on non-associated regions from the reference mask
  3. **Errors** due to **over-segmentation** computed from non-associated regions from the evaluated mask

Telefónica

Evaluation of Image Segmentation
Global (GCE) and Local (LCE) Consistency Error

- Based on a definition of local refinement error [Martin et al. 2001]:

$$E(S_1, S_2, p_i) = \frac{|R(S_1, p_i) \setminus R(S_2, p_i)|}{|R(S_1, p_i)|}$$

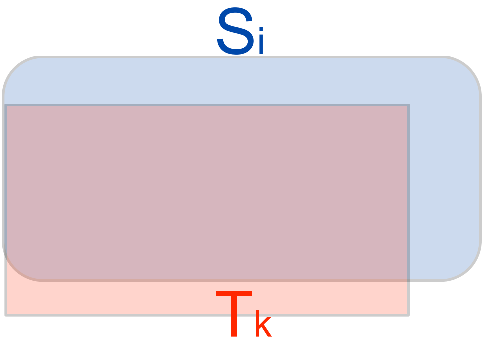


- In the range [0..1]
- Values closer to zero denote a better segmentation
- Correspond well with human perception – very low values when comparing different human segmentations of the same scene
- Not sensitive to over- and under-segmentation

Evaluation of Image Segmentation
Huang-Dom measure (HD)

- Hamming distance between non-maximally intersecting regions [Huang et al. 1995]:

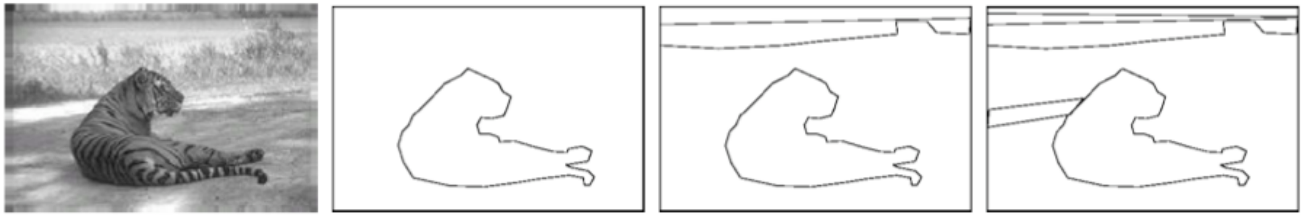
$$D_H(T \Rightarrow S) = \sum_{S_i \in S} \sum_{T_j \neq T_k} |S_i \cap T_j|$$



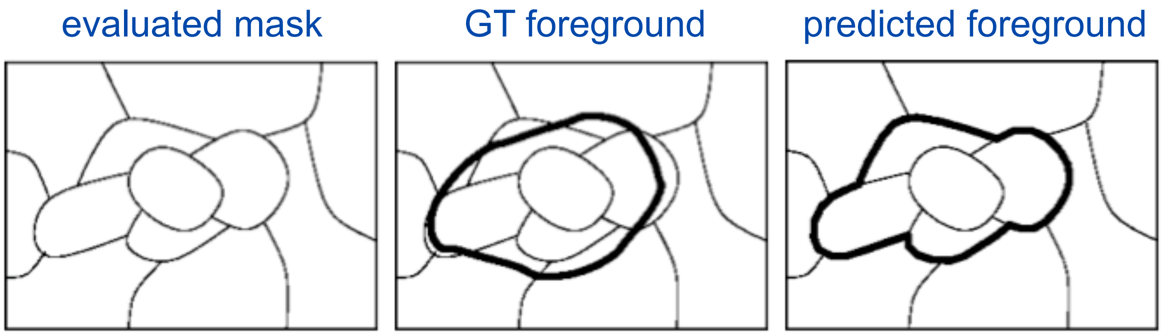
- Where S and T are two segmentations of the same image
- Association of a region Tk with each region Si such that Si ∩ Tk is maximal
- Final HD measure is in the range [0..1]
- Values closer to one denote a better segmentation
- Takes into account over- and under-segmentation

Evaluation of Image Segmentation
Figure ground segmentation evaluation

- The ground truth of general purpose segmentation evaluation may be ambiguous and differ between persons or applications



- Figure ground segmentation evaluation considers only the most salient foreground object for the evaluation [Ge et al. 2006]



Evaluation of Image Segmentation
The problem starts to attract attention

- Very recently the problem started to attract deserved attention within the research community resulting in a several articles devoted solely to this problem:
  - [Chabrier et al. 2006] S. Chabrier, B. Emile, C. Rosenberger, and H. Laurent, “Unsupervised performance evaluation of image segmentation,” EURASIP Journal on Applied Signal. Processing, vol. 2006.
  - [Correia et al. 2006] P. Correia and F. Pereira, “Video object relevance metrics for overall segmentation quality evaluation,” EURASIP Journal on Applied Signal. Processing, vol. 2006.
  - [Jiang et al. 2006] X. Jiang, C. Marti, C. Irniger, and H. Bunke, “Distance measures for image segmentation evaluation,” EURASIP Journal on Applied Signal. Processing, vol. 2006.
  - [Piroddi et al. 2006] R. Piroddi and T. Vlachos, “A method for single-stimulus quality assessment of segmented video,” EURASIP Journal on Applied Signal. Processing, vol. 2006.

Overview

- Introduction
- Review of the main categories of image segmentation approaches
- Evaluation of Image Segmentation
- **RSST with syntactic visual features**
- Semi-automatic Segmentation
- Segmentation in K-Space participation in TRECVid'07

RSST with Syntactic Visual Features
Focus

- Focus on automatic visual feature-based segmentation of images:
  - Partitioning of images into large, visually coherent regions reflecting real objects
  - Generic and broadly applicable, i.e. does not require models of individual objects

RSST with Syntactic Visual Features
Motivations

- The problem is ill posed

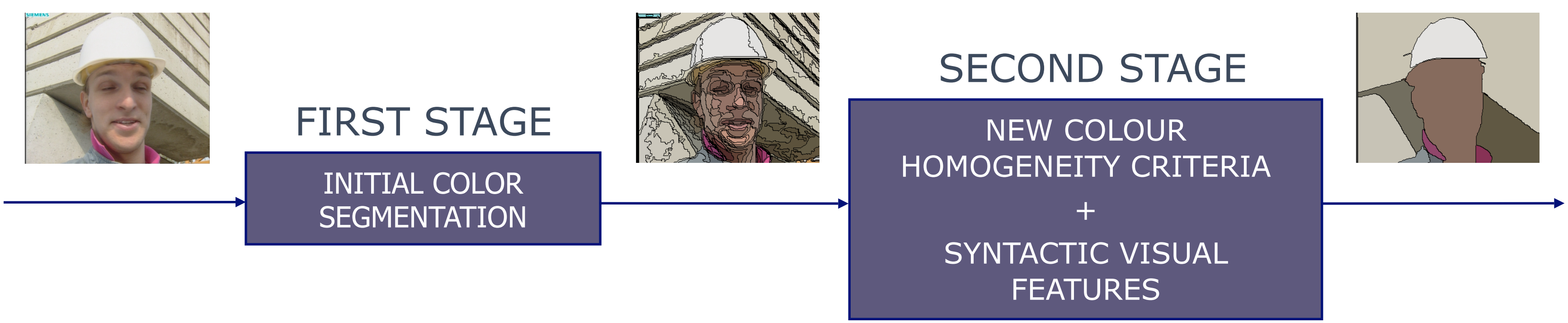
- Then, why are we working on it?
- Initial regions and their local low-level features needed before we can start semantic analysis
- Initial regions needed for speeding-up detection of semantic objects
- Currently automatic segmentation often fails to produce meaningful results

RSST with Syntactic Visual Features
Research objectives

- How far we can get until we need semantic (contextual) information?
- Explore feasibility of utilizing the geometrical properties of regions for improving bottom-up segmentation via region merging
- Investigate new ways of integrating evidence from multiple sources
- Investigate stopping criteria for region merging process aiming at producing partitions containing the most salient regions

RSST with Syntactic Visual Features
Overview of the approach

- Segmentation performed in two stages:
  1. Initial partitioning via RSST algorithm [Alatan et al. 1998]
  2. Merging based on new homogeneity criterion:
     - New colour homogeneity criterion
     - Regions’ geometric properties

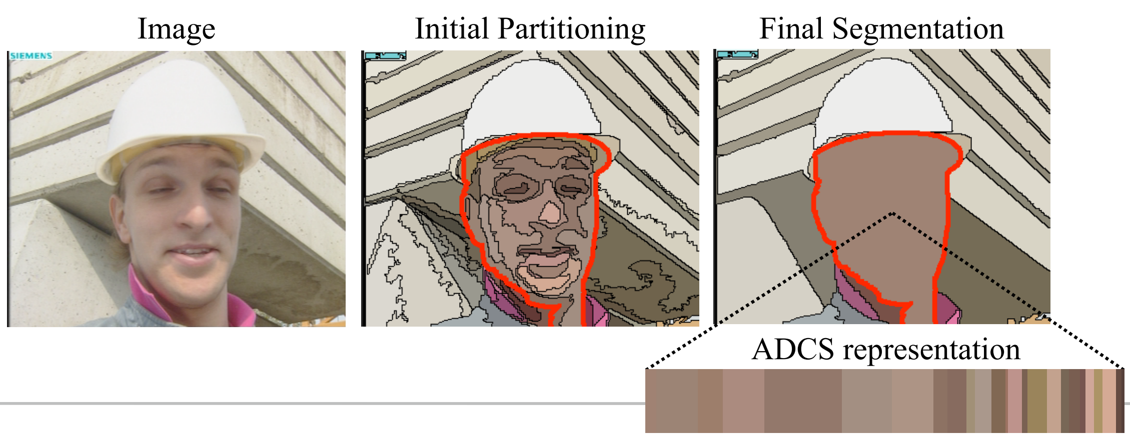


RSST with Syntactic Visual Features
Recursive Shortest Spanning Tree (RSST)

- RSST algorithm [Morris et al. 1986]
  - Image mapped into a weighted graph
  - Regions form the nodes and links between neighboring regions represent merging costs
  - Merging is performed iteratively:
    - At each iteration two regions connected by the least cost link are merged
    - Merging involves creating a joint representation for the new region and updating its links with its neighbors
  - The process continues until a stopping criterion is fulfilled
- All merges can be stored in a Binary Partition Tree (BPT)
- The homogeneity criterion plays a key role in obtaining meaningful hierarchical representation of the scene

RSST with Syntactic Visual Features
Features: extended colour representation

- Average values unsuitable for effective characterization of colour of large regions
- **Adaptive Distribution of Colour Shades (ADCS)** [Frauqueur & Boujemaa 2004]
- Balance between colour distance and the size dependent scaling factor found experimentally
- Two ADCS representations efficiently compared using **Quadratic Distance**



RSST with Syntactic Visual Features
Features: syntactic visual features

- **Syntactic Visual Features**: geometric properties of regions and their spatial configurations, e.g. homogeneity, compactness, regularity, inclusion or symmetry [Bennstrom & Casas 2004]
- Provide additional cues whenever colour information is not sufficient
- The resulting approach is still generic
- More stable basis for semantic knowledge extraction
- Proposed geometric measures:
  - Compactness (adjacency): “real world” objects exhibit adjacency of their constituent parts
  - Regularity (changes in global shape complexity): shape complexity of most objects tends to be rather low

RSST with Syntactic Visual Features
Integration of evidence using DS theory

- Different sources of information provide different accuracy and reliability
- Evidence provided by colour homogeneity and the syntactic features combined using **Dempster-Shafer (DS) theory** [Dempster 1967], [Shafer 1976]:
  - Convenient management of uncertainties
  - Ability of taking into account reliability of information sources
  - Distinct representation of ignorance, imprecision and conflict
  - Convenient incorporation of statistical and/or expert knowledge

RSST with Syntactic Visual Features
Application of DS theory to region merging

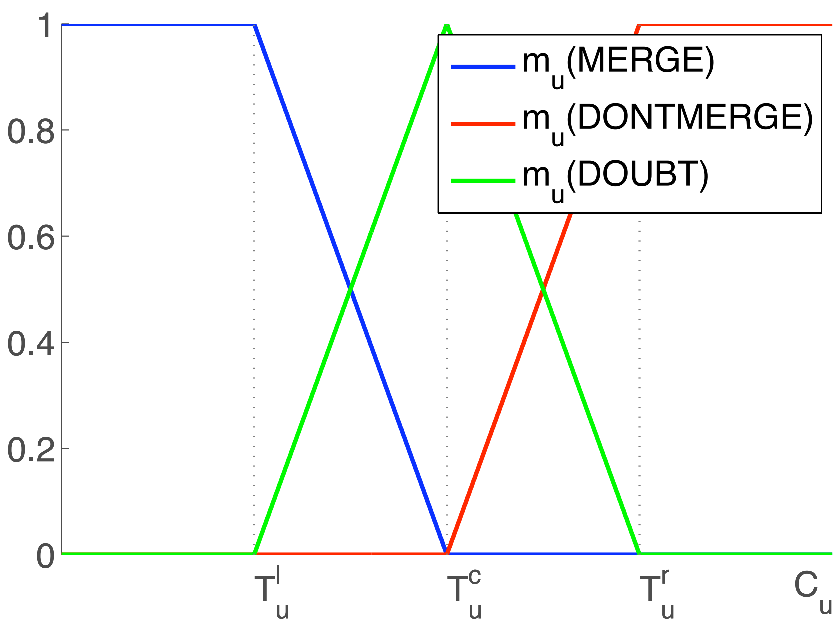
- Frame of discernment: $\Omega = \{ MERGE, DONTMERGE \}$
- Power set 2^{Ω} is composed of four propositions:

$$2^{\Omega} = \{\emptyset, \{MERGE\}, \{DONTMERGE\}, \{MERGE \cup DONTMERGE\}\}$$

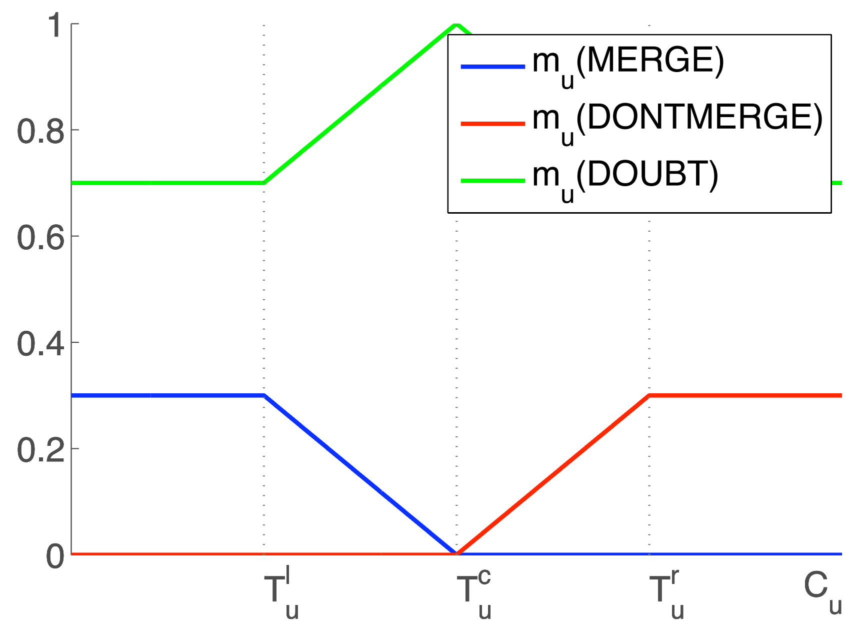
- Basic Belief Assignment (BBA) – models the way a piece of evidence is brought by a source on a proposition

RSST with Syntactic Visual Features
Application of DS theory to region merging, cont.

- $2^{\Omega} = \{\emptyset, \{MERGE\}, \{DONTMERGE\}, \{MERGE \cup DONTMERGE\}\}$
- Basic Belief Assignment (BBA) – associates a belief mass to each of the symbols of 2^{Ω} for each value of the evidence from a given source



BBA for a reliable source



BBA for a unreliable source

RSST with Syntactic Visual Features
Beliefs from multiple sources combined

- Beliefs from multiple sources are combined using Dempster's rule of combination

$$\forall A \subseteq \Omega \qquad m^{\oplus}(A) = \frac{1}{1-K} \sum_{B \cap C = A} m_u(B) \cdot m_v(C)$$

$$K = \sum_{B \cap C = \emptyset} m_u(B) \cdot m_v(C)$$

m_v \ m_u	MERGE	DOUBT	DONTMERGE
MERGE	MERGE	MERGE	∅ (K)
DOUBT	MERGE	DOUBT	DONTMERGE
DONTMERGE	∅ (K)	DONTMERGE	DONTMERGE

Telefónica

RSST with Syntactic Visual Features
New cost of merging of two neighboring regions

- New cost of merging of two neighboring regions r_i and r_j based on evidence from all sources supporting hypothesis MERGE and DONTMERGE:

$$C_{\text{total}}(i,j) = m^{\oplus}_{(i,j)}(DONTMERGE) - m^{\oplus}_{(i,j)}(MERGE)$$

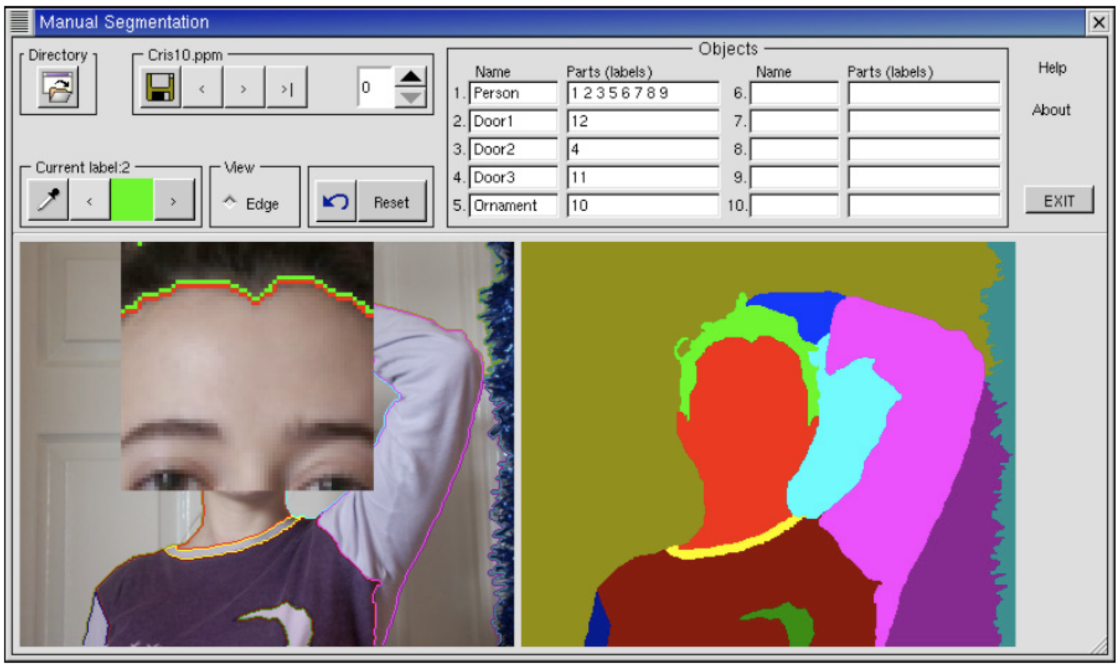
RSST with Syntactic Visual Features
Designing belief structures

- Designing belief structures for each source of information involves:
  - Selection of the thresholds T_u^l, T_u^c, T_u^r and the discounting factor α_u
- Estimation using a training collection of manually segmented images
  - examples of links with an associated set of features and classified as either "**link to be merged**" or "**link which should not be merged**" are automatically generated based on ground-truth segmentation

Telefónica

RSST with Syntactic Visual Features
Experimental setup

- Collection of 120 manually segmented images:
  - 100 images from Corel dataset,
  - 20 images from other sources, MPEG-4 test seq., digital cameras, etc.



RSST with Syntactic Visual Features
Experimental setup, cont.

- 100 images from the Corel gallery and 20 images from well known MPEG-4 test seq. and private collections
- Ground-truth segmentation masks created manually

RSST with Syntactic Visual Features
Experimental setup, cont.

- Relative evaluation in terms of spatial accuracy error, [Mezaris et al. 2003]
- Two-fold cross validation
  - The dataset is divided into two subsets
  - Each time one of them is used for parameter tuning, the other subset is used as the test set
  - The final result is computed as the average error from the two trials
- Merging costs and stopping criteria are developed and tested separately
- Merging criteria evaluated using the “optimal” stopping criterion

RSST with Syntactic Visual Features
Evaluation of merging criteria

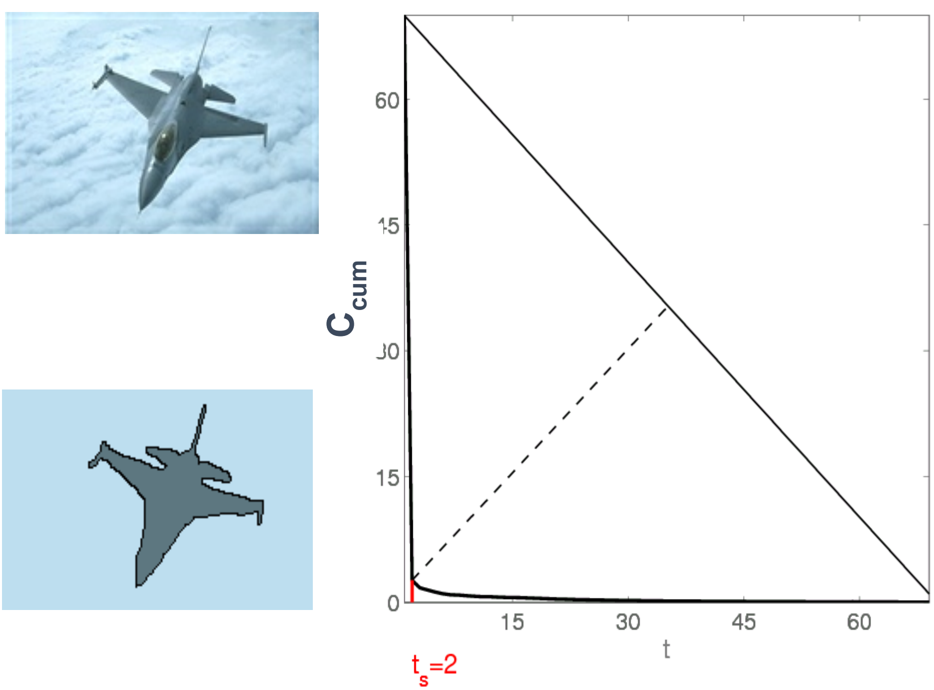
INTEGRATED FEATURES	AVG. SPATIAL SEGMENTATION ERROR
Original colour homogeneity	0.99
New colour homogeneity	0.63
New colour homogeneity + adjacency	0.56
New colour homogeneity + complexity	0.53
New colour homogeneity + adjacency + complexity	0.51

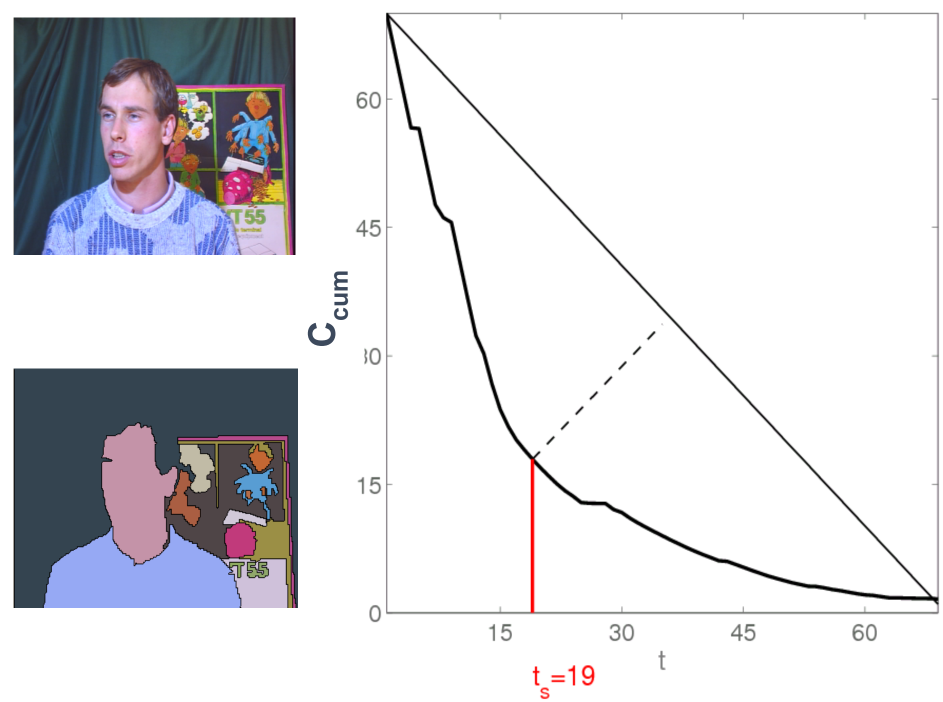
RSST with Syntactic Visual Features
Commonly used stopping criteria

- Required number of regions:
  - Merging process continues until a desired number of regions is reached
  - Independent of the image content
- Maximum value of the least link cost:
  - Merging process is stopped when the merging cost exceeds a predefined threshold
- Value of Peak Signal to Noise Ratio (PSRN) between the orig. and segmented image:
  - All merges stored in a Binary Partition Tree
  - Deactivating nodes following the merging sequence until the value of PSNR falls below a predefined threshold

RSST with Syntactic Visual Features
Proposed stopping criterion

- Accumulated merging cost measure Ccum(t)
  - Total cost of all mergings performed to produce t regions
- Bi-level thresholding, [Rosin 2001]





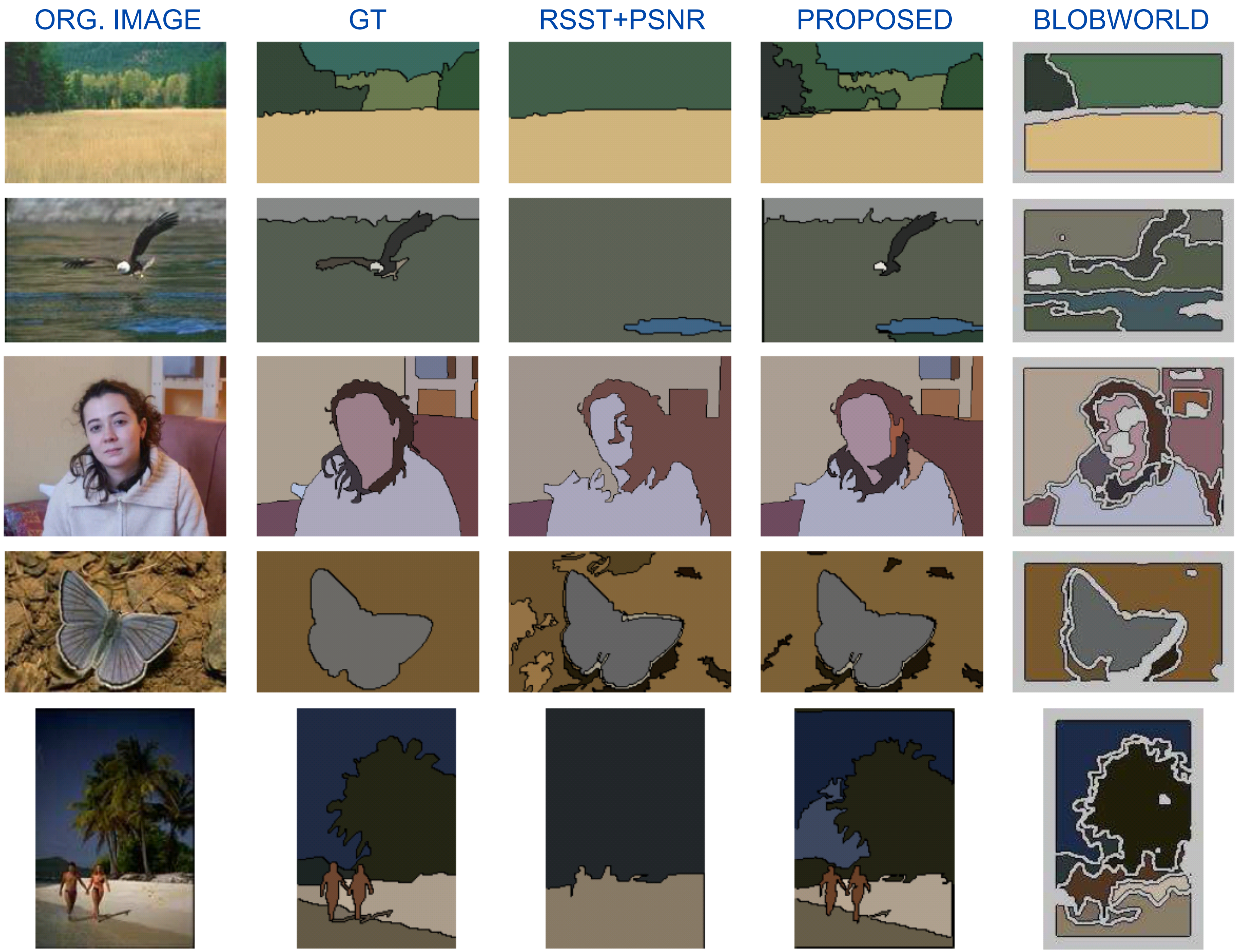
RSST with Syntactic Visual Features
Evaluation of stopping criterion

- Average spatial accuracy error

STOPPING CRITERION	Original RSST	Proposed homogeneity criterion
“Manual"	0.99	0.50
Number of regions	0.93	0.72
Merging cost	0.77	0.69
PSNR	0.84	0.82
Accum. merging cost	0.81	0.63

RSST with Syntactic Visual Features

Examples showing lower spatial segmentation error



RSST with Syntactic Visual Features

Examples showing higher spatial segmentation error

ORG. IMAGE | GT | RSST+PSNR | PROPOSED | BLOBWORLD

RSST with Syntactic Visual Features
Evaluation with Berkeley dataset – K-Space framework

- K-Space general framework for running and evaluating automatic image and video segmentation algorithms [McGuinness 2007]
  - Effortless integration of existing and forthcoming image segmentation algorithms
  - Allows focus on the development and evaluation of segmentation methods

RSST with Syntactic Visual Features
Evaluation with Berkeley dataset – K-Space framework

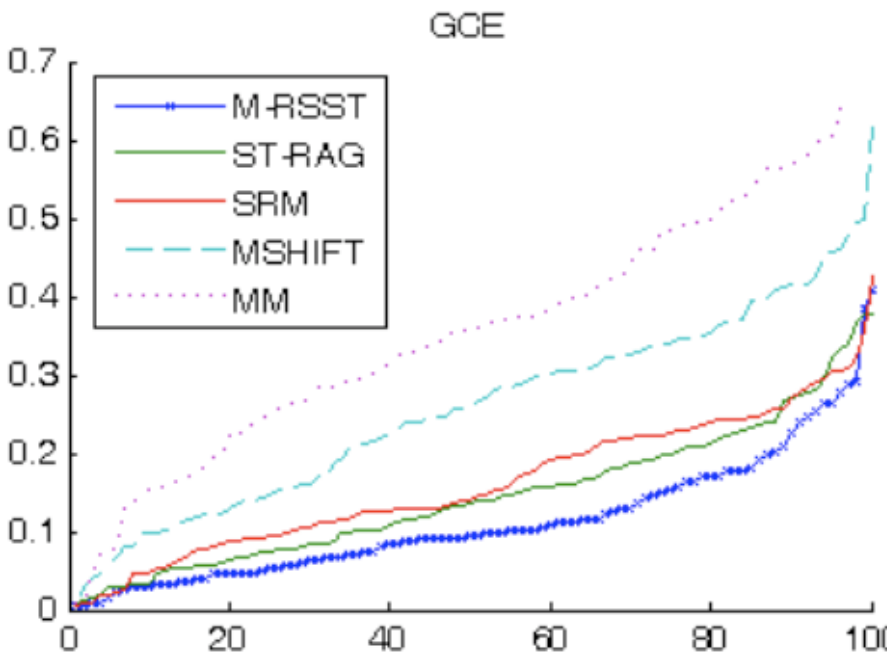
- Automatic approaches integrated with the tool:
  - MRSST: Modified RSST using Syntactic Features
  - ST-RAG: Spatio-Temporal Region Adjacency Graphs
  - MSHIFT : Optimized Mean Shift
  - SRM: Statistical Region Merging

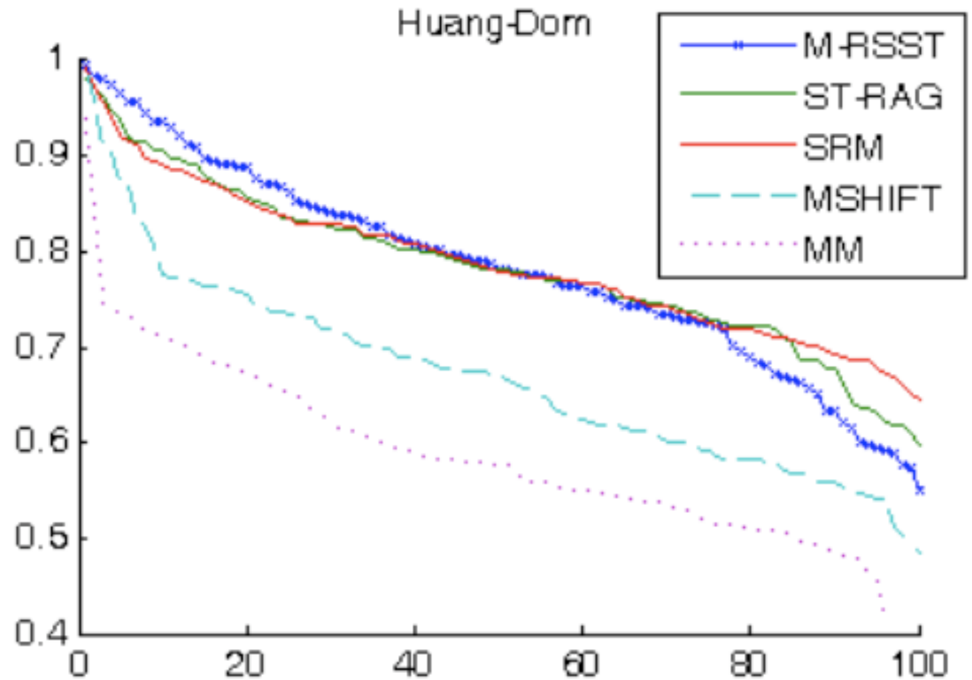
RSST with Syntactic Visual Features
Database

- This excellent dataset is comprised of 300 images
  - A test set of 100 images and a training set of 200 images
  - Each image has ≈ 11 ground truth images
- Each machine segmented image evaluated against all available ground-truths, presenting both the best-match and mean results
- A “worst-case” baseline – mismatched segmentations

RSST with Syntactic Visual Features
Comparison

- Evaluation measures:
  - Local (LCE) and Global Consistency Errors (GCE)
  - Huang-Dom measure (HD)
  - Correlation of GCE and LCE values





RSST with Syntactic Visual Features
Summary

- Works well on (most) images from different collections, i.e. identifies the most salient regions
- Syntactic features can improve the quality of the segmentation produced by region merging
- Better preservation of semantic boundaries
- It takes less than 3 seconds to segment a CIF image (352x288) on a standard PC

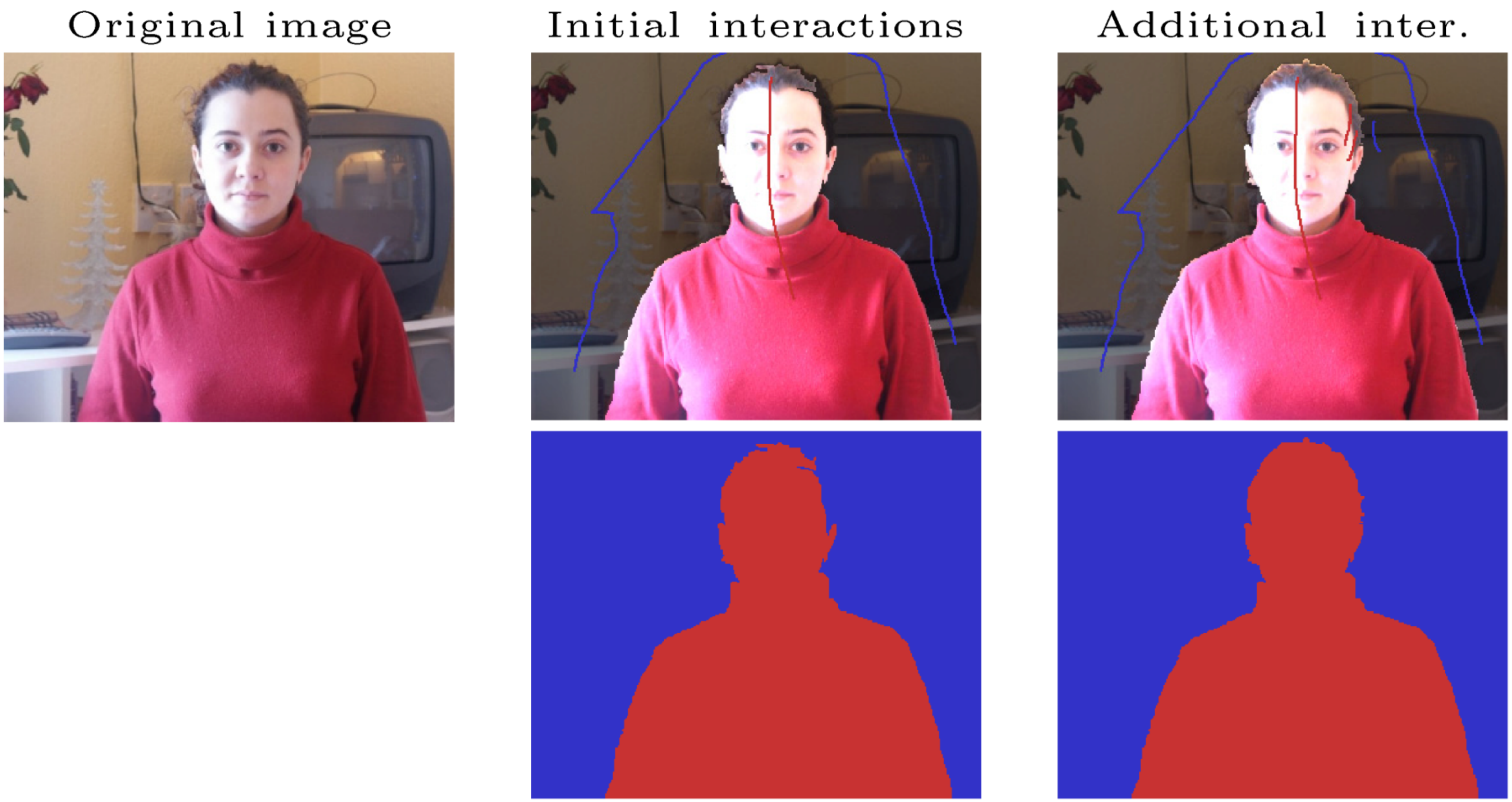
RSST with Syntactic Visual Features
Future work

- What could be done differently?
  - Use better/more assessment methods
  - Use more heterogenous test collection
- Future work:
  - Use more geometric properties, e.g. texture, co-linearity, etc.
  - Better exploration of the solution space
  - Syntactic and semantic information are not mutually exclusive
  - Investigate different techniques for the initial partitioning

Overview

- Introduction
- Review of the main categories of image segmentation approaches
- Evaluation of image segmentation
- RSST with syntactic visual features
- **Semi-automatic segmentation**
- Segmentation in K-Space participation in TRECVid'07

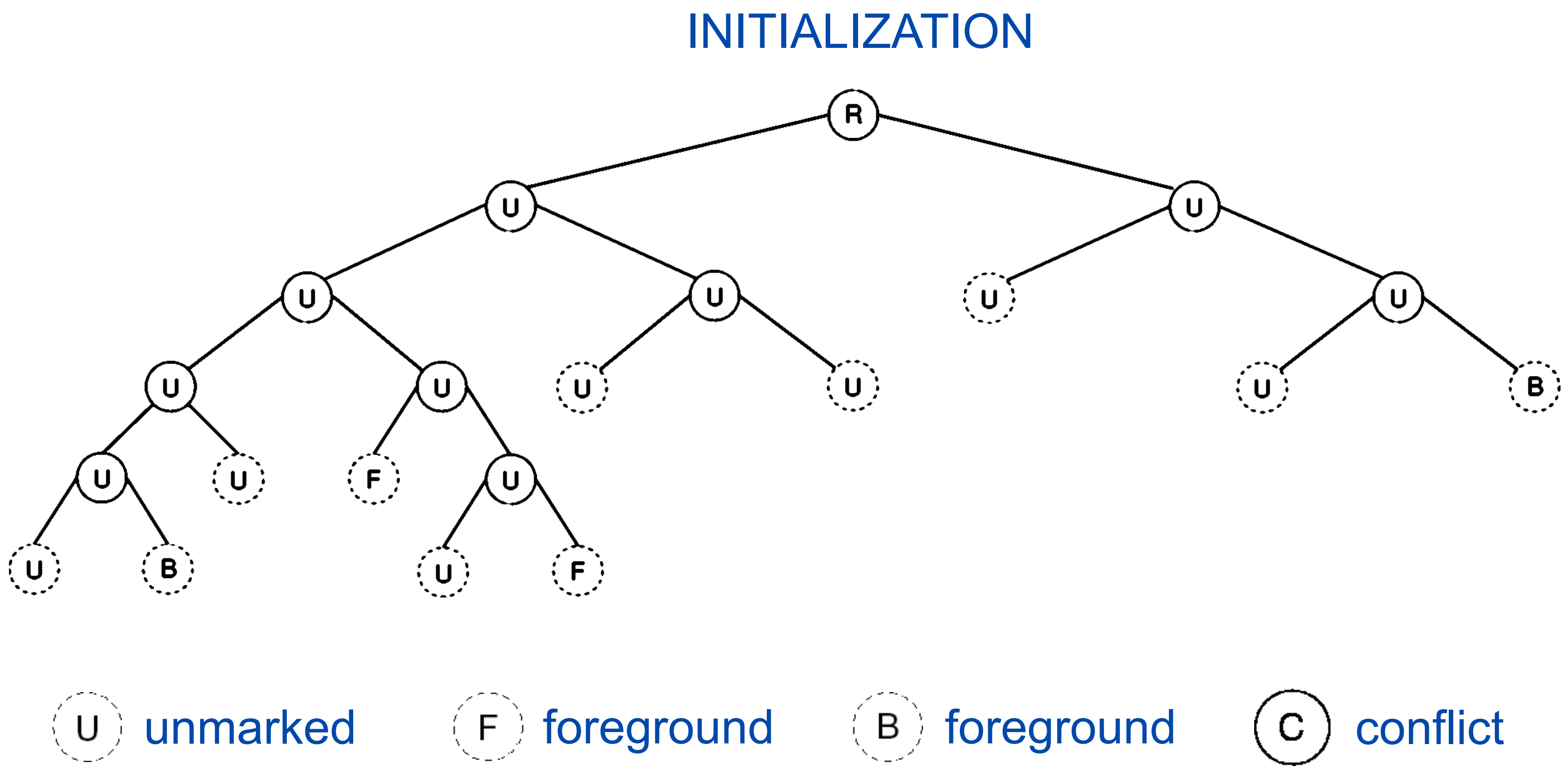
Semi-automatic Segmentation
Intuitive and effective scribble-based interactions



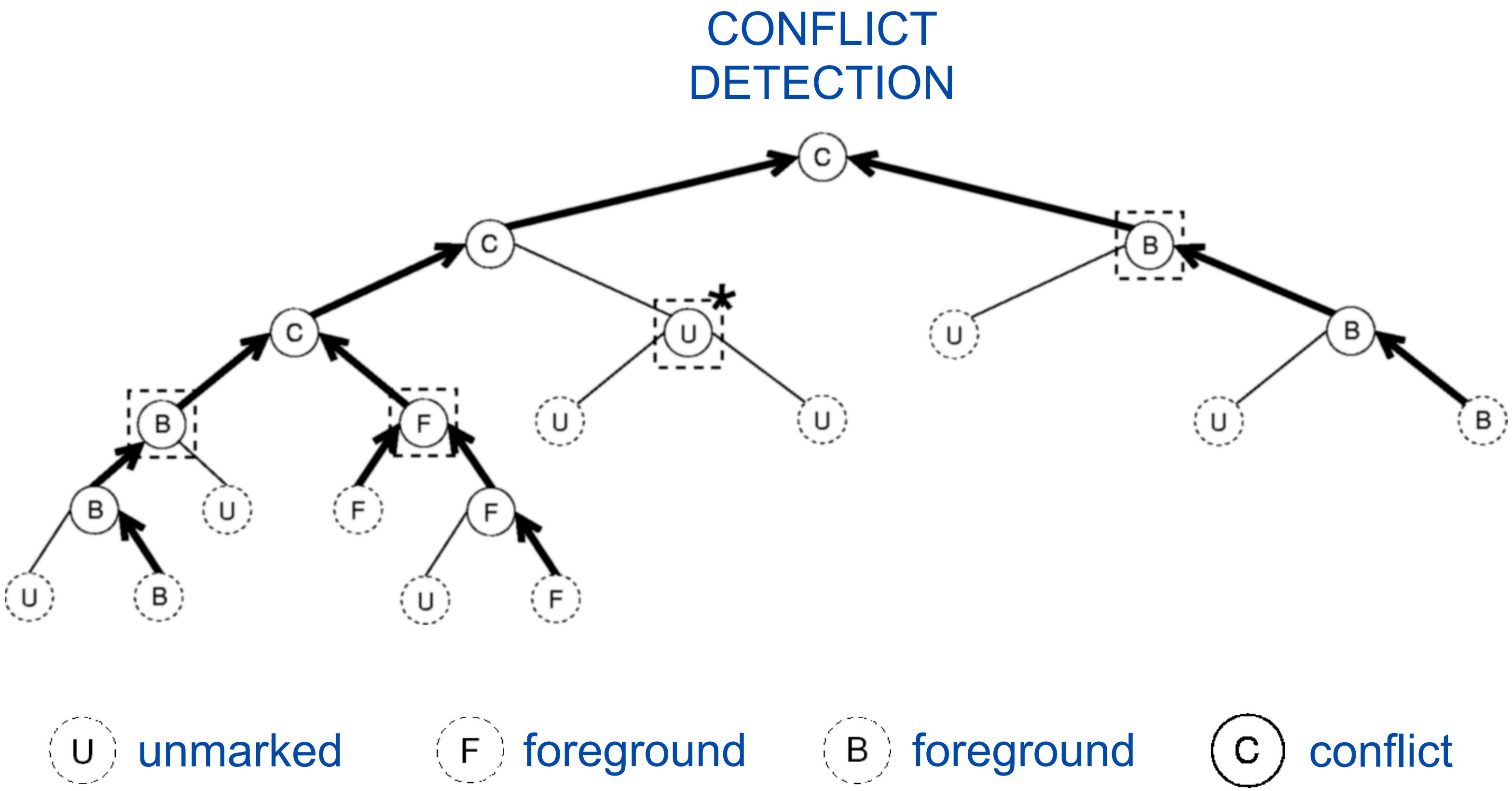
Semi-automatic Segmentation
Scribble Based Partitioning

- Automatic process uses scribbles to partition the image
- Based on scribble-based segmentation proposed in [Salembier & Garrido 2000]
- Prior to interactions image is automatically pre-segmented into regions and represented as Binary Partition Tree (BPT)
- The tree structure encodes similarities between regions pre-computed during the automatic segmentation
- BPT is used for rapid propagation of labels from scribbles(leafs) to large areas(regions) of the image
- Allows convenient integration of an automatic segmentation tool shown earlier

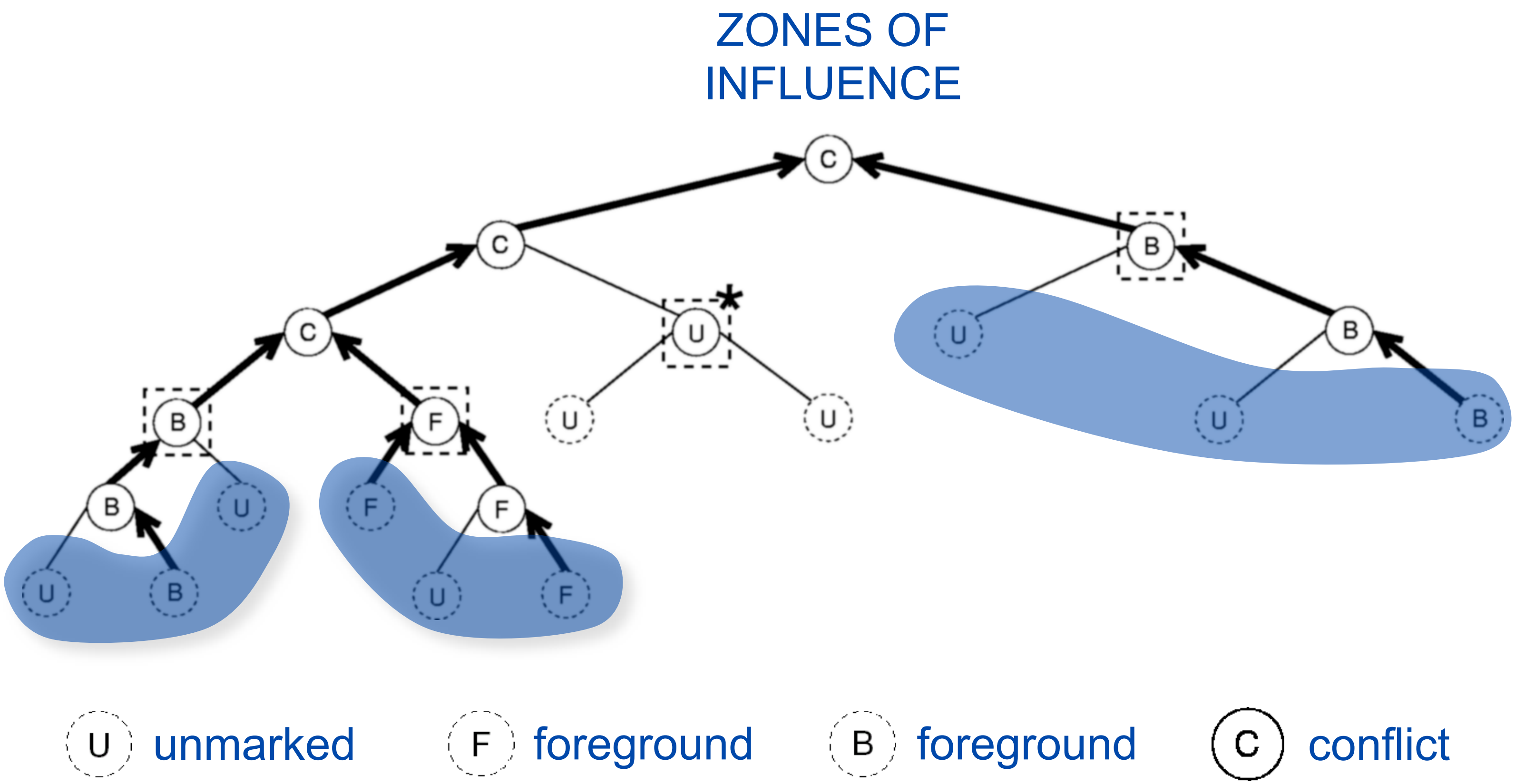
Semi-automatic Segmentation
Tree initialization



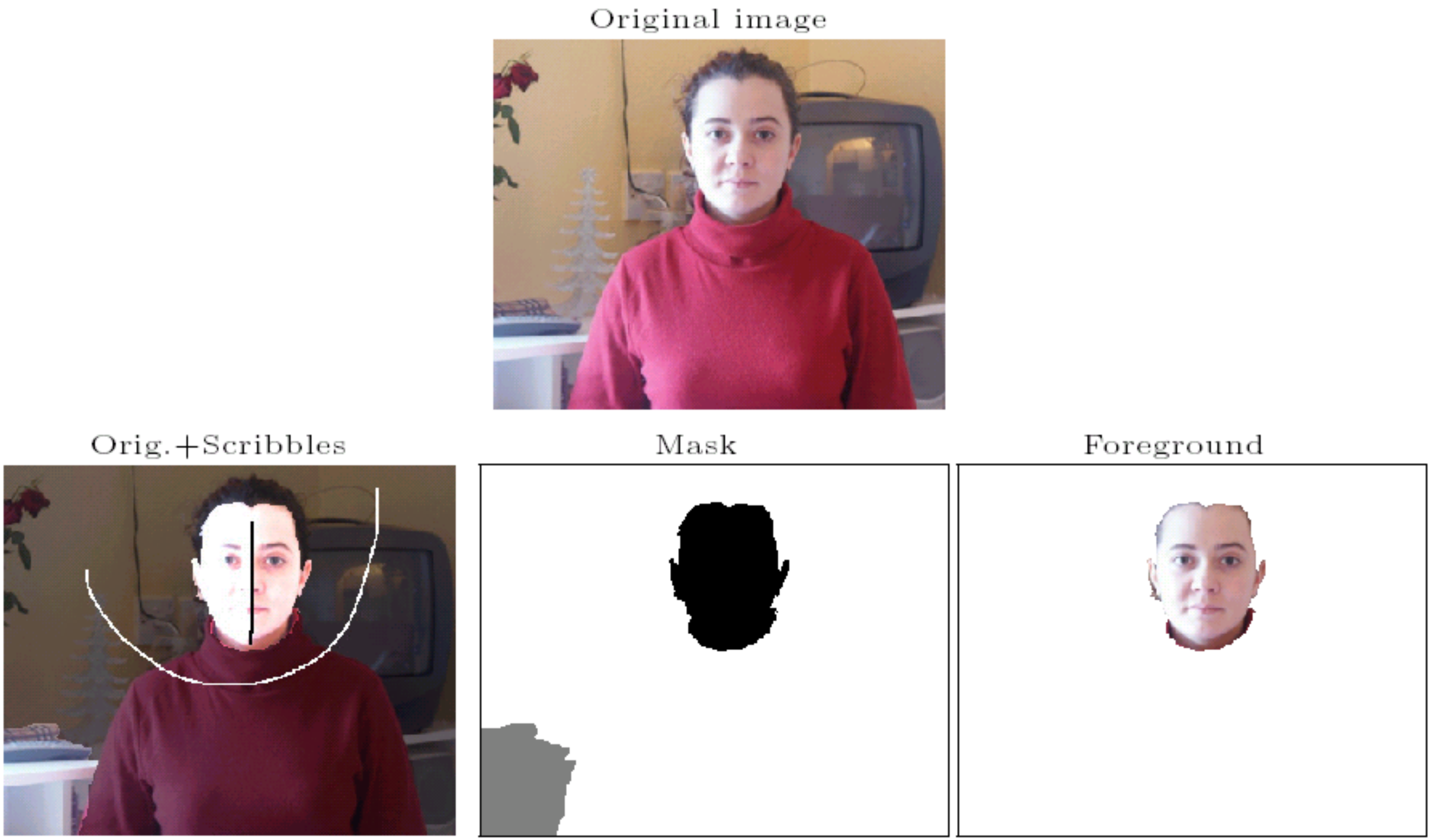
Semi-automatic Segmentation
Detection of conflicts



Semi-automatic Segmentation
Labeling entire image



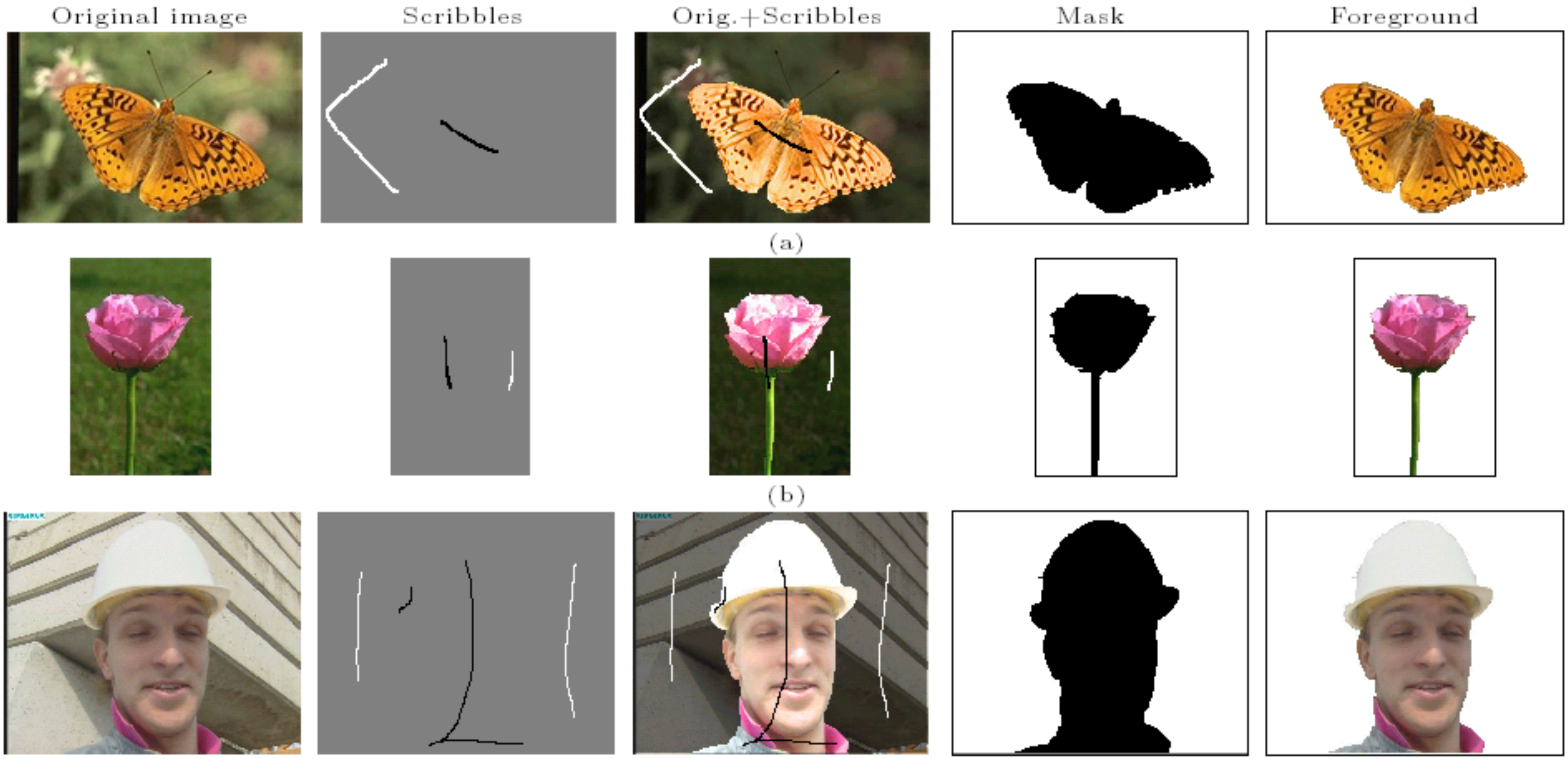
Semi-automatic Segmentation
Labeling entire image, cont.



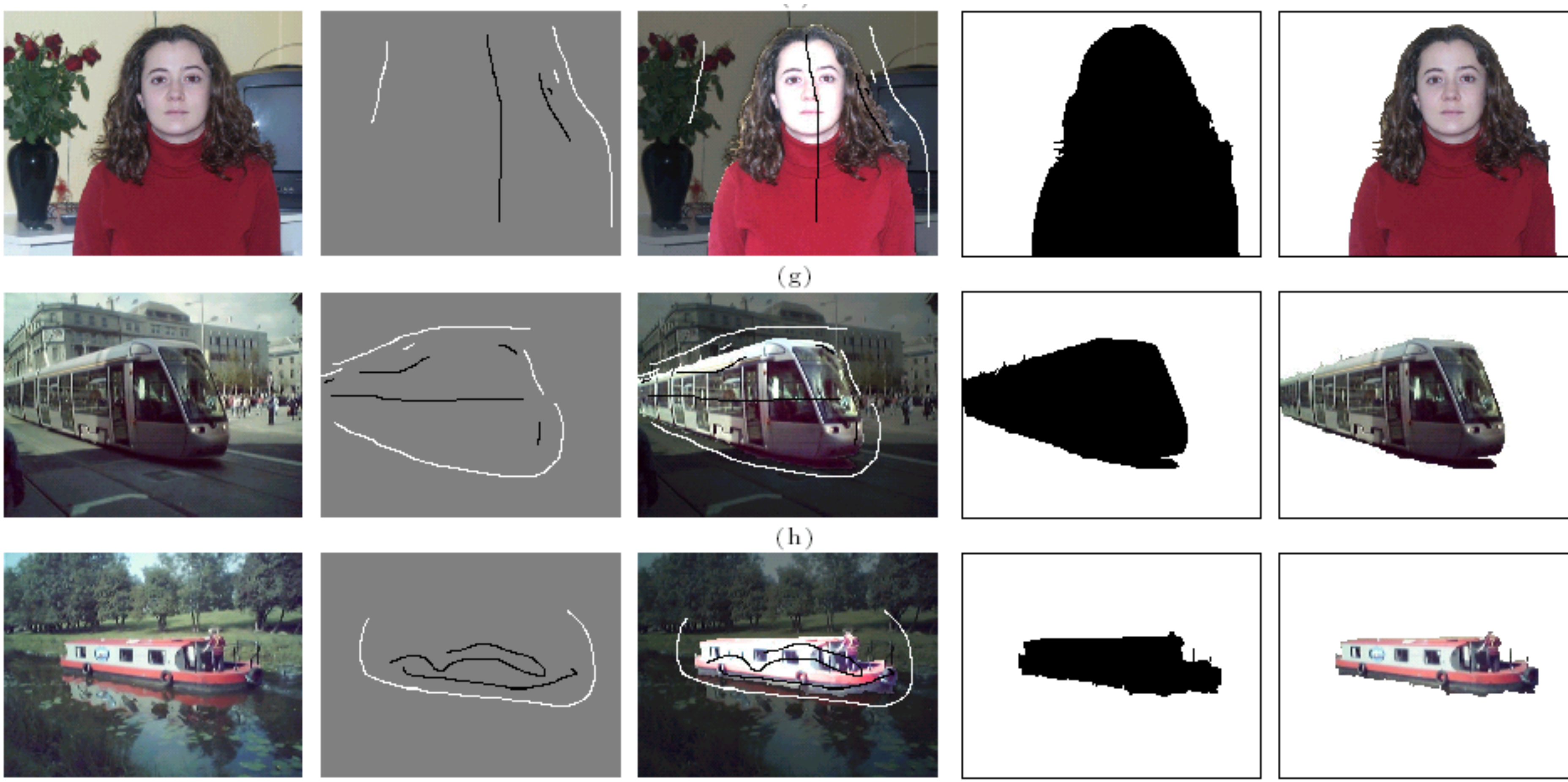
Semi-automatic Segmentation
Labeling entire image, cont.

- The filling is performed by an iterative merging
- The unlabeled nodes (regions) are iteratively assigned to the competing objects, formed by already classified regions
- Distance between an unclassified region and an object is computed as the shortest distance between the region and one of its neighbors already assigned to that object
- At each iteration, only one region, with the highest confidence, is assigned to an appropriate object
- The representation of the extended object is updated and the process continues until all regions are labeled

Semi-automatic Segmentation
Results



Semi-automatic Segmentation
Results

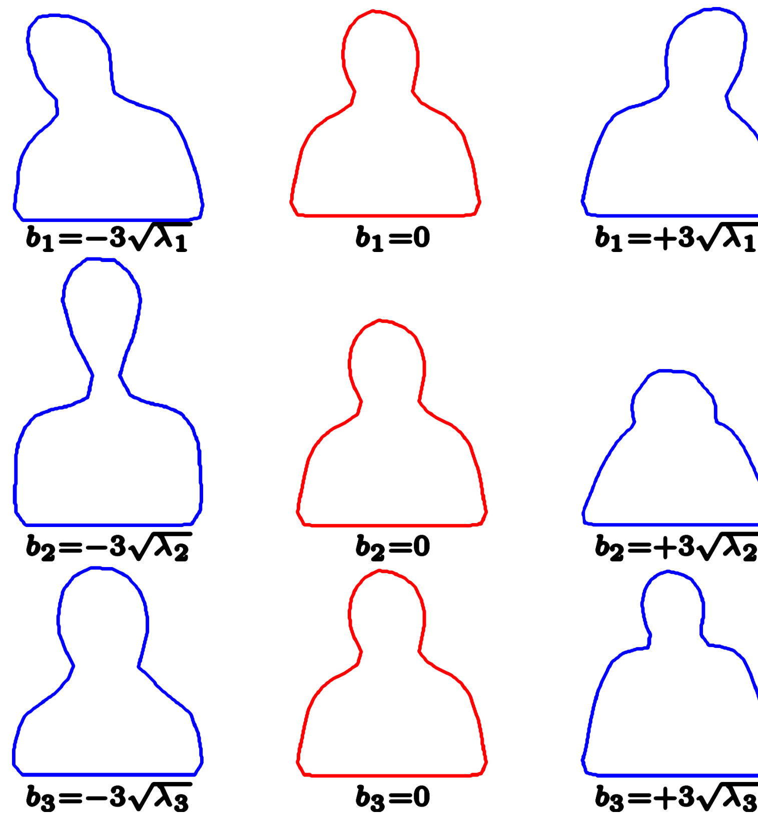


Semi-automatic Segmentation
Example application

⋮

Statistical Model

Principal Modes Of Variation



Semi-automatic Segmentation
Summary

- Pre-segmentation of a CIF (352x288) image takes under 3 seconds on a standard PC
- Instantaneous responses to interactions
- Typical time spent by the user to obtain a required segmentation is between 5-10 seconds
- Could find application in:
  - Query formulation in object-based information retrieval
  - Annotation tools
- Creation of models of objects

Overview

- Introduction
- Review of the main categories of image segmentation approaches
- Evaluation of image segmentation
- RSST with syntactic visual features
- Semi-automatic segmentation
- **Segmentation in K-Space participation in TRECVid'07**

Telefónica

Segmentation in K-Space participation in TRECVid
TRECVid'07

- TRECVid is run by NIST (National Institute of Standards and Technologies, USA) and uses a common corpus of video data for evaluation, on which a set of tasks are defined
- TRECVid High-level feature extraction task
  - Identification of shots which contain one of 36 pre-defined semantic features, e.g. 'people','snow', 'military', 'urban' etc.
- TRECVID 2007 video data: ~100 hours of news magazine, science news, news reports, documentaries, educational programming and archival video in MPEG-1

Segmentation in K-Space participation in TRECVid
K-Space participation

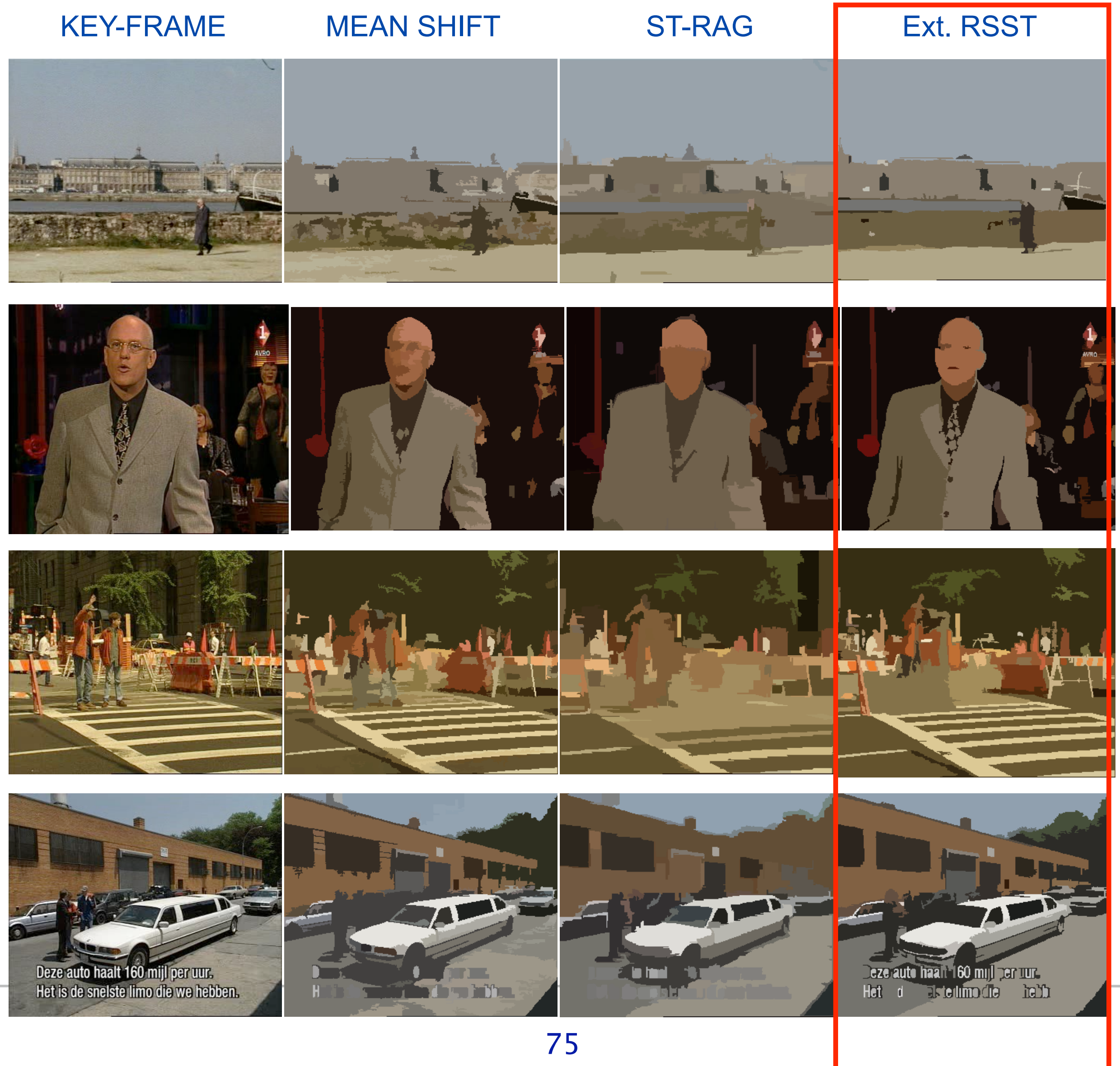
- Combined effort of nine K-Space partners
- Common set of features:
  - 7 global low-level features: color (4), edges(1), texture(2)
  - 10 region-based low-level features: color (5), edges(1), texture(1), shape(1), geometric properties (2)
  - 25 audio features
  - 2 temporal features (motion activity, camera motion)

Segmentation in K-Space participation in TRECVid
How we segmented the key-frames?

- We considered several approaches from the K-Space segmentation tool
- Accuracy of segmented regions in terms of how well they mapped to object
- Typical number of regions produced by a given algorithm
  - large regions are typically more suited to subsequent robust feature estimation
- Computational cost to execute

Segmentation in K-Space participation in TRECVid

How we segmented the key-frames?



Segmentation in K-Space participation in TRECVid
Using segmentation for high-level feature detection

- Two partners used region-based low-level features:
  - E. Spyrou, Y. Avrithis, **National Techn. University of Athens**
  - R. Benmokhtar, E. Galmar, B. Huet, **Eurecom**
- General Idea
  - The number of regions for each image may vary and there is no ordering to the regions
  - SVM classifier requires feature vectors of fixed dimensions
  - A dictionary of visual terms built by k-means clustering of all training feature vectors
  - Shot represented as “Bags of Visual Words” by mapping regions to visual words from the vocabulary
  - The signature built by counting the number of occurrences of each term in the shot

Telefónica

Segmentation in K-Space participation in TRECVid
Segmentation did not help much!

- The region-based approaches performed worse than variants based solely on MPEG-7 global features

Concept	median	KSpace 1	KSpace 2	KSpace 3	KSpace 4	KSpace 6
1	0.028	0.015	0.004	0.008	0.000	0.022
3	0.002	0.003	0.000	0.003	0.000	0.005
5	0.061	0.046	0.023	0.079	0.015	0.030
6	0.053	0.065	0.086	0.122	0.024	0.022
10	0.008	0.001	0.000	0.001	0.001	0.005
12	0.003	0.009	0.003	0.028	0.006	0.021
17	0.167	0.151	0.151	0.149	0.021	0.157
23	0.003	0.002	0.002	0.012	0.005	0.006
24	0.005	0.012	0.007	0.008	0.002	0.001
26	0.074	0.077	0.005	0.074	0.047	0.108
27	0.051	0.035	0.035	0.021	0.003	0.019
28	0.000	0.000	0.000	0.000	0.000	0.000
29	0.022	0.045	0.002	0.009	0.006	0.013
30	0.082	0.076	0.035	0.117	0.010	0.059
32	0.026	0.018	0.018	0.036	0.016	0.026
33	0.083	0.020	0.020	0.071	0.071	0.033
35	0.028	0.002	0.012	0.015	0.012	0.013
36	0.005	0.004	0.001	0.006	0.007	0.001
38	0.035	0.015	0.002	0.018	0.002	0.071
39	0.017	0.006	0.006	0.021	0.004	0.014
average	0.039	0.030	0.021	0.040	0.013	0.031

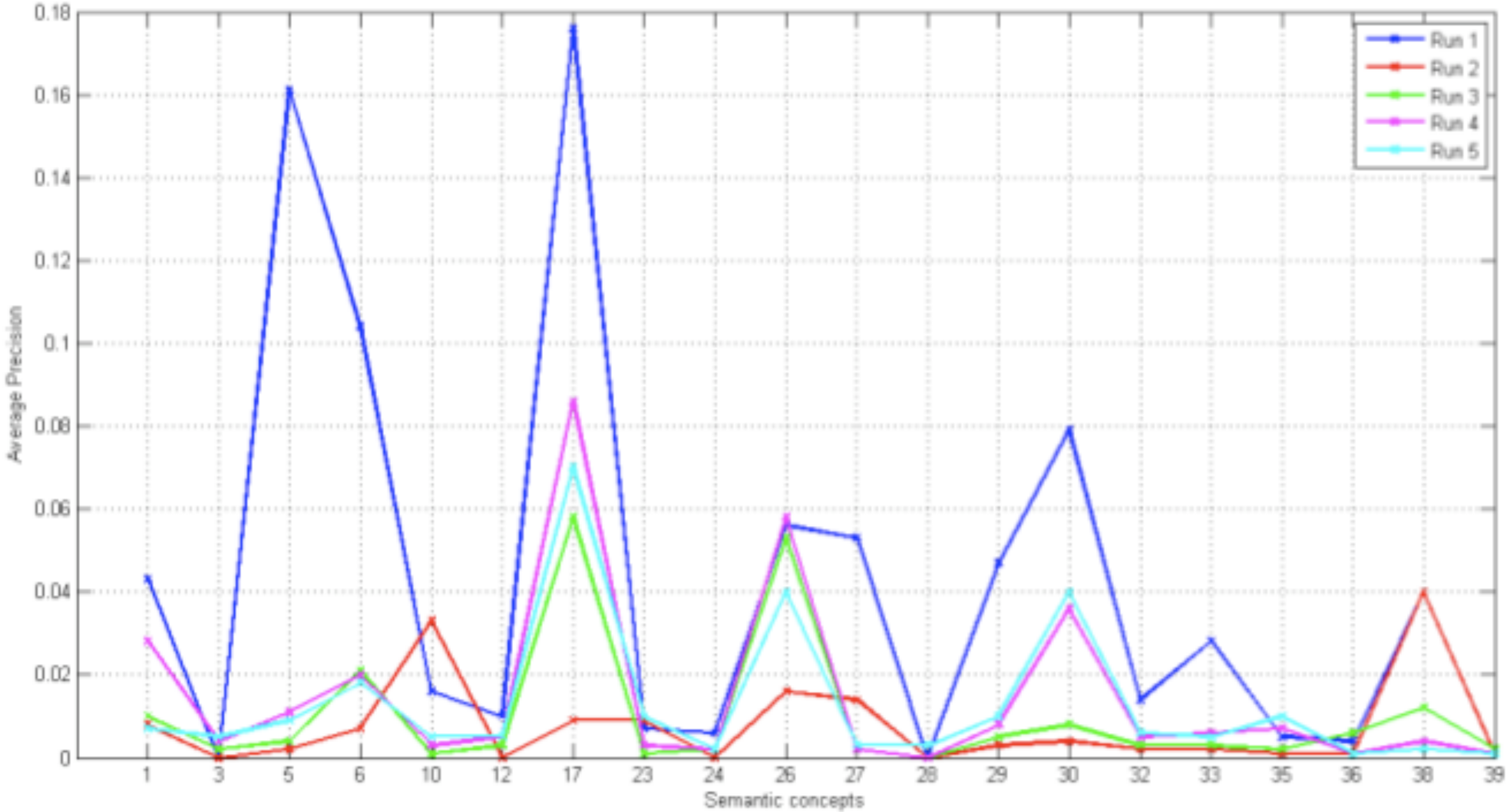
GLOBAL-BASED

REGION-BASED

MULTIMODAL

Segmentation in K-Space participation in TRECVid
Segmentation did not help much!

- Results from [Benmokhtar et al. 2007]:



Segmentation in K-Space participation in TRECVid
Why is it difficult to use segmentation for HLE

- The “Bag of Regions” approach might suffer from the lack of spatial information
- The best parameters are still to be found
  - E.g. not sufficient number of “visual words”
- Poor repeatability of the automatic segmentation algorithms
- Some MPEG-7 features are not adequate to represent regions
- But there is a hope!
  - The winners of TRECVid 07 used segmentation [Tsinghua 2007]
  - Fusion with other features

References

- [Wollborn'98] M. Wollborn and R. Mech, "Refined procedure for objective evaluation of video object generation algorithms," MPEG Meeting, Tokyo, Japan, Mar. 1998.
- [Villegas et al. 99] P. Villegas, X. Marichal, and A. Salcedo, "Objective evaluation of segmentation masks in video sequences," WIAMIS'99.
- [Correia et al. 2000] P. Correia and F. Pereira, "Objective evaluation of relative segmentation quality," ICIP'00.
- [Gelasca et al. 2004] E. D. Gelasca, T. Ebrahimi, M. C. Q. Farias, and S. K. Mitra, "Impact of topology changes in video segmentation evaluation," WIAMIS'04.
- [Mezaris et al. 2003] V. Mezaris, I. Kompatsiaris, and M. Strintzis, "Still image objective segmentation evaluation using ground truth," COST 276 Workshop 2003.
- [Martin et al. 2001] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics", ICCV 2001.
- [Huang et al. 1995] Qian Huang and B. Dom, "Quantitative methods of evaluating image segmentation", ICIP 1995.
- [Ge et al. 2006] Ge, F., Wang, S., Liu, T., "Image segmentation evaluation from the perspective of salient object extraction", CVPR 2006.

References

- [Chabrier et al. 2006] S. Chabrier, B. Emile, C. Rosenberger, and H. Laurent, “Unsupervised performance evaluation of image segmentation,” EURASIP Journal on Applied Signal. Processing, vol. 2006.
- [Correia et al. 2006] P. Correia and F. Pereira, “Video object relevance metrics for overall segmentation quality evaluation,” EURASIP Journal on Applied Signal. Processing, vol. 2006.
- [Jiang et al. 2006] X. Jiang, C. Marti, C. Irniger, and H. Bunke, “Distance measures for image segmentation evaluation,” EURASIP Journal on Applied Signal. Processing, vol. 2006.
- [Piroddi et al. 2006] R. Piroddi and T. Vlachos, “A method for single-stimulus quality assessment of segmented video,” EURASIP Journal on Applied Signal. Processing, vol. 2006.
- [Benmokhtar 2007] R. Benmokhtar, E. Galmar, B. Huet, “Eurecom at TRECVid 2007: Extraction of High-level Features”, TREC Video Retrieval Evaluation Online Proceedings.
- [Tsinghua 2007] Tsinghua University-Intel China, “THU and ICRC at TRECVID 2007”
- [Morris et al. 1986] O. Morris, M. Lee, and A. Constantinides, “Graph theory for image analysis: an approach based on the shortest spanning tree,” in IEE Proceedings, vol. 133, Apr. 1986
- [Frauqueur & Boujemaa 2004] J. Fauqueur and N. Boujemaa, “Region-based image retrieval: Fast coarse segmentation and fine color description,” Journal of Visual Languages and Computing, special issue on Visual Information Systems, vol. 15, pp. 69-95, 2004.
- [Bennstrom & Casas 2004] C. F. Bennstrom and J. R. Casas, “Binary-partition-tree creation using a quasi-inclusion criterion,” IV’04, IEEE.

References

- [Dempster 1967] A. Dempster, “Upper and lower probabilities induced by multivalued mapping,” Annals of Mathematical Statistics, vol. 38, pp. 325-339, 1967.
- [Shafer 1976] G. Shafer, “A mathematical theory of evidence,” Princetown Univ. Press, Princetown New Jersey, 1976.
- [Rosin 2001] P. Rosin, “Unimodal thresholding,” Pattern Recognition, vol. 34, no. 11, pp. 2083-2096, Nov. 2001.
- [McGuinness 2007] K. McGuinness, G. Keenan, T. Adamek and N. O’Connor, “Image Segmentation Evaluation Using An Integrated Region Based Segmentation Framework”
- [Salembier & Garrido 2000] P. Salembier and L. Garrido, “Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval,” IEEE Trans. on Image Processing, vol. 9, no. 4, pp. 561-576, Apr. 2000.
- [Mezaris et al. 2004] V. Mezaris, I.Kompatsiaris, and M.G.Strintzis, “Still image segmentation tools for object-based multimedia applications,” Int’l Journal of Pattern Recognition and Artificial Intell., vol. 18, no. 4, pp. 701-725, June 2004.
- [Carson et al. 2002] C. Carson, S. Belongie, H. Greenspan, and J. Malik, “Blobworld: Color- and texture-based image segmentation using EM and its application to image querying and classification,” IEEE Trans. Pattern Anal. and Machine Intell., vol. 24, no. 8, pp. 1026-1037, Aug. 2002.
- [Comaniciu & Meer 2002] D. Comaniciu and P. Meer, “Mean shift: A robust approach toward feature space analysis,” IEEE Trans. Pattern Anal. and Machine Intell., vol. 24, no. 5, 2002.

References

- [Vincent & Soille 1991] L. Vincent and P. Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations," IEEE Trans. Pattern Anal. and Machine Intell., vol. 13, no. 6, pp. 583-598, June 1991.
- [Alatan et al. 1998] A. Alatan, L. Onural, M. Wollborn, R. Mech, E. Tuncel, and T. Sikora, "Image sequence analysis for emerging interactive multimedia services - the European COST 211 Framework," IEEE Trans. Circuits Syst. Video Technol., vol. 8, no. 7, pp. 802-813, Nov. 1998.
- [Shi & Malik 2000] J. Shi and J. Malik, "Normalized cuts and image segmentation," IEEE Trans. Pattern Anal. and Machine Intell., vol. 22, no. 8, pp. 888-905, Aug. 2000.
- [Cootes & Taylor 1992] T. F. Cootes and C. J. Taylor, "Active shape models - smart snakes," in Proc. British Machine Vision Conf. (BMVC'92), Springer Verlag, 1992, p. 266.
- [Cootes & Taylor 2004] T. Cootes and C.J.Taylor, "Statistical models of appearance for computer vision," University of Manchester, http://www.isbe.man.ac.uk/˜bim, Tech. Rep., Mar. 2004.
- [Adams & Bischof 1994] R. Adams and L. Bischof, "Seeded region growing," IEEE Trans. Pattern Anal. and Machine Intell., vol. 16, no. 6, pp. 641-647, June 1994.
- [Boykov & Jolly 2001] Y. Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images," in Proc. Int'l Conf. on Computer Vision, Vancouver, Canada, vol. I, July 2001.
- [Kass & Witkin 1988] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," Int'l Journal of Computer Vision, vol. 1, pp. 312-331, 1988.

References

- [Caselles et al. 1997] V. Caselles, R. Kimmel, and G. Sapiro, “Geodesic active contours,” Int’l Journal of Computer Vision (IJCV), vol. 22, no. 1, pp. 61-79, 1997.
- [Zhu & Yuille 1996] S. C. Zhu and A. Yuille, “Region competition: Unifying snakes, region growing and bayes/MDL for multiband image segmentation,” IEEE Trans. Pattern Anal. and Machine Intell., vol. 18, no. 9, pp. 884-900, 1996.

Telefónica