PATCH-BASED IMAGE CLASSIFICATION AND RETRIEVAL

Giuseppe Passino
(Supervisors: Ebroul Izquierdo, Ioannis Patras)
School of Electronic Engineering and Computer Science
Queen Mary, University of London, London, UK
giuseppe.passino@elec.qmul.ac.uk

1. Content-based Image Analysis

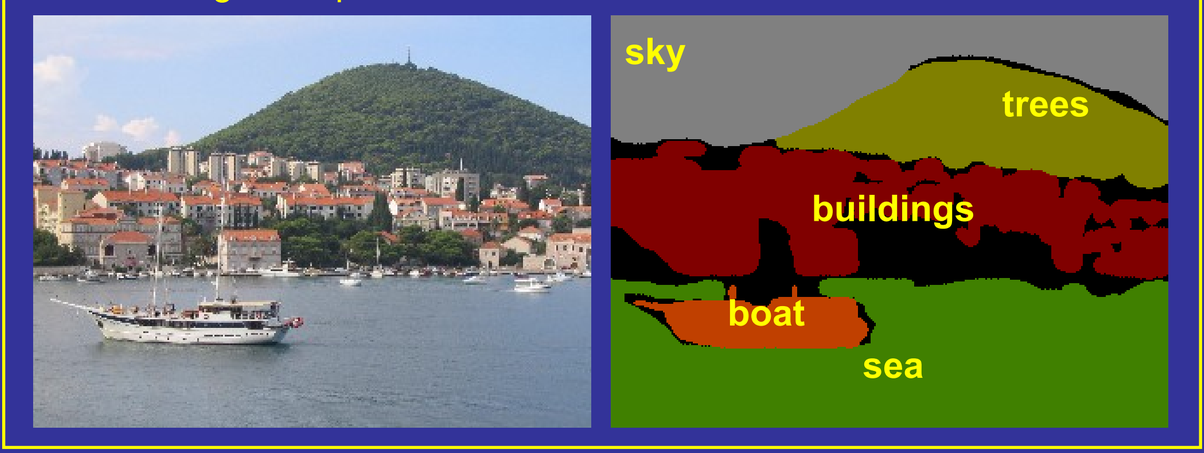
In a word of automated multimedia content, the challenge is to provide means of extract image semantics

Applications: image retrieval, Human-Computer Interaction,

e.g.: *sailboat at sunset*

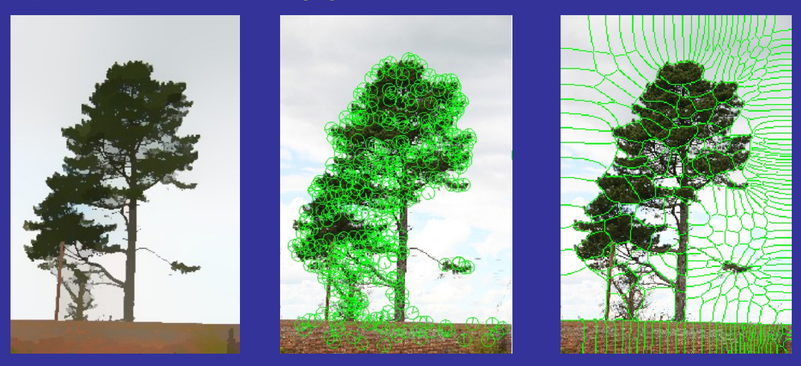
2. Problem: Understanding Image Content

My aim is to devise a system to automatically interpret different parts of an image and put them in context



3. Solution: Patch-based Image Analysis

The strategy is to extract parts from images, and associate categories to them, according to **(1)** appearance, and **(2)** context.



cartooning | salient points | spectral clustering

Context is considered linking parts in a graph and using a discriminative probabilistic model to infer parts categories

Conditional Random Fields: $p(y|X;\theta) = \frac{\exp \Psi(y,X,\theta)}{Z(X,\theta)}$

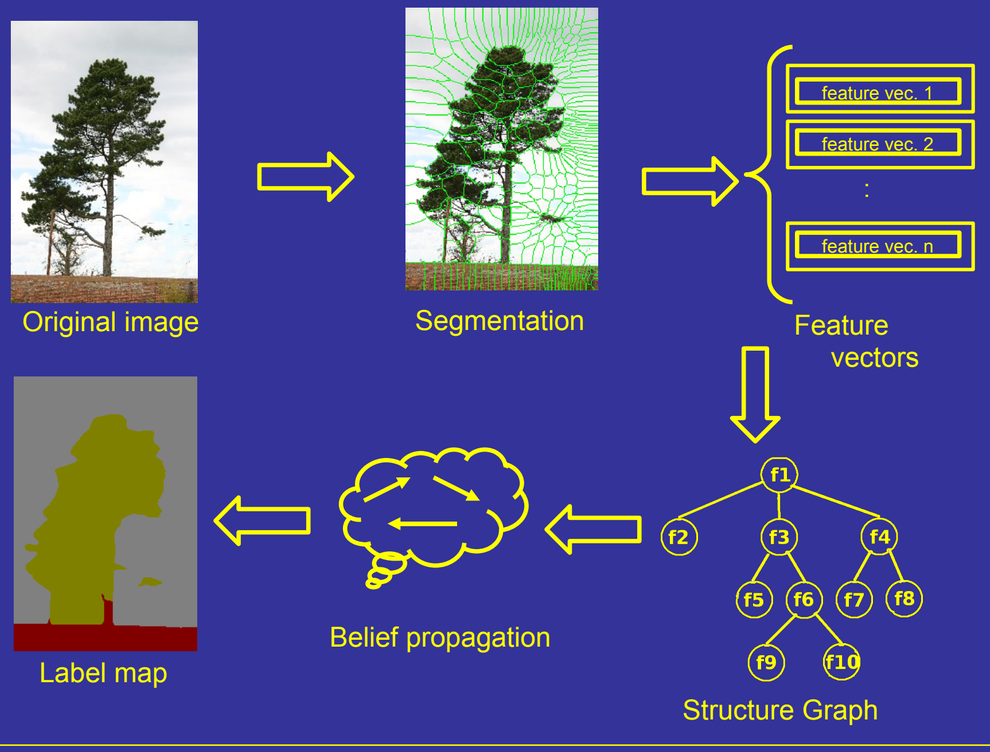
$$\Psi(y,X,\theta) = \sum_{v\in V}\sum_{k\in K_1}\theta_k\phi_k^1(y_v,X) + \sum_{(i,j)\in E}\sum_{k\in K_2}\theta_k\phi_k^2(y_i,y_j,X)$$

appearance | relationships

CRF is an unconditional graphical model, is which inference can be performed via Belief Propagation

4. Contributions

4.1 System Model

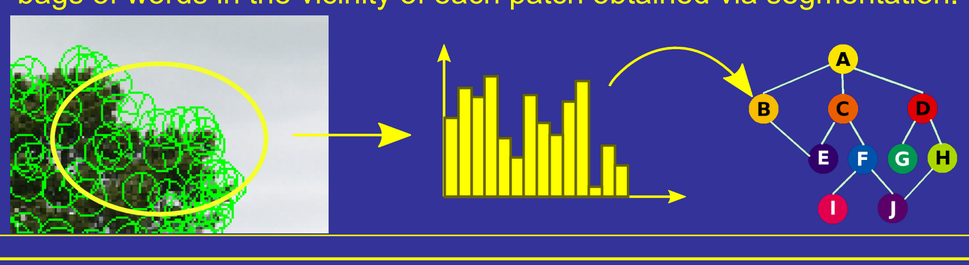


4.2 Aspect-coherence for Tree Structure

The structure connecting parts has to be a tree to perform exact inference exactly. I connect parts that have similar aspect, thus highly correlated.

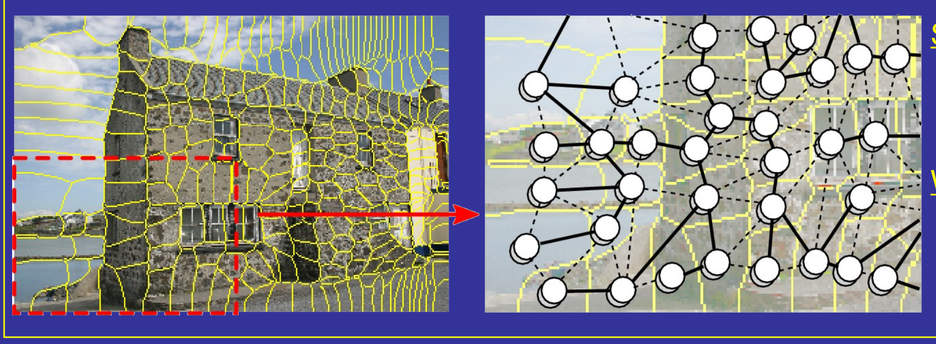
4.3 Salient Points for Context Awareness

I extract salient points, associate them with "visual words" and I consider bags of words in the vicinity of each patch obtained via segmentation.



4.4 Weak Neighbours for Dense Local Area Coverage

The patches left unconnected in the tree are pre-classified as considered in inference.



Strong Neighbours: considered in the aspect-coherent spanning tree
Weak Neighbours: pre-classified and considered as fixed

5. System Performance

System	building	grass	tree	sky	cow	plane	face	car	cycle	avg
no conn.	57.0%	93.1%	67.8%	53.0%	94.2%	37.2%	60.2%	51.5%	60.3%	73.4%
CRF_{AC}	63.0%	94.2%	68.9%	84.4%	93.7%	75.8%	92.9%	76.4%	86.5%	82.9%
CRF_{WN}	72.0%	**94.8%**	71.6%	77.3%	95.3%	**80.3%**	92.1%	82.3%	89.9%	85.2%
CRF_{WWH}	**76.7%**	94.6%	71.4%	**86.3%**	95.0%	73.1%	99.3%	73.2%	**93.7%**	**86.2%**
Lit.	73.6%	91.1%	**82.1%**	73.6%	**95.7%**	78.3%	**89.5%**	**84.5%**	81.4%	84.9%

no conn.: no connections considered
CRF_AC: Aspect-Coherence tree
CRF_WN: Weak Neighbours for dense local coverage
CRF_WWH: Salient Points for Context Awareness
LIT: J. Verbeek and B. Triggs, "Scene segmentation learned from partially labeled images," in *NIPS*, 2007

6. Results examples