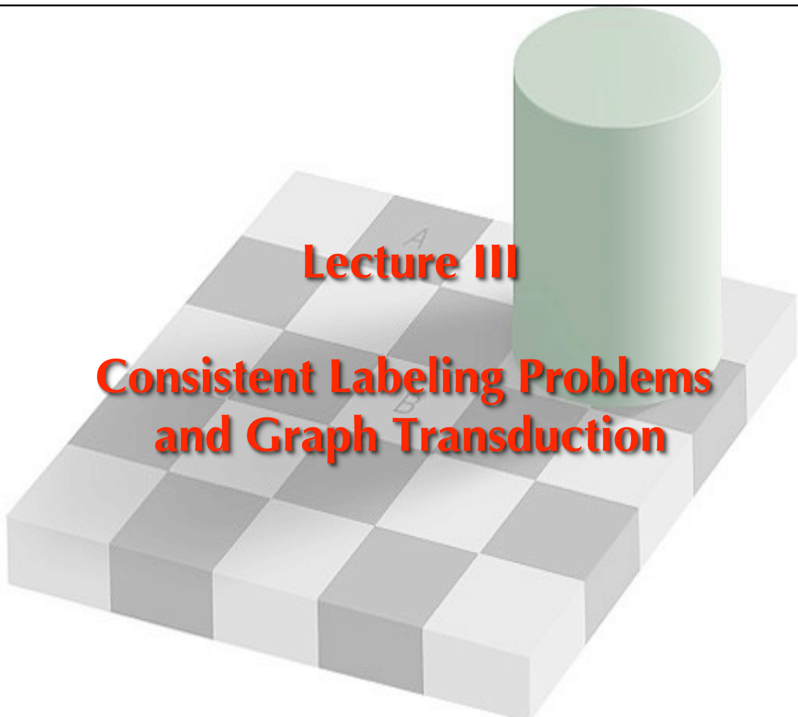
Game Theory in Graph-Based Computer Vision and Pattern Recognition

Marcello Pelillo

Ca' Foscari University, Venice

pelillo@dsi.unive.it

Summer School on Graphs in Computer Graphics, Image and Signal Analysis
Bornholm, Denmark, August 2011



Lecture III

Consistent Labeling Problems and Graph Transduction

Succinct Games

Describing a game in normal form entails listing all payoffs for all players and strategy combinations. In a game with n players, each facing m pure strategies, one need to store nm^n numbers!

A **succinct game** (or a **succinctly representable game**) is a game which may be represented in a size much smaller than its normal-form representation.

Examples.

Sparse games. Most of the payoffs are zero.

Graphical games. The payoffs of each player depends on the actions of very few (at most d) other players. The number of payoffs needed to describe this game is nm^{d+1}.

Symmetric games. All players are identical, so in evaluating the payoff of a combination of strategies, all that matters is how many of the n players play each of the s strategies.

Polymatrix Games

A **polymatrix game** (a.k.a. **multimatrix game**) is a non-cooperative game in which the relative influence of the selection of a pure strategy by any one player on the payoff to any other player is always the same, regardless of what the rest of the players do.

Formally:

- ✓ There are n players each of whom can use m pure strategies
- ✓ For each pair (i,j) of players there is an $m \times m$ payoff matrix A^{ij}
- ✓ The payoff of player i for the strategy combination $s_1,\ldots,s_n$ is given by

$$u_i(s_1,\cdots,s_n) = \sum_{j \neq i} A^{ij}_{s_i s_j}$$

The number of payoff values required to represent such a game is $O(n^2m^2)$.

The problem of finding a Nash equilibrium in a polymatrix game is PPAD-complete.

Context helps …

12
A 13 C
14

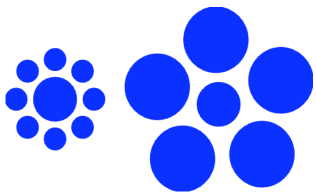
c → cat
→ circus

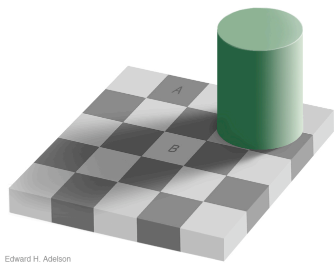
i → sin
→ fine

e → red
→ read

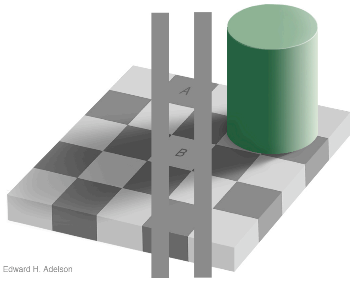
festival
graphics

… but can also deceive!





Edward H. Adelson



Edward H. Adelson

What do you see?

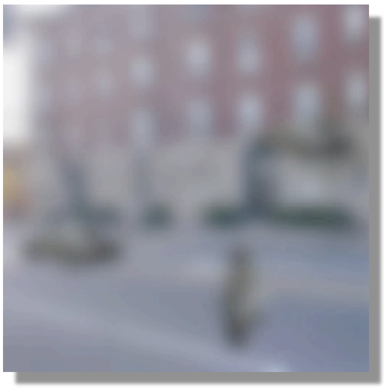
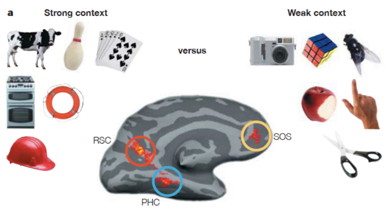


Figure 2. The strength of context. The visual system makes assumptions regarding object identities according to their size and location in the scene. In this picture, observers describe the scene as containing a car and pedestrian in the street. However, the pedestrian is in fact the same shape as the car, except for a 90° rotation. The atypicality of this orientation for a car within the context defined by the street scene causes the car to be recognized as a pedestrian.

From: A. Oliva and A. Torralba, “The role of context in object recognition”, *Trends in Cognitive Sciences*, 2007.

Context and the Brain



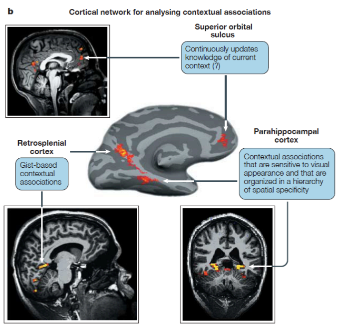


Figure 3 | **Cortical areas involved in processing context. a** | A functional magnetic resonance imaging (fMRI) statistical activation map representing the difference between perceiving objects that are strongly associated with a specific context and perceiving objects that are not associated with a unique context. This is a medial view of the left hemisphere, shown using a precise computer reconstruction where the sulci have been exposed by ‘inflation’. The parahippocampal cortex (PHC) is circled in blue; the retrosplenial cortex (RSC) is circled in red; the superior orbital sulcus (SOS) is circled in yellow. Note that in all experimental conditions, subjects viewed similar looking colour photographs of meaningful, everyday common objects that were equally recognizable. Consequently, activation due to low-level processes was presumably subtracted out, and the differential activation map shown here represents only processes that are related to the level of contextual association. **b** | The cortical network for contextual associations among visual objects, suggested on the basis of existing evidence. Other types of context might involve additional regions (for example, hippocampus for navigation[125] and Broca’s area for language-related context). Modified, with permission, from REF. 12 © (2003) Elsevier Science.

From: M. Bar, “Visual objects in context”, *Nature Reviews Neuroscience*, August 2004.

The (Consistent) Labeling Problem

A **labeling problem** involves:

- ✓ A set of n **objects** $B = \{b_1,\ldots,b_n\}$
- ✓ A set of m **labels** $\Lambda = \{1,\ldots,m\}$

The goal is to label each object of B with a label of Λ.

To this end, two sources of information are exploited:

- ✓ Local measurements which capture the salient features of each object viewed in isolation
- ✓ Contextual information, expressed in terms of a real-valued $n^2 \times m^2$ matrix of **compatibility coefficients** $R = \{r_{ij}(\lambda, \mu)\}$.

The coefficient $r_{ij}(\lambda, \mu)$ measures the strenght of compatibility between the two hypotheses: "b_i is labeled λ" and "b_j is labeled μ".

Relaxation Labeling Processes

The initial local measurements are assumed to provide, for each object $b_i \in B$, an m-dimensional (probability) vector:

$$p_i^{(0)} = \left(p_i^{(0)}(1), \cdots, p_i^{(0)}(m)\right)^T$$

with $p_i^{(0)}(\lambda) \geq 0$ and $\sum_\lambda p_i^{(0)}(\lambda) = 1$. Each $p_i^{(0)}(\lambda)$ represents the initial, non-contextual degree of confidence in the hypothesis "b_i is labeled λ".

By concatenating vectors $p_1^{(0)},\ldots,p_n^{(0)}$ one obtains an (initial) **weighted labeling assignment** $p^{(0)} \in \Re^{nm}$.

The space of weighted labeling assignments is

$$\mathrm{IK} = \underbrace{\Delta \times \ldots \times \Delta}_{m \text{ times}}$$

where each Δ is the standard simplex of $\Re^n$. Vertices of IK represent unambiguous labeling assignments

A **relaxation labeling process** takes the initial labeling assignment $p^{(0)}$ as input and iteratively updates it taking into account the compatibility model R.

Relaxation Labeling Processes

In a now classic 1976 paper, Rosenfeld, Hummel, and Zucker introduced heuristically the following update rule (assuming a non-negative compatibility matrix):

$$p_i^{(t+1)}(\lambda) = \frac{p_i^{(t)}(\lambda)q_i^{(t)}(\lambda)}{\sum_\mu p_i^{(t)}(\mu)q_i^{(t)}(\mu)}$$

where

$$q_i^{(t)}(\lambda) = \sum_j \sum_\mu r_{ij}(\lambda,\mu)p_i^{(t)}(\mu)$$

quantifies the support that context gives at time t to the hypothesis "b_i is labeled with label λ".

See (Pelillo, 1997) for a rigorous derivation of this rule in the context of a formal theory of consistency.

Applications

Since their introduction in the mid-1970's relaxation labeling algorithms have found applications in virtually all problems in computer vision and pattern recognition:

- ✓ Edge and curve detection and enhancement
- ✓ Region-based segmentation
- ✓ Stereo matching
- ✓ Shape and object recognition
- ✓ Grouping and perceptual organization
- ✓ Graph matching
- ✓ Handwriting interpretation
- ✓ …

Further, intriguing similarities exist between relaxation labeling processes and certain mechanisms in the early stages of biological visual systems (see Zucker, Dobbins and Iverson, 1989, for physiological and anatomical evidence).

Hummel and Zucker's Consistency

In 1983, Bob Hummel and Steve Zucker developed an elegant theory of consistency in labeling problem.

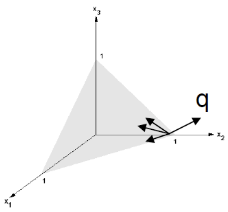
By analogy with the unambiguous case, which is easily understood, they define a weighted labeling assignment $p \in \mathrm{IK}$ **consistent** if:

$$\sum_{\lambda} p_i(\lambda) q_i(\lambda) \geq \sum_{\lambda} v_i(\lambda) q_i(\lambda) \qquad i = 1 \ldots n$$

for all labeling assignments $v \in \mathrm{IK}$.
If strict inequalities hold for all $v \neq p$, then p is said to be **strictly consistent**.

Geometrical interpretation.
The support vector q points away from all tangent vectors at p (it has null projection in IK).

Generalization of classical constraint satisfaction problems!



Characterizations

Theorem (Hummel and Zucker, 1983). A labeling $p \in \mathrm{IK}$ is consistent if and only if, for all i = 1…n, the following conditions hold:

1. $q_i(\lambda) = c_i$ whenever $p_i(\lambda) > 0$
2. $q_i(\lambda) \leq c_i$ whenever $p_i(\lambda) = 0$

for some constants $c_1 \ldots c_n$.

The "average local consistency" of a labeling $p \in \mathrm{IK}$ is defined as:

$$A(p) = \sum_i \sum_{\lambda} p_i(\lambda) q_i(\lambda)$$

Theorem (Hummel and Zucker, 1983). If the compatibility matrix R is symmetric, i.e., $r_{ij}(\lambda, \mu) = r_{ji}(\mu, \lambda)$, then any local maximizer $p \in \mathrm{IK}$ of A is consistent.

Understanding the "1976-rule"

Using the Baum-Eagon inequality it is easy to prove the following result, concerning the original Rosenfeld-Hummel-Zucker (RHZ) update rule.

Theorem (Pelillo, 1997). The RHZ relaxation operator is a "growth transformation" for the average local consistency *A*, provided that compatibility coefficients are symmetric. In other words, the algorithm strictly increases the average local consistency on each iteration, i.e.,

$$A(p^{(t+1)}) > A(p^{(t)})$$

for $t = 0,1,\ldots$ until a fixed point is reached.

Theorem (Elfving and Eklundh, 1982; Pelillo, 1997). Let $p \in \mathbb{K}$ be a strictly consistent labeling. Then p is an asymptotically stable equilibrium point for the RHZ relaxation scheme, whether or not the compatibility matrix is symmetric.

Relaxation Labeling and Polymatrix Games

As observed by Miller and Zucker (1991) the consistent labeling problem is equivalent to a polymatrix game.

Indeed, in such formulation we have:

- ✓ Objects = players
- ✓ Labels = pure strategies
- ✓ Weighted labeling assignments = mixed strategies
- ✓ Compatibility coefficients = payoffs

and:

- ✓ Consistent labeling = Nash equilibrium
- ✓ Strictly consistent labeling = strict Nash equilibrium

Further, the RHZ update rule corresponds to discrete-time multi-population "replicator dynamics" used in evolutionary game theory (see previous talk).

Semi-Supervised Learning

Unsupervised learning

- Learning with unlabeled data $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$

Supervised learning

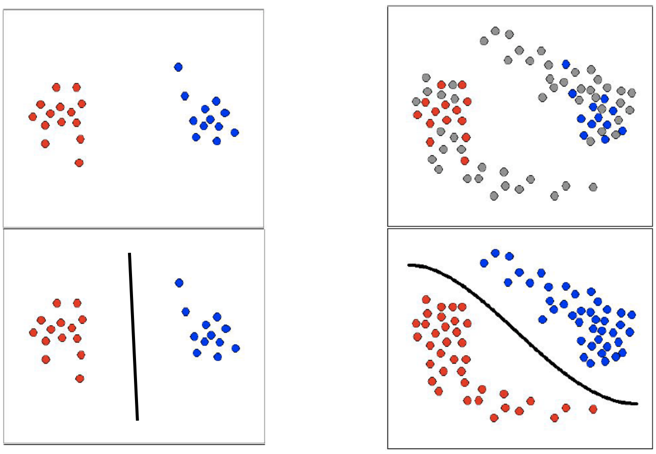
- Learning with labeled data $\{(\mathbf{x}_1, y_1, \ldots, \mathbf{x}_n, y_n)\}$
- Finding a mapping from the feature space to the label space $f : \mathcal{X} \rightarrow \mathcal{Y}$

Semi-supervised learning

- Learning with labeled and unlabeled data
  - labeled data: $\{(\mathbf{x}_1, y_1), \ldots, \mathbf{x}_\ell, y_\ell)\}$
  - unlabeled data: $\{\mathbf{x}_{\ell+1}, \ldots, \mathbf{x}_n\}$

Can we find a better classifier from both labeled and unlabeled data?

Unlabeled Points Can Help…



Adapted from: O. Duchene, J.-Y. Audibert, R. Keriven, J. Ponce, and F. Ségonne. Segmentation by transduction. *CVPR 2008*.

Graph Transduction

Given a set of data points grouped into:

- ✓ labeled data: $\{(\mathbf{x}_1, y_1), \ldots, \mathbf{x}_\ell, y_\ell)\}$
- ✓ unlabeled data: $\{\mathbf{x}_{\ell+1}, \ldots, \mathbf{x}_n\}$ $\qquad \ell \ll n$

Express data as a graph *G=(V,E)*

- ✓ *V* : nodes representing labeled and unlabeled points
- ✓ *E* : pairwise edges between nodes weighted by the similarity between the corresponding pairs of points

Goal: Propagate the information available at the labeled nodes to unlabeled ones in a "consistent" way.

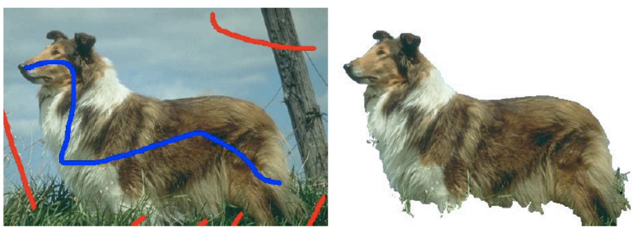
Cluster assumption:

- ✓ The data form distinct clusters
- ✓ Two points in the same cluster are expected to be in the same class

An Application: Interactive Image Segmentation

Segmentation by transduction: "Given a set of user-supplied seeds representative of each region to be segmented in an image, generate a segmentation of the entire image that is consistent with the seeds."

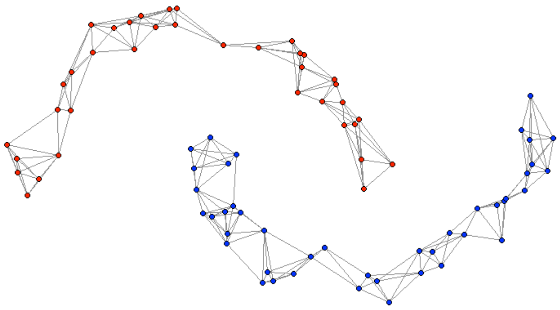
From: O. Duchene, J.-Y. Audibert, R. Keriven, J. Ponce, and F. Ségonne. Segmentation by transduction. *CVPR 2008*.



A Special Case: Unweighted Undirected Graphs

A simple case of graph transduction in which the graph *G* is an unweighted undirected graph:

- ✓ An edge denotes perfect similarity between points
- ✓ The adjacency matrix of *G* is a 0/1 matrix



The cluster assumption: Each node in a connected component of the graph should have the same class label.

A Special Case: Unweighted Undirected Graphs

This toy problem can be formulated as a (binary) **constraint satisfaction problem** (CSP) as follows:

- ✓ The set of variables: $V = \{v_1, \ldots, v_n\}$
- ✓ Domains: $D_{v_i} = \begin{cases} \{y_i\} & \text{for all } 1 \le i \le l \\ Y & \text{for all } l+1 \le i \le n \end{cases}$
- ✓ Binary constraints: $\forall i,j$: if $a_{ij} = 1$, then $v_i = v_j$
  e.g. for a 2-class problem $R_{ij} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Each assignment of values to the variables satisfying all the constraints is a solution of the CSP, thereby providing a consistent labeling for the unlabeled points.

Goal: Generalize to real-valued (soft) constraints

Idea: Use consistency criterion of relaxation labeling (= Nash equilibrium)

The Graph Transduction Game

Assume:

- ✓ the players participating in the game correspond to the vertices of the graph
- ✓ the set of strategies available to each player denote the possible hypotheses about its class membership

- labeled players $\mathcal{I}_\ell = \{\mathcal{I}_{\ell|1}, \ldots, \mathcal{I}_{\ell|c}\}$
- unlabeled players $\mathcal{I}_u$

Labeled players choose their strategies at the outset:

- ✓ each player $i \in \mathcal{I}_{l|k}$ always play its k^{th} pure strategy.

The transduction game is in fact played among the unlabeled players to choose their memberships.

By assuming that only pairwise interactions are allowed, we obtain a polymatrix game that can be solved used standard relaxation labeling / replicator algorithms.

Defining the Payoffs

If the fixed choices of labeled players are considered, the payoff function is:

$$u_i(x) = \sum_{j \in \mathcal{I}_\mathcal{U}} x_i^T A_{ij} x_j + \sum_{k=1}^{c} \sum_{j \in \mathcal{I}_{\mathcal{D}|k}} x_i^T (A_{ij})_k$$

But how to specify partial payoff matrices?

If $A = (A_{ij})$ represent partial payoff matrices in block form, we define

$$A = I_c \otimes W$$

e.g., for a 3-class problem:

$$A_{ij} = \begin{bmatrix} w_{ij} & 0 & 0 \\ 0 & w_{ij} & 0 \\ 0 & 0 & w_{ij} \end{bmatrix}$$

We end up with a generalization of the binary CSP for the toy transduction problem!

Example Results: Symmetric Symilarities

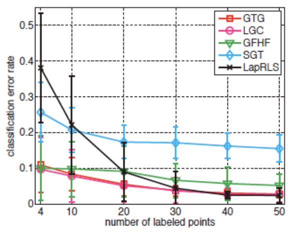
Data set used: *USPS*, *YaleB*, *Scene*, *20-news*

	USPS	*YaleB*	*Scene*	*20-news*
# objects	3874	1755	2688	3970
# dimensions	256	1200	512	8014
# classes	4	3	8	4

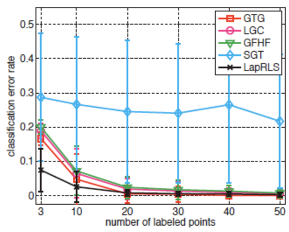
Methods compared:

- ✓ *Gaussian fields and harmonic functions* (GFHF) (Zhu *et al.*, 2003)
- ✓ *Spectral Graph Transducer* (SGT) (Joachims, 2003)
- ✓ *Local and global consistency* (LGC) (Zhou *et al.*, 2004)
- ✓ *Laplacian Regularized Least Squares* (LapRLS) (Belkin *et al.*, 2006)

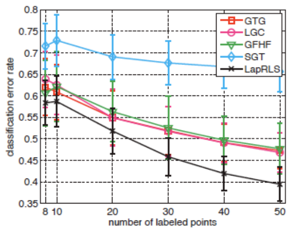
Example Results: Symmetric Symilarities



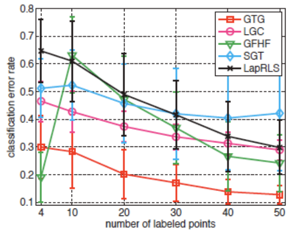
(a) *USPS*



(b) *YaleB*



(c) *Scene*



(d) *20-news*

In short…

Graph transduction can be formulated as a (polymatrix) non-cooperative game (i.e., a consistent labeling problem).

The proposed game-theoretic framework can cope with **symmetric, negative and asymmetric similarities** (none of the existing techniques is able to deal with all three types of similarities).

Experimental results on standard datasets show that our approach is not only more general but also competitive with standard approaches.

A. Erdem and M. Pelillo. Graph transduction as a non-cooperative game. *Neural Computation* (in press) *(preliminary version in GbR 2011)*.

Extensions

The approach described here can be naturally extended along several directions:

- ✓ Using more powerful algorithms than "plain" replicator dynamics (e.g., Porter et al., 2008; Rota Bulò and Bomze, 2010)
- ✓ Dealing with high-order interactions (i.e., hypergraphs) (e.g., Agarwal et al., 2006; Rota Bulò and Pelillo, 2009)
- ✓ From the "homophily" to the "Hume" similarity principle?
- ✓ Introducing uncertainty in "labeled" players

References

J. T. Howson. Equilibria of polymatrix games. *Management Science* (1972).

A. Rosenfeld, R. A. Hummel, and S. W. Zucker. Scene labeling by relaxation operations. *IEEE Trans. Syst. Man. Cybern* (1976).

R. A. Hummel and S. W. Zucker. On the foundations of relaxation labeling processes. *IEEE Trans. Pattern Anal. Machine Intell.* (1983).

M. Pelillo. The dynamics of nonlinear relaxation labeling processes. *J. Math. Imaging and Vision* (1997).

J. Kittler and J. Illingworth. Relaxation labeling algorithms – A review. *Image and Vision Computing* (1986).

D. A. Miller and S. W. Zucker. Copositive-plus Lemke algorithm solves polymatrix games. *Operation Research Letters* (1991).

D. A. Miller and S. W. Zucker. Efficient simplex-like methods for equilibria of nonsymmetric analog networks. *Neural Computation (*1992).

S. W. Zucker. Relaxation labeling: 25 years and still iterating. In: L. S. Davis (Ed.) *Foundations of Image Understanding*. Kluwer Academic Publisher (2001).

A. Erdem and M. Pelillo. Graph transduction as a non-cooperative game. *Neural Computation* (in press) *(preliminary version in GbR 2011)*.

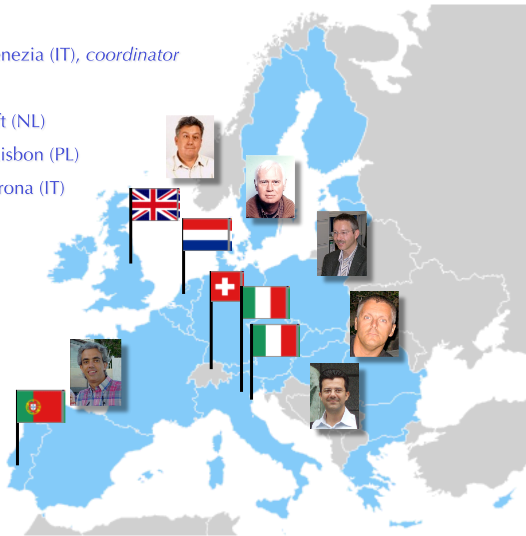
Contributors

- ✓ Massimiliano Pavan (University of Venice)
- ✓ Samuel Rota Bulò (University of Venice)
- ✓ Andrea Torsello (University of Venice)
- ✓ Aykut Erdem (Hacettepe University, Ankara)
- ✓ Immanuel Bomze (University of Vienna)
- ✓ Steve Zucker (Yale University)
- ✓ Kaleem Siddiqi (McGill University)

The SIMBAD FP7 Project

Beyond Features:
Similarity-Based Pattern Analysis and Recognition

1. Università Ca' Foscari di Venezia (IT), *coordinator*
2. University of York (UK)
3. Technische Universität Delft (NL)
4. Insituto Superior Técnico, Lisbon (PL)
5. Università degli Studi di Verona (IT)
6. ETH Zürich (CH)



SIMBAD

http://simbad-fp7.eu

SIMBAD 2011

http://www.dsi.unive.it/~simbad