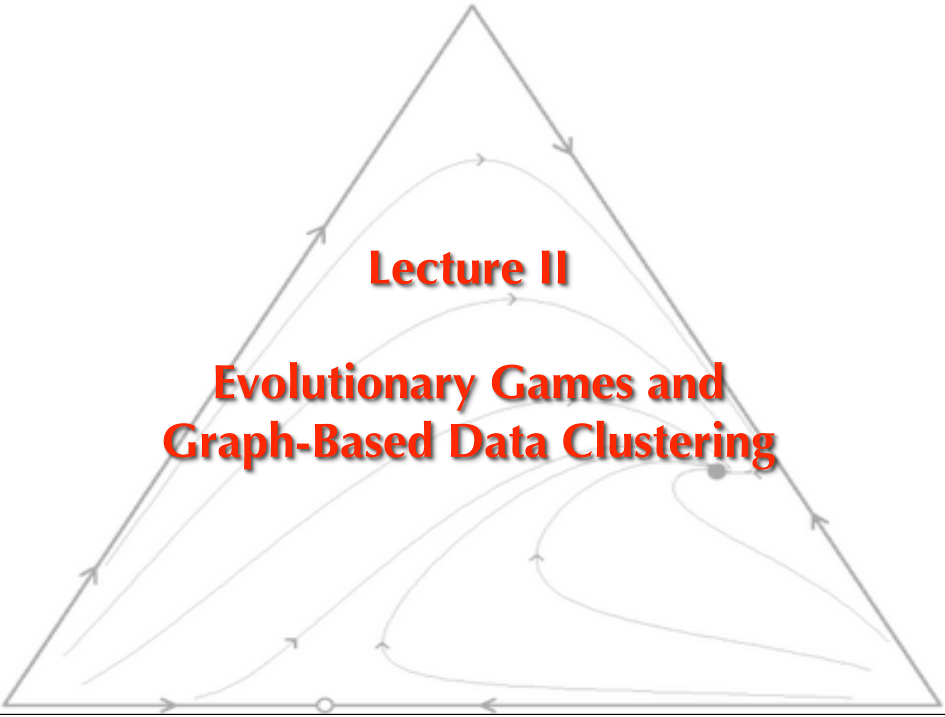
Game Theory in Graph-Based Computer Vision and Pattern Recognition

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 II

Evolutionary Games and Graph-Based Data Clustering

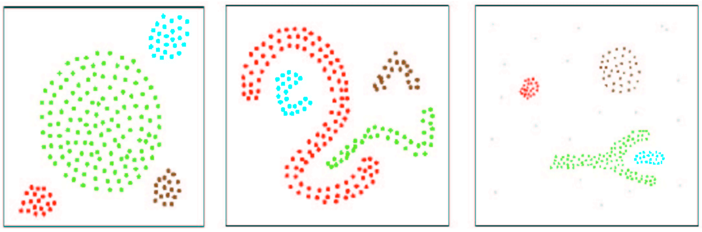
The "Classical" Clustering Problem

Given:

- a set of n "objects"
- an $n \times n$ matrix A of pairwise similarities

} = an edge-weighted graph

Goal: *Partition* the input objects (the vertices of the graph) into maximally homogeneous groups (i.e., clusters).



Applications

Clustering problems abound in many areas of computer science and engineering.

A short list of applications domains:

Image processing and computer vision
Computational biology and bioinformatics
Information retrieval
Document analysis
Medical image analysis
Data mining
Signal processing
…

For a review see, e.g., A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters* 31(8):651-666, 2010.

The Need for Non-exhaustive Clusterings

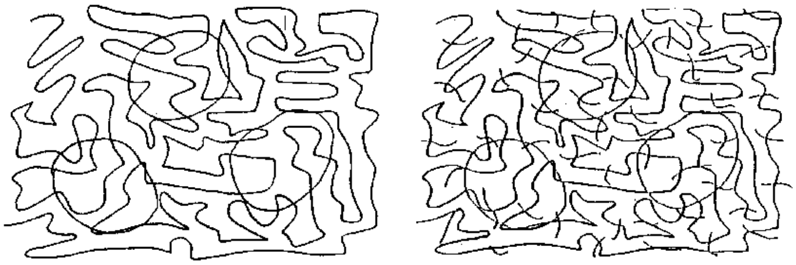


Figure 1a. Three prominent blobs are perceived immediately and with little effort. Locally, the blobs are similar to the background contours. (adopted from Mahoney (1986)

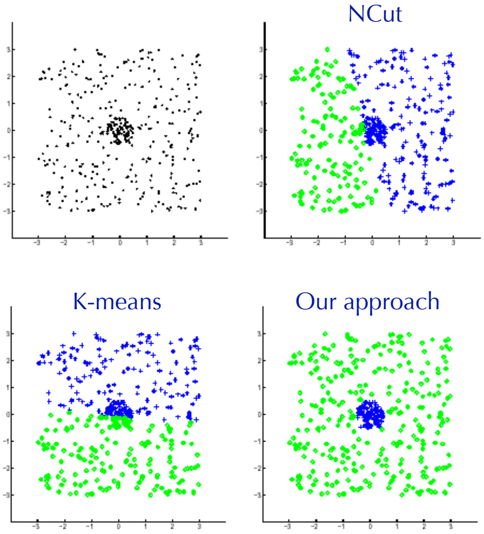
Figure 1b. Intersections were added to illustrate that the blobs are not distinguished by virtue of their intersections with the background curves.

Separating Structure from Clutter

Figure 2. A circle in a background of 200 randomly placed and oriented segments The circle is still perceived immediately although its contour is fragmented.

Figure 3. An edge image of a car in a cluttered background. Our attention is drawn immediately to the region of interest. It seems that the car need not be recognized to attract our attention. The car also remains salient when parallel lines and small blobs are removed, and when the less textured region surrounding parts of the car is filled in with more texture.

Separating Structure from Clutter



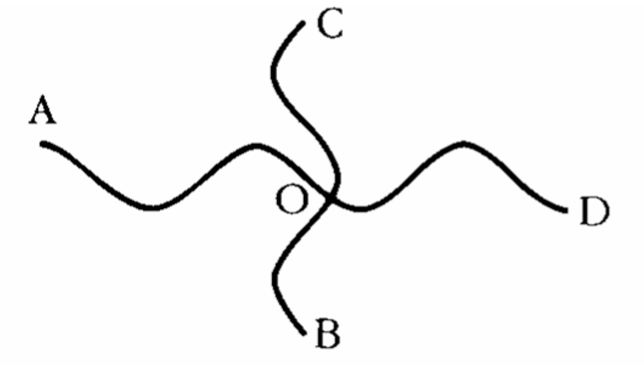
One-class Clustering

"[…] in certain real-world problems, natural groupings are found among only on a small subset of the data, while the rest of the data shows little or no clustering tendencies.

In such situations it is often more important to cluster a small subset of the data very well, rather than optimizing a clustering criterion over all the data points, particularly in application scenarios where a large amount of noisy data is encountered."

G. Gupta and J. Ghosh. Bregman bubble clustering: A robust framework for mining dense cluster. *ACM Trans. Knowl. Discov. Data* (2008).

When Groups Overlap



Does *O* belong to *AD* or to *BC* (or to none)?

The Need for Overlapping Clusters

Partitional approaches impose that each element cannot belong to more than one cluster. There are a variety of important applications, however, where this requirement is too restrictive.

Examples:

- ✓ clustering micro-array gene expression data
- ✓ clustering documents into topic categories
- ✓ perceptual grouping
- ✓ segmentation of images with transparent surfaces

References:

- ✓ N. Jardine and R. Sibson. The construction of hierarchic and non-hierarchic classifications. *Computer Journal*, 11:177–184, 1968
- ✓ A. Banerjee, C. Krumpelman, S. Basu, R. J. Mooney, and J. Ghosh. Model-based overlapping clustering. *KDD 2005*.
- ✓ K. A. Heller and Z. Ghahramani. A nonparametric Bayesian approach to modeling overlapping clusters. *AISTATS 2007*.

The Symmetry Assumption

«Similarity has been viewed by both philosophers and psychologists as a prime example of a symmetric relation. Indeed, the assumption of symmetry underlies essentially all theoretical treatments of similarity.

Contrary to this tradition, the present paper provides empirical evidence for asymmetric similarities and argues that **similarity should not be treated as a symmetric relation**.»

Amos Tversky
"Features of similarities," *Psychol. Rev.* (1977)

Examples of asymmetric (dis)similarities

- ✓ Kullback-Leibler divergence
- ✓ Directed Hausdorff distance
- ✓ Tversky's contrast model

What is a Cluster?

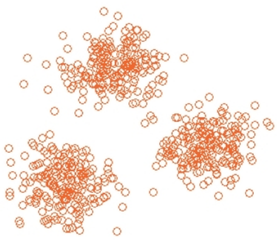
No universally accepted (formal) definition of a "cluster" but, informally, a cluster should satisfy two criteria:

Internal criterion

all "objects" *inside* a cluster should be highly similar to each other

External criterion

all "objects" *outside* a cluster should be highly dissimilar to the ones inside



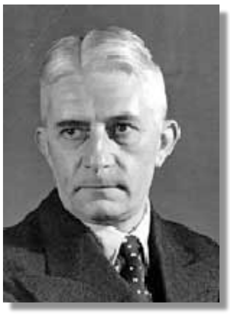
The Notion of "Gestalt"

«In most visual fields the contents of particular areas "belong together" as circumscribed units from which their surrounding are excluded.»

W. Köhler, *Gestalt Psychology* (1947)

«In gestalt theory the word "Gestalt" means any segregated whole.»

W. Köhler (1929)



Data Clustering: Old *vs.* New

By answering the question "what is a cluster?" we get a novel way of looking at the clustering problem.

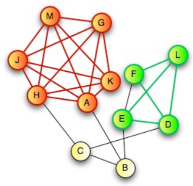
```
Clustering_old(V,A,k)
   V1,V2,...,Vk <- My_favorite_partitioning_algorithm(V,A,k)
   return V1,V2,...,Vk

------

Clustering_new(V,A)
   V1,V2,...,Vk <- Enumerate_all_clusters(V,A)
   return V1,V2,...,Vk

Enumerate_all_clusters(V,A)
   repeat
      Extract_a_cluster(V,A)
   until all clusters have been found
   return the clusters found
```

A Special Case: Binary Symmetric Affinities

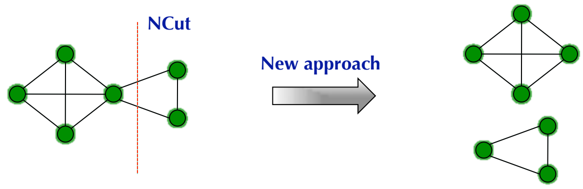


Suppose the similarity matrix is a binary (0/1) matrix.

Given an unweighted undirected graph $G=(V,E)$:

A *clique* is a subset of mutually adjacent vertices
A *maximal clique* is a clique that is not contained in a larger one

In the 0/1 case, a meaningful notion of a cluster is that of a *maximal clique*.



Advantages of the New Approach

- ✓ No need to know the number of clusters in advance (since we extract them sequentially)
- ✓ Leaves clutter elements unassigned (useful, e.g., in figure/ground separation or one-class clustering problems)
- ✓ Allows extracting overlapping clusters

Need a partition?

```
Partition_into_clusters(V,A)
  repeat
     Extract_a_cluster
     remove it from V
  until all vertices have been clustered
```

ESS's as Clusters

We claim that ESS's abstract well the main characteristics of a cluster:

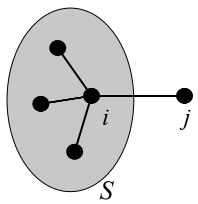
- ✓ **Internal coherency**: High mutual support of all elements within the group.
- ✓ **External incoherency**: Low support from elements of the group to elements outside the group.

Basic Definitions

Let $S \subseteq V$ be a non-empty subset of vertices, and $i \in S$.

The **(average) weighted degree** of i w.r.t. S is defined as:

$$\text{awdeg}_S(i) = \frac{1}{|S|} \sum_{j \in S} a_{ij}$$



Moreover, if $j \notin S$, we define:

$$\phi_S(i,j) = a_{ij} - \text{awdeg}_S(i)$$

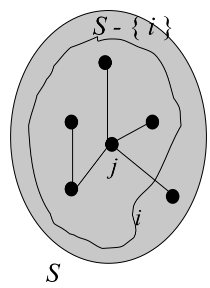
Intuitively, $\phi_S(i,j)$ measures the similarity between vertices j and i, with respect to the (average) similarity between vertex i and its neighbors in S.

Assigning Weights to Vertices

Let $S \subseteq V$ be a non-empty subset of vertices, and $i \in S$.

The **weight** of i w.r.t. S is defined as:

$$w_S(i) = \begin{cases} 1 & \text{if } |S| = 1 \\ \sum_{j \in S-\{i\}} \phi_{S-\{i\}}(j,i) w_{S-\{i\}}(j) & \text{otherwise} \end{cases}$$

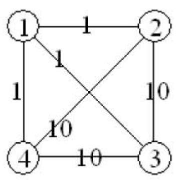


Further, the **total weight** of S is defined as:

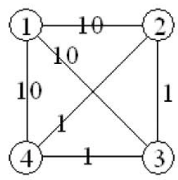
$$W(S) = \sum_{i \in S} w_S(i)$$

Interpretation

Intuitively, $w_S(i)$ gives us a measure of the overall (relative) similarity between vertex i and the vertices of S-{i} with respect to the overall similarity among the vertices in S-{i}.



$w_{\{1,2,3,4\}}(1) < 0$

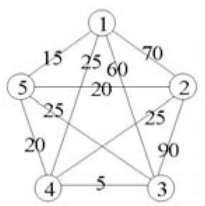


$w_{\{1,2,3,4\}}(1) > 0$

Dominant Sets

Definition (Pavan and Pelillo, 2003, 2007). A non-empty subset of vertices $S \subseteq V$ such that $W(T) > 0$ for any non-empty $T \subseteq S$, is said to be a **dominant set** if:

1. $w_S(i) > 0$, for all $i \in S$ (internal homogeneity)
2. $w_{S \cup \{i\}}(i) < 0$, for all $i \notin S$ (external homogeneity)



The set {1,2,3} is dominant.

Dominant sets ≡ clusters

The Clustering Game

Consider the following "clustering game."

- ✓ Assume a preexisting set of objects O and a (possibly asymmetric) matrix of affinities A between the elements of O.
- ✓ Two players with complete knowledge of the setup play by simultaneously selecting an element of O.
- ✓ After both have shown their choice, each player receives a payoff, monetary or otherwise, proportional to the affinity that the chosen element has with respect to the element chosen by the opponent.

Clearly, it is in each player's interest to pick an element that is strongly supported by the elements that the adversary is likely to choose.

Hence, in the (pairwise) clustering game:

- ✓ There are 2 players
- ✓ The objects to be clustered are the pure strategies
- ✓ The (null-diagonal) affinity matrix coincides with the similarity matrix

Dominant Sets are ESS's

Theorem (Torsello, Rota Bulò and Pelillo, 2006). Evolutionary stable strategies of the clustering game with affinity matrix A are in a one-to-one correspondence with dominant sets.

Note. Generalization of well-known Motzkin-Straus theorem from graph theory.

Dominant-set clustering

- ✓ To get a single dominant-set cluster use, e.g., replicator dynamics (but see Rota Bulò, Pelillo and Bomze, *CVIU* in press, for faster dynamics)
- ✓ To get a partition use a simple *peel-off* strategy: iteratively find a dominant set and remove it from the graph, until all vertices have been clustered
- ✓ To get overlapping clusters, enumerate dominant sets (see Bomze, 1992; Torsello, Rota Bulò and Pelillo, 2008)

Special Case: Symmetric Affinities

Given a symmetric real-valued matrix A (with null diagonal), consider the following Standard Quadratic Programming problem (StQP):

$$\begin{array}{ll} \text{maximize} & f(x) = x^T A x \\ \text{subject to} & x \in \Delta \end{array}$$

Note. The function $f(x)$ provides a measure of cohesiveness of a cluster (see Pavan and Pelillo, 2003, 2007; Sarkar and Boyer, 1998; Perona and Freeman, 1998).

ESS's are in one-to-one correspondence to (strict) local solutions of StQP

Note. In the 0/1 (symmetric) case, ESS's are in one-to-one correspondence to (strictly) maximal cliques (Motzkin-Straus theorem).

Measuring the Degree of Cluster Membership

The components of the converged vector give us a measure of the participation of the corresponding vertices in the cluster, while the value of the objective function provides of the cohesiveness of the cluster.

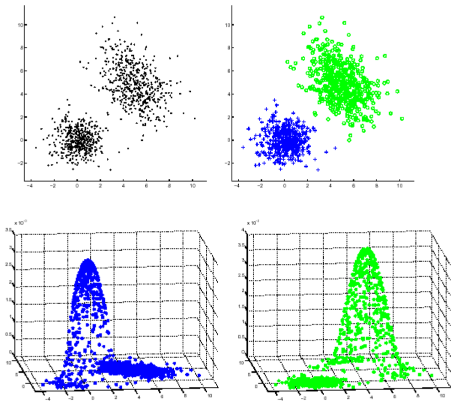


Image Segmentation

Image segmentation problem:

Decompose a given image into *segments*, i.e. regions containing "similar" pixels.

First step in many computer vision problems

Example: Segments might be regions of the image depicting the same object.

Semantics Problem: *How should we infer objects from segments?*

Application to Image Segmentation

An image is represented as an edge-weighted undirected graph, where vertices correspond to individual pixels and edge-weights reflect the "similarity" between pairs of vertices.

For the sake of comparison, in the experiments we used the same similarities used in Shi and Malik's normalized-cut paper (PAMI 2000).

To find a hard partition, the following *peel-off* strategy was used:

```
Partition_into_dominant_sets(G)
Repeat
   find a dominant set
   remove it from graph
until all vertices have been clustered
```

To find a single dominant set we used replicator dynamics (but see Rota Bulò, Pelillo and Bomze, *CVIU 2011*, for faster game dynamics).

Experimental Setup

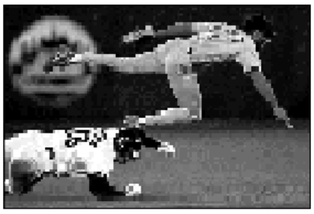
The similarity between pixels i and j was measured by:

$$w(i,j) = \exp\left(\frac{-\|\mathbf{F}(i) - \mathbf{F}(j)\|_2^2}{\sigma^2}\right)$$

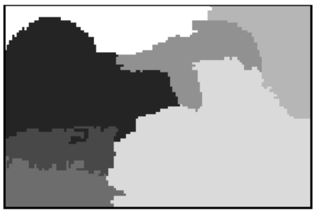
where σ is a positive real number which affects the decreasing rate of w, and:

- $\mathbf{F}(i) \equiv$ (normalized) intensity of pixel i, for intensity segmentation
- $\mathbf{F}(i) = [v, vs\sin(h), vs\cos(h)](i)$, where h, s, v are the HSV values of pixel i, for color segmentation
- $\mathbf{F}(i) = [|I * f_1|, \ldots, |I * f_k|](i)$ is a vector based on texture information at pixel i, the f_i being DOOG filters at various scales and orientations, for texture segmentation

Intensity Segmentation Results

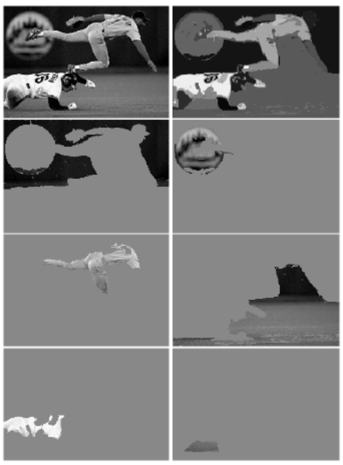


Dominant sets



Ncut

Intensity Segmentation Results

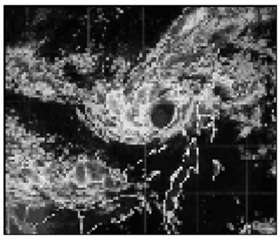


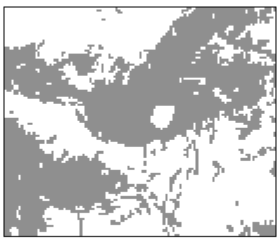
Felzenszwalb and Huttenlocher (2003).

Intensity Segmentation Results

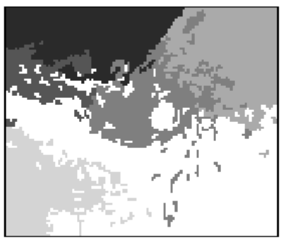
Gdalyahu, Weinshall, and Werman (2001).

Intensity Segmentation Results





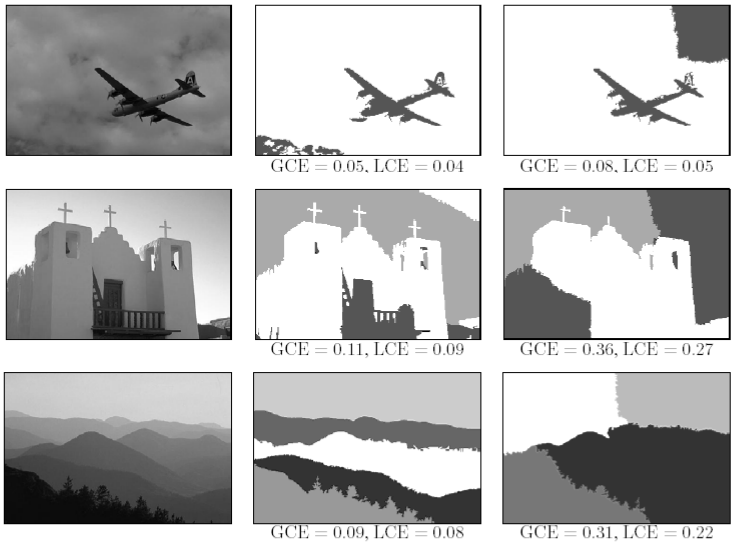
Dominant sets



Ncut

Results on the Berkeley Dataset

Dominant sets | Ncut

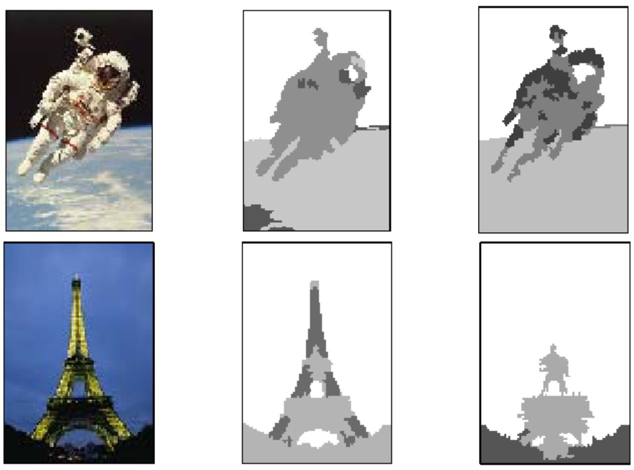


GCE = 0.05, LCE = 0.04 | GCE = 0.08, LCE = 0.05

GCE = 0.11, LCE = 0.09 | GCE = 0.36, LCE = 0.27

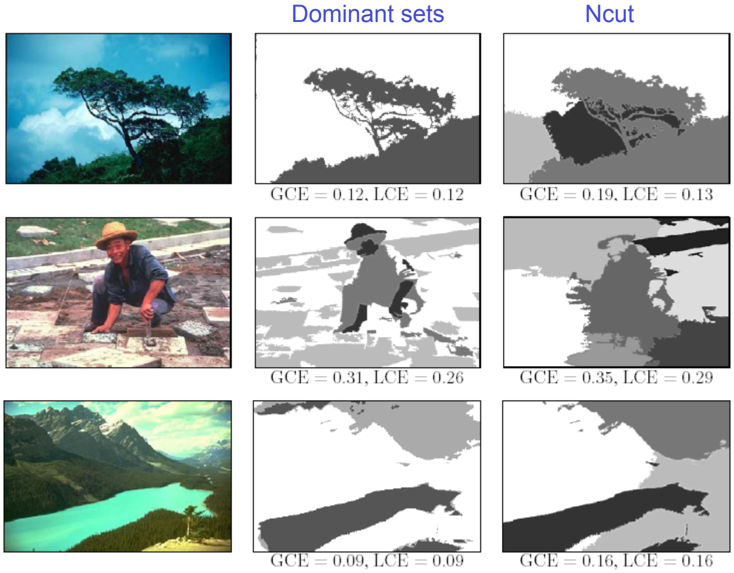
GCE = 0.09, LCE = 0.08 | GCE = 0.31, LCE = 0.22

Color Segmentation Results

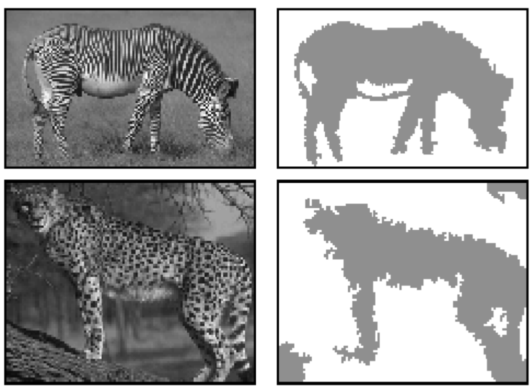


Original image | Dominant sets | Ncut

Results on the Berkeley Dataset

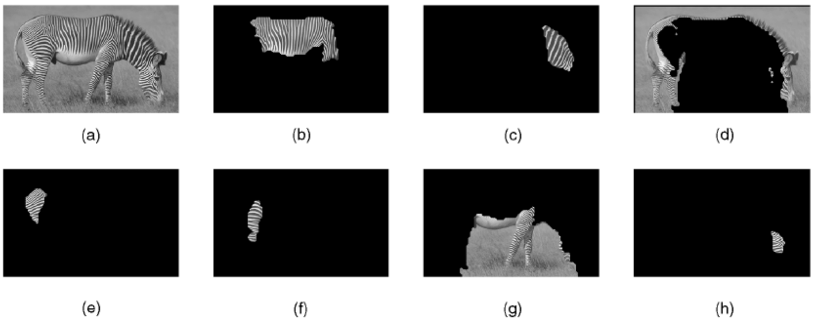


Texture Segmentation Results



Dominant sets

Texture Segmentation Results



NCut

Other Applications of Dominant-Set Clustering

Bioinformatics
Identification of protein binding sites (*Zauhar and Bruist*, 2005)
Clustering gene expression profiles (*Li et al*, 2005)
Tag Single Nucleotide Polymorphism (SNPs) selection (*Frommlet*, 2010)

Security and video surveillance
Detection of anomalous activities in video streams (*Hamid et al.*, CVPR'05; AI'09)
Detection of malicious activities in the internet (*Pouget et al.*, J. Inf. Ass. Sec. 2006)

Content-based image retrieval
Wang et al. (Sig. Proc. 2008); *Giacinto and Roli (*2007)

Analysis of fMRI data
Neumann et al (NeuroImage 2006); *Muller et al* (J. Mag Res Imag. 2007)

Video analysis, object tracking, human action recognition
Torsello et al. (EMMCVPR'05); *Gualdi et al.* (IWVS'08); *Wei et al.* (ICIP'07)

Multiple instance learning
Erdem and Erdem (SIMBAD'11)

Feature selection
Hancock et al. (GbR'11; ICIAP'11; SIMBAD'11)

Image matching and registration
Torsello et al. (IJCV 2011, ICCV'09, CVPR'10, ECCV'10)

In a nutshell…

The dominant-set (ESS) approach:

- ✓ makes no assumption on the underlying (individual) data representation
- ✓ makes no assumption on the structure of the affinity matrix, being it able to work with asymmetric and even negative similarity functions
- ✓ does not require *a priori* knowledge on the number of clusters (since it extracts them sequentially)
- ✓ leaves clutter elements unassigned (useful, e.g., in figure/ground separation or one-class clustering problems)
- ✓ allows principled ways of assigning out-of-sample items *(NIPS'04)*
- ✓ allows extracting overlapping clusters *(ICPR'08)*
- ✓ generalizes naturally to hypergraph clustering problems, i.e., in the presence of high-order affinities, in which case the clustering game is played by more than two players *(NIPS'09)*
- ✓ extends to hierarchical clustering *(ICCV'03: EMMCVPR'09)*

References

M. Pavan and M. Pelillo. A new graph-theoretic approach to clustering and segmentation. *CVPR 2003*.

M. Pavan and M. Pelillo. Dominant sets and hierarchical clustering. *ICCV 2003*.

M. Pavan and M. Pelillo. Efficient out-of-sample extension of dominant-set clusters. *NIPS 2004*.

A. Torsello, S. Rota Bulò and M. Pelillo. Grouping with asymmetric affinities: A game-theoretic perspective. *CVPR 2006*.

M. Pavan and M. Pelillo. Dominant sets and pairwise clustering. *PAMI 2007*.

A. Torsello, S. Rota Bulò and M. Pelillo. Beyond partitions: Allowing overlapping groups in pairwise clustering. *ICPR 2008*.

S. Rota Bulò and M. Pelillo. A game-theoretic approach to hypergraph clustering. *NIPS 2009*.

M. Pelillo. What is a cluster? Perspectives from game theory. *NIPS 2009 Workshop on "Clustering: Science or Art?"* (talk available on videolectures.net).

S. Rota Bulò, M. Pelillo and I. M. Bomze. Graph-based quadratic optimization: A fast evolutionary approach. *CVIU 2011*.