Multi-labelled Image Segmentation in Irregular, Weighted Networks:
A Spatial Autocorrelation Approach

Raphaël Ceré and François Bavaud
Department of Geography and Sustainability, University of Lausanne, Switzerland

Keywords: Free Energy, Image Segmentation, Iterative Clustering, K-means, Laplacian, Modularity, Multivariate Features, Ncut, Soft Membership, Spatial Autocorrelation, Spatial Clustering.

Abstract: Image segmentation and spatial clustering both face the same primary problem, namely to gather together spatial entities which are both spatially close and similar regarding their features. The parallelism is particularly obvious in the case of irregular, weighted networks, where methods borrowed from spatial analysis and general data analysis (soft K-means) may serve at segmenting images, as illustrated on four examples. Our semi-supervised approach considers soft memberships (fuzzy clustering) and attempts to minimize a free energy functional made of three ingredients: a within-cluster features dispersion (hard K-means), a network partitioning objective (such as the Ncut or the modularity) and a regularizing entropic term, enabling an iterative computation of the locally optimal soft clusters. In particular, the second functional enjoys many possible formulations, arguably helpful in unifying various conceptualizations of space through the probabilistic selection of pairs of neighbours, as well as their relation to spatial autocorrelation (Moran’s I).

1 INTRODUCTION

Regional data analysis, as performed on geographic information systems, deals with a notion of “where” (the spatial disposition of regions), a notion of “what” (the regional features), and a notion of “how much” (the relative importance of regions, as given by their surface or the population size). The data define a marked, weighted network, generally irregular (think e.g. of administrative units): weighted vertices represent the regions, weighted edges measure the proximity between regions, on which uni- or multivariate features (the marks) are defined.

Much the same can be said of an image made of pixels, that is a collection of elements embedded in a bidimensional layout. The regularity of the setup (regular grid, uniform weights, binary regular adjacencies) is exploited in most segmentation algorithms, but the latter may become inadapted, precisely, under irregular situations, such as pixels of various sizes or importance, aggregated pixels, irregular boundaries or connectivities, multi-layered or partially missing data.

This paper proposes an iterative algorithm for semi-supervised image segmentation, directly adapted from regional clustering procedures in spatial analysis, themselves originating in (non-marked) network clustering. In a nutshell, the purely spatial standard procedures aimed at Ncut minimisation (Shi and Malik, 2000; Grady and Schwartz, 2006) or modularity maximization (Newman, 2006) are enriched with a features dissimilarity term, central in the K-means approach, and further regularized by an entropy term, favoring the emergence of soft clusters, and allowing an iterative computation of locally optimal solutions.

The fields of spatial analysis, in particular spatial clustering, on one hand, and image segmentation on the other hand, seem currently to be investigated by distinct, non-overlapping communities. Yet, both communities arguably share the same what-versus-where-trading challenge, aimed at obtaining clusters both homogeneous and connected.

In this work in progress, we first define the main quantities of interest and the iterative algorithm itself (section 2). The definition of the canonical measure of image homogeneity, namely Moran’s I in the present spatial context, is recalled in section 2.3. Section (3) illustrates the method on four case studies. A conclusion (section 4) lists some further research lines, to be investigated in a close future. The appendix details the construction of the so-called exchange matrix, as well as the test of spatial autocorrelation in a weighted setting.
2 DEFINITIONS AND
FORMALISM

The formalism we consider extends the spatial autocorrelation formalism used in Quantitative Geography and Spatial Econometrics to the case of weighted, irregular regions, as well as to multivariate features. It turns out to be broad enough to provide a flexible framework for unsupervised or semi-supervised generalized image segmentation, where the "generalized images" under consideration can be made of irregular pixels, irregularly inter-connected, and endowed with multivariate features.

Space as a weighted network: the exchange matrix E

Specifically, consider $n$ regions (generalized pixels) with relative weights $f_i > 0$, summing to $\sum_{i=1}^n f_i = 1$, together with an $n \times n$ symmetric non-negative exchange matrix $E = (e_{ij})$, and weight-compatible in the sense $e_{ii} = \sum_{i=1}^n e_{ij} = f_i$. The exchange matrix $E$ interprets as a joint probability $p(i, j) = e_{ij}$ to select the pair of regions $i$ and $j$ (edges), and defines a weighted unoriented network. Its margins interpret as the probability $p(i) = f_i$ to select region $i$ (vertices).

Weight-compatible exchange matrices $E$ define a continuous neighborhood relation between regions. They can be constructed from $f$ and the adjacency matrix $A$, or from another spatial proximity of distance matrix (see the appendix). The row-standardized matrix of spatial weights $W = (w_{ij})$ of spatial autoregressive models obtains as $w_{ij} = e_{ij} / f_i$, and constitutes the transition matrix of a Markov chain with stationary distribution $f$.

Multivariate features: the dissimilarity matrix $D$

Regional features can consist of univariate grey levels, multivariate color or spectral intensities, or (in a geographical context) any regional variable such as the proportions of specific land uses, population density, proportion of party B voters, etc. Multivariate characteristics $x_i$ are suitably combined into an $n \times n$ squared Euclidean dissimilarities $D_{ij} = \|x_i - x_j\|^2$.

2.1 Image Segmentation (Regional Clustering)

A soft regional clustering or image segmentation into $m$ groups is described by non-negative $n \times m$ membership matrix $Z = (z_{ig})$ with $z_{ig} = p(g|i)$ denotes the probability that region (pixel) $i$ belongs to group $g$, and obeys $z_{ig} = \sum_{g=1}^m z_{ig} = 1$. The weights $p$ of the corresponding groups obtain as $p_g = \sum_i f_i z_{ig}$, and the regional distribution of groups as $f^g_i = f_i z_{ig} / p_g = p(g|i)$, obeying $f^g_i = 1$. The region-group dependency can be measured by the mutual information

$$I[Z] = \sum_{ig} p(i, g) \ln \frac{p(i, g)}{p(i) p(g)} = \sum_{ig} f_i z_{ig} \ln \frac{z_{ig}}{p_g}$$

A good clustering should consist of homogeneous groups made of regions not too dissimilar regarding their features, that is insuring a low value of the within-group inertia (Bavaud, 2009)

$$\Delta w[Z] = \sum_{g=1}^m \rho_g \Delta_g$$

where $\Delta_g = \frac{1}{2} \sum_{ij} f^g_i f^g_j D_{ij}$

A good clustering should also avoid to separate a pair of spatially strongly connected regions, that is to insure a low value of the generalized cut

$$G[Z] = \frac{1}{2} \sum_{ig} G(p_g) \sum_{ij} e_{ij} (z_{ig} - z_{jg})^2$$

where $G(p) \geq 0$ is non-increasing in $p$. The choice $G(p) = 1/p$ amounts to the N-cut (Shi and Malik, 2000), while the choice $G(p) = 1$, we shall adopt here, is equivalent to the modularity criterion (Newman, 2006).

We consider the regularized clustering problem, consisting in finding a clustering $Z$ minimizing the free energy functional

$$F[Z] = \beta \Delta w[Z] + \alpha \frac{1}{2} G[Z] + I[Z]$$

The terms $\Delta w$, respectively $G$, behaves as a features dissimilarity energy, respectively a spatial energy, favoring hard partitions obeying $z_{ig} = 0$ or $z_{ig} = 1$. By contrast, the regularizing entropy term $I$ favors the emergence of soft clusterings. Setting $\alpha = 0$ yields the soft $K$-means algorithm (Gaussian mixtures), where $\beta$ is an inverse (dissimilarity) temperature. Canceling the first-order derivative of the free energy with respect to $z_{ig}$ under the constraints $z_{ig} = 1$ yields the minimization condition

$$z_{ig} = \frac{p_g \exp(-\beta D_{ig}^g - \alpha (Lz_{ig})_i)}{\sum_g p_g \exp(-\beta D_{ig}^g - \alpha (Lz_{ig})_i)}$$

where $D_{ig}^g = \sum_j f^g_i f^g_j D_{ij} - \Delta_g$ is the squared Euclidean dissimilarity from $i$ to the centroid of group $g$, and $(Lz_{ig})_i = z_{ig} - \sum_k w_{ik} z_{kg}$ is the Laplacian of $z$ at $i$, comparing its value to the average value of its neighbors.
2.2 Iterative Procedure

Equation (5) can be solved iteratively, updating $\rho_g$, $D_{ig}$ and $L_{zg}$ at each step, and converges towards a local minimum $Z^*$ of $\mathcal{F}[Z]$, which constitutes the searched for soft spatial partition or image segmentation. $Z^*$ can be further hardened by assigning $i$ to group $g = \arg \max_{g} z_{ig}^{(w)}$.

A semi-supervised implementation of the procedure, imposing the membership of a few pixels (and possibly breaking down the monotonic decrease of $\mathcal{F}[Z]$: see figure 5) goes as follows: first, the set $\Omega$ of the $n$ regions is partitioned into two disjoint, non-empty sets, namely the user-defined tagged regions $T$, and the free regions $F$, with $\Omega = T \cup F$ and $T \cap F = \emptyset$. The tagged set $T$ itself consists of $m$ non-empty disjoint subregions $T = \bigcup_{i=1}^{m} T_i$ initially tagged with $m$ distinct strokes applied on a small number of pixels: they form the seeds of the $g = 1, \ldots, m$ figures to be extracted, while the remaining regions will be assigned to the background numbered $g = 0$.

Memberships $Z = \{z_{ig}\}$ consist of $n \times (m + 1)$ non-negative matrices obeying $\sum_{g=0}^{m} z_{ig} = 1$. Their initial value $Z^0$ is set as

$$
\begin{align*}
    z_{ig}^{(0)} &= 1 & \text{if } i \in F \text{ and } g = 0 \\
    &= 1 & \text{if } i \in T_i \text{ and } g = \tau \\
    &= 0 & \text{otherwise}.
\end{align*}
$$

Iteration (5) is then performed. At the end of each loop, the tagged regions are reset to their initial values $z_{ig}^{(0)}$ for all $i \in T_i$. After convergence, one expects the hardened clusters obtained by assigning $i$ to group $g = \arg \max_{g=0}^{m} z_{ig}^{(w)}$ to consist of $m$ connected figures $g = 1, \ldots, m$ each containing the tagged set $T_g$, as well as a remaining background support on $F$.

The iterative image segmentation algorithm summarized below (table 1) requires:

1) a vector of $n$ weights $f_i > 0$ associated to each pixel or region
2) a vector of $n$ grey levels or multivariate characteristics $x_i$
3) a $n \times n$ binary, symmetric, off-diagonal adjacency matrix $A$.

Retaining the soft memberships $Z^{(w)}$ after convergence could possibly provide a novel edge detection mechanism (figure 5(b)), to be investigated in a subsequent work. Namely, one expects $z_{ig}^{(w)} = 0$ or $z_{ig}^{(w)} = 1$ for pixels $i$ located within the interior of homogeneous groups, while $0 < z_{ig}^{(w)} < 1$ for pixels $i$ located at the boundary of two or more groups. As a consequence, the value of the entropy

<table>
<thead>
<tr>
<th>Table 1: Summary of the iterative algorithm.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Begin</strong></td>
</tr>
<tr>
<td>Compute the weight vector ( f_i = 1/n ) for regular grids</td>
</tr>
<tr>
<td>Compute the binary adjacency matrix $A$</td>
</tr>
<tr>
<td>For given $i &gt; 0$, compute the weight-compatible exchange matrix $E(f, A, t)$ by (9), and the matrix of spatial weights as $w_{ij} = c_i f / f_j$</td>
</tr>
<tr>
<td>Compute the features dissimilarity matrix $D_{ij} =</td>
</tr>
<tr>
<td>Initialize the $n \times (m + 1)$ membership matrix $Z^{0}$ as: $z_{ig}^{(0)} = 1$ if $i \in F$ and $g = 0$ $z_{ig}^{(0)} = 1$ if $i \in T_i$ and $g = \tau$ $z_{ig}^{(0)} = 0$ otherwise.</td>
</tr>
<tr>
<td><strong>Loop</strong> for the $r$-th iteration, stop after convergence</td>
</tr>
<tr>
<td>Group weight: $\rho_g := \sum_i f_i z_{ig}$</td>
</tr>
<tr>
<td>Emission probabilities $f_i^g := z_{ig}^{(w)}</td>
</tr>
<tr>
<td>Dissimilarity to the centroid: $D_{ig}^r := \sum_j f_j D_{ij} - \Delta_g$</td>
</tr>
<tr>
<td>Laplacian: $(\mathcal{L} z_{ig}^{(w)}) = z_{ig} - \sum_j w_{ij} z_{ij}^{(w)}$</td>
</tr>
<tr>
<td>Compute $z_{ig}^{(r+1)}$ by (5)</td>
</tr>
<tr>
<td>Re-initialize $z_{ig}^{(w)} = \delta_{ig}$ for $i \in T_i$</td>
</tr>
<tr>
<td>Attribute $i \in F$ to the group $g = \arg \max_{g=0}^{m} z_{ig}^{(w)}$</td>
</tr>
<tr>
<td><strong>End</strong></td>
</tr>
</tbody>
</table>

$$
H_i = -\sum_g z_{ig}^{(w)} \ln z_{ig}^{(w)}
$$

is presumably large for $i$ located at the group frontiers.

Deterministic profiling

The above iterative algorithm has been developed with Python 2.7.12 (van Rossum, 1995) and performed on a CPU Intel Core i7 two Core with a frequency 3.1GHz (Mac OS X 10.10.5). Figure 1 summarizes the “plain” computational performances (i.e. without further optimizing such as Parallel computing or Analysis of complexity) for the main ingredients at stake, as a function of the number of pixels $n$ in a regular setting.

2.3 Spatial Autocorrelation: Moran’s $I$

Average multivariate dissimilarities between regions are expressed by inertias, generalizing the univariate variances. The inertia between randomly selected regions, and the local inertia between neighbours, are respectively defined as

$$
\Delta := \frac{1}{2} \sum_{i,j=1}^{n} f_i f_j D_{ij} \\
\Delta_{loc} := \frac{1}{2} \sum_{i,j=1}^{n} c_i D_{ij}
$$

Comparing the global versus local inertiis provides a multivariate generalization of Moran’s $I$, namely,

$$
I := \frac{\Delta - \Delta_{loc}}{\Delta}
$$

whose values range in $[-1, 1]$. A large positive $I$ is expected for an image made of large patches characterized with constant features, or at least varying smoothly on average (spatial continuity = positive autocorrelation). A large negative $I$ characterizes an image whose pixel features are contrasted, opposite to
We first study the segmentation procedure, before turing to spatial autocorrelation.

Figure 2 depicts the proportion of “yes” $x_i$, respectively $y_i$, for each canton $i$ ($n = 26$) for two emblematic swiss popular initiatives, namely (a) the Initiative against mass immigration which was accepted February the 9th, 2014, on the tie (50.3% of the citizen, and the majority of cantons), respectively (b) the Initiative for minimum wages which was widely refused May the 25, 2014 (23.7% of the citizens, and the majority of cantons). Both results have been aggregated into the dissimilarity $D_{ij} = (x_i - x_j)^2 + (y_i - y_j)^2$, further rescaled in the range $[0, 1]$. Non-uniform canton weights are defined relatively to the population 2015 density as $f_i = \frac{\text{POP}_i}{\text{POP}_{\text{total}}}$, and the “queen” adjacency scheme has been adopted.

Figure 1: CPU times for a range of $n \in [400, 10000]$ pixels. (a) Computing of the binary “rook” adjacency matrix $A$ (with the Scikit-learn package (Pedregosa et al., 2011)), the squared Euclidean profiles dissimilarity $D$, the Markov transition matrix $W$ (for $E$ given), as well as one single iteration $Z(t) \rightarrow Z(t+1)$ of the membership matrix. (b) The computation of the exchange matrix $E(f, A, t)$ by (9) involves a time-consuming complete eigen-decomposition.


Figure 4 depicts a regular multivariate image (RGB) of size $50 \times 54$ with uniform weight vector $f_i = 1/n$ for $n = 2700$. Also, the binary adjacency matrix has been built under the “queen” scheme (8 neighbours), and $D_{ij}$ is the sum of the squared differences between color intensities.

3 ILLUSTRATIONS

We consider four small datasets:

- **Swiss votes**, an irregular, bivariate vector image defined by two votations results at the canton level (figure 2 and 3)
- **the Portrait**, a regular, trivariate (levels of red green blue) image (figure 4)
- **the Geometer**, a regular, univariate (grey levels) image (figure 5)
- **Thun**, a geographical univariate image depicting the building density at the hectometre scale (census blocks) in the region of Thun, Switzerland (figure 6).
cluster dissimilarity, and $b$ the mean “nearest-cluster” dissimilarity (Rousseeuw, 1987).

Table 2 gives the silhouette coefficients $s$ for three groups (including the background group) under four segmentation algorithms, possibly converted to hard clustering after convergence, namely: hard K-means with random initialization of centroids, EM Gaussian Mixture modeling with diagonal covariance matrix (Pedregosa et al., 2011), our own iterative clustering algorithm (table 1) with $\alpha = 0$ and $\beta$ optimized so as to maximize $s$, and finally the iterative clustering algorithm (table 1) with the same value of $\beta$, but with $\alpha = 1$ in order to take into account the spatial contribution $L_z^2$. Spatial autocorrelation

Table 2: Segmentation evaluation: silhouette coefficients $s$ under four datasets and four clustering algorithms (see text). The iterative algorithm with $\alpha = 0$ yields the optimal values $\beta_0 = 1/0.0052$ (Swiss votes), $\beta_0 = 1/0.0052$ (Portrait), $\beta_0 = 1/0.00342$ (Geometer) and $\beta_0 = 1/0.0052$ (Thun), and should be equivalent to the EM algorithm. The lower performance of the iterative algorithm (table 1) for the Swiss votes dataset is unexpected, and under current investigation.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>$s_{K\text{-means}}$</th>
<th>$s_{EM}$</th>
<th>$s^{\beta_0=0}_{\alpha=0}$</th>
<th>$s^{\beta_0=1}_{\alpha=1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swiss votes</td>
<td>3</td>
<td>0.41</td>
<td>0.40</td>
<td>0.28</td>
</tr>
<tr>
<td>Portrait</td>
<td>3</td>
<td>0.57</td>
<td>0.57</td>
<td>0.56</td>
</tr>
<tr>
<td>Geometer</td>
<td>3</td>
<td>0.66</td>
<td>0.63</td>
<td>0.64</td>
</tr>
<tr>
<td>Thun</td>
<td>3</td>
<td>0.77</td>
<td>0.70</td>
<td>0.76</td>
</tr>
</tbody>
</table>

4 DISCUSSION

The present study constitutes a first attempt aimed to unify two little interacting domains – image segmentation and regional clustering – yet arguably very similar in their aims, and both relevant to spatial analysis. Spatial autocorrelation, in its general formulation (irregular, weighted networks) emerges as a common unifying paradigm, worth extending in the future.
Figure 5: the Geometer: (a) raster univariate image (grey levels $x_i$ range in the interval $[0, 255]$) of size $n = 50 \times 50 = 2500$ with initial strokes; (b) background group $g = 0$; (c) $g = 1$; (d) $g = 2$, and (e) $g = 3$, obtained after 60 iterations, with parameters $\beta = 1/0.003$ and $\alpha = 0.017$. Bottom: behavior of the free energy functional during the iteration, initially increasing.

Figure 6: Thun example. Top: raster univariate image of size $n = 20 \times 24 = 480$, depicting the population density per hectometer, rescaled in the interval of $[0, 27]$. Middle: groups $g = 0$ (blue, background), $g = 1$ (green) and $g = 2$ (purple), obtained after 50 iterations, with parameters $\beta = 1/0.003$ and $\alpha = 0.17$. Bottom: behavior of the free energy.

Figure 7: Moran’s $I$ and standardized test value $z$ for the four illustrations.

Figure 8: Swiss votes: Moran’s $I$ (top) and standardized $z$ normal test value (bottom) for $t \in [1, 50]$.

Alternative spatial or “Dirichlet-like” functionals $G[Z]$, such as the $N$-cut specification mentioned in section (2), or “$p$-Laplacians” (Bougleux et al.,... to local indicators (LISA) (Anselin, 1995). The proposed iterative approach can be considered as a generalization of the EM soft clustering for mixture models, containing an additional spatial term.
2009) should be considered, as is the relation to the Mumford-Shah functional (Chambolle et al., 2012) (Bar et al., 2011). Note that $G[Z]$ can be adjusted in two obvious, independent ways, namely by considering alternative weight-compatible specifications to $E(A, f)$, and by considering alternative forms $G(p)$ in (3).

Besides those energy-based approaches, the entropic part $\mathcal{K}[Z]$ can be introduced as a simple regularization term, or, as in the EM soft clustering, as originating from maximizing the log-likelihood of a statistical model (MAP approach), as in (Besag, 1986) or (Greig et al., 1989); see also (Couprie et al., 2011). Probabilistic formulations of the image segmentation problem should, as attested from spatial autocorrelation studies, benefit from results and experience in model-based clustering and latent modeling.

Alternatively, the virtue of random walk approaches, intimately connected to Markov chains and weighted networks, electric of not, has been attested in semi-supervised clustering in general, and image segmentation in particular (Grady, 2006). One key quantity is the probability that a random walk starting from a free pixel reaches first each of the $m$ tagged regions – a problem well-known to be related to the Dirichlet differential problem with suitable boundary conditions, and to the computation of electric potentials (Doyle and Snell, 1984).

Yet, in the probabilistic formulations of the multi-source, multi-target random walks (Guex, 2016), the capacities and the edge resistances turn out to be freely, separately adjustable (Bavaud and Guex, 2012). In the image segmentation context, it is tempting to identify the capacity contribution as a spatial term enabling transitions between neighbors, and the resistance contribution as a barrier preventing transitions between too dissimilar pixels; to which extent this line of research could prove itself innovative and efficient should be addressed in a near future.

REFERENCES


Bavaud, F. (2014). Spatial weights: constructing weight-compatible exchange matrices from proximity matri-


APPENDIX

Computing $E(A,f)$

Constructing an exchange matrix $E$ (section 2) both weight-compatible (that is obeying $E1 = f$, where the regional weights $f$ are given) and reflecting the spatial structure contained in the binary adjacency matrix $A = (a_{ij})$ is not trivial, nor that difficult either.

A natural attempt consists in determining a vector $c$ such that $e_{ij} = c_i a_{ij} c_j$. The problem can be essentially solved by Sinkhorn iterative fitting (Schneider and Zenios, 1990).

In this paper, we alternatively consider $A$ as the infinitesimal generator of a continuous Markov chain at time $t > 0$, which yields the diffusive specification (Bavaud, 2014)

$$E = E(A,f,t) = \Pi^{1/2} \exp(-t\Psi) \Pi^{1/2}$$

where $\Pi = \text{diag}(f)$, and

$$\Psi = \Pi^{-1/2} \frac{LA}{\text{trace}(LA)} \Pi^{-1/2}$$

$LA$ is the Laplacian of matrix $A$, and matrix exponentiation (9) can be performed by spectral decomposition of $\Psi$.

The resulting $E$ is semi-definite positive, with limits $E = \Pi$ for $t \to 0$ (diagonal spatial weights $W$, expressing complete spatial autarchy), and $E = \Pi f f'$ for $t \to \infty$ (constant spatial weights $W$, expressing complete mobility). Identity $\text{trace}(E(t)) = 1 - t + O(t^2)$ (Bavaud, 2014) shows $t$ to measure, for $t \ll 1$, the proportion of distinct regional pairs in the joint distribution $E$.

Testing spatial autocorrelation

Under the null hypothesis $H_0$ of absence of spatial autocorrelation, and under normal approximation, the expected value of the multivariate Moran’s $I$ reads (Bavaud, 2014)

$$E_0(I) = \frac{\text{tr}(W^2) - 1}{n - 1} \quad \text{where} \quad w_{ij} = \frac{e_{ij}}{f_i}$$

and its the variance reads

$$\text{Var}_0(I) = \frac{2}{n^2 - 1} \left[ \text{trace}(W^2) - 1 - \frac{(\text{trace}(W^2) - 1)^2}{n - 1} \right]$$

1. Spatial autocorrelation is thus significant at level $\alpha$ if $z := |I - E_0(I)|/\sqrt{\text{Var}_0(I)} \geq u_{1-\alpha}$, where $u_{1-\alpha}$ is the $\alpha$th quantile of the standard normal distribution.