Multiscale, Multigranular Image Segmentation

Eric D. Kolaczyk
kolaczyk@math.bu.edu

Dept of Mathematics and Statistics, Boston University
Credit Where Credit is Due

Collaborators:

- Sucharita Gopal (BU, Geography & CRS)
- Junchang Ju (BU, Geography & CRS)

Work supported in part by NSF grants BCS 0079077 and BCS 0318209, and ONR award N00014-99-1-0219.
Motivation: Land Cover Characterization

(a) Landsat Composite;
(b) Pixel-wise characterization
Scale versus Granularity

(a) (b)

Figure 1: $H = \text{hardwood}$, $C = \text{conifer}$, and $F = \text{forest}$
Image Segmentation

Goal: Partition an image region into sub-regions that are
1. homogeneous as individual elements;
2. meaningful as parts of the overall scene.
Takes on a classification aspect if goal includes labeling sub-regions.

Challenge: Equipping with a precise mathematical formulation a task influenced by cognitive aspects like vision and semantics.
MS-MG Image Segmentation

Contribution: A framework for producing labeled image segmentations from pixel-wise measurements that allows for simultaneous, adaptive choice of

(i) spatial resolution of sub-regions
(ii) categorical granularity of sub-region labels.

Response to aspect of UCGIS challenge in ‘scale’.
Summary of our Framework

Four Key Components:

1. Mixlets – A blending of recursive partitioning and mixture models.

2. Model selection through complexity-penalized likelihood.

3. Rule-based decision system for extraction of image segmentation.

4. Redundant (translation invariant) implementation.
Overview for Remainder of Talk

- Derivation of the mixlet-based framework.

- Evaluation of the framework:
  - Theoretical Evaluation: Horizon Model.
  - Empirical Evaluation:
    - Simulated landscapes
    - Sierra Data

- Discussion on open problems.
Some Notation

- Image region: \([0, 1]^2 \subset \mathbb{R}^2\)

- Pixelization: \(\{I_{i_1,i_2}\}^{n}_{i_1,i_2=1} \quad n = 2^J, J > 0\).

- Pure Pixels: \(c_{i_1,i_2} \in \{1, \ldots, C\}\)

- Pure-Class Densities: \(g(x|c)\)

- Measurements: \(x_{i_1,i_2} \in \mathbb{R}^k\)
Example

Region: Sierra, Nevada; 14.7 km$^2$

Pixelization: $128 \times 128 @ 30m$

Measurements: Landsat TM
6 Reflective Bands

Categories: Conifer, Hardwood, Brush, Grass

Pure-Class Densities: Trained as Mixtures of Gaussians.
Mixlets: Basic Idea

Combine

- Recursive dyadic partitions (quad-tree)
- Classes in a pre-specified categorical hierarchy
- Known or learned pure-class densities

and equip each sub-region of a given RDP with a mixture of densities specified by a given label in the hierarchy.
RDPs (Quad-Trees)

Recursive dyadic partitions (RDP) \( \mathcal{P} \) are partitions of \([0, 1]^2\) resulting from application of two rules:

(i) \( \mathcal{P}_0 \equiv [0, 1]^2 \), the trivial RDP, and

(ii) if \( \mathcal{P} \) an RDP, a quad-split of one \( R \in \mathcal{P} \) results in a new RDP \( \mathcal{P}' \) s.t. \( \mathcal{P} \prec \mathcal{P}' \).

Call \( \mathcal{P}_n^* \equiv \{I_{i_1,i_2}\} \) the C-RDP.
Categorical Hierarchies

Assume a pre-specified hierarchy (e.g., a tree) $\mathcal{T}_{Cat}$ of class labels $C(\mathcal{T}_{Cat}) \equiv \{\{c^{(d,l)}\}_{l=1}^{L_d}\}_{d=0}^{D}$, where $d$ denotes depth, from root to leaves, and $l$ denotes position within depth, from left to right.

$$\alpha_{d'}(c^{(d,l)}) \equiv \text{Ancestor of } c^{(d,l)} \text{ at depth } d' \leq d$$

$$\delta_{d'}(c^{(d,l)}) \equiv \text{All descendents of } c^{(d,l)} \text{ at depth } d' \geq d$$
Mixture Models

- Natural trade-off between spatial resolution versus categorical granularity.

- Coarser granularity implies multiple pure classes.
  E.g., \( \{H, C\} \rightarrow F \)

- Multiple classes suggests use of finite mixture models.
Definition: Mixlets

A mixlet model $\mathcal{M}$ is a triple $\{ \mathcal{P}, c(\mathcal{P}), \pi(\mathcal{P}) \}$, where

- $\mathcal{P} \subseteq \mathcal{P}_n$,
- $c(\mathcal{P}) \equiv \{ c(R) \}_{R \in \mathcal{P}}$,
- and $\pi(\mathcal{P}) \equiv \{ \pi(R) \}_{R \in \mathcal{P}}$,

along with densities $g(x|c)$, for $c \in \{ c^{(D,1)}, \ldots, c^{(D,L_D)} \}$, such that $x_{i_1,i_2} \sim f_{\mathcal{M}}(x|c_{i_1,i_2})$, $c_{i_1,i_2} = c(R(i_1,i_2))$, and

$$f_{\mathcal{M}}(x|c_{i_1,i_2}) = \sum_{c \in \delta_D(c_{i_1,i_2})} \pi_c(R(i_1,i_2)) \times g(x|c).$$
Identifiability

Each mixlet is a mixture of some allowable subset of densities \( g(x|c) \), for pure classes \( c \) in the leaves of \( \mathcal{T}_{Cat} \).

Define
\[
n(R) = |\{(i_1, i_2) : I_{i_1,i_2} \subseteq R\}|
\]
and
\[
l(R) = |\delta_D(c(R))|.
\]

Identifiability on each \( R \) follows from
- \( l(R) \leq n(R) \),
- ordering inherent in \( \mathcal{T}_{Cat} \),
- mild assumptions on \( g(x|c) \).
Related Literature

- Multiscale Segmentation
- Hierarchical Segmentation
- Multiscale Geometric Analysis
- Ecological Hierarchy Theory
Selection of Mixlet Model

A segmentation is based up an adaptively selected mixlet model:

$$\hat{M} \equiv \arg \max_M \{ \ell (x \mid M) - 2 \text{pen}(M) \} ,$$

where

- $$\ell(x \mid M) = \text{log-likelihood of } M$$
- $$\text{pen}(M) = \text{penalty function.}$$
Recalling that \( \mathcal{M} = \{ \mathcal{P}, c(\mathcal{P}), \pi(\mathcal{P}) \} \) we define

\[
pen(\mathcal{M}) = \left( \frac{4m}{3} \right) \log 2 + m \log L_C + \sum_{i=1}^{m} [l(R_i) - 1] \beta \log N ,
\]

where

- \( m = |\mathcal{P}| \),
- \( L_C = |\{ C(\mathcal{T}_{Cat}) \}| \),
- \( l(R_i) = |\delta_D(c(R_i))| \).

(Code length for uniquely decodable code for \( \mathcal{M} \in \mathcal{M}_N^*(\beta) \).)
Implementation

Log-likelihood and penalty are additive in components $R \in \mathcal{P}, \forall \mathcal{P} \in \mathcal{P}_n^*.$

- Optimization solvable using a bottom-up tree-pruning algorithm on the underlying quad-tree;
- Requires $\mathcal{O}(N)$ comparison stages;
- No more than $L_C$ mixture models fit in each stage.
Theoretical Evaluation: Horizon Model

We consider the performance of the estimator $\hat{M}$ in the context of a natural extension of the horizon model. (Korostelev and Tsybakov 1993; Donoho 1999)

Let $[0, 1]^2 = G \cup \bar{G}$, where $G \in \mathcal{G}$, and

$\mathcal{G} \equiv \{(t_1, t_2) \in [0, 1]^2 : t_1 \in [0, 1], 0 \leq t_2 \leq h(t_1), h \in \text{Lip}_A[0, 1]\}$

with

$h \in \text{Lip}_A[0, 1] \iff |h(x) - h(y)| \leq A|x - y|$.
Horizon Model (cont)

Let \( \pi : [0, 1]^2 \rightarrow [0, 1]^{L_D} \), such that

(i) \( \pi_c(\cdot) \) constant on each of \( G \) and \( \bar{G} \), \( \forall c \)

(ii) \( \pi_c \) not constant on \([0, 1]^2\) for all \( c \)

(iii) \( \pi_1(t) + \cdots + \pi_{L_D}(t) = 1, \forall t \in [0, 1]^2. \)

\[ \Rightarrow \] A constrained set of \( L_D \) standard horizon models.
Sample $x_{i_1,i_2}$ independently with respect to

$$f(x) = \sum_{c=1}^{L_D} \pi_c(I_{i_1,i_2}) \times g(x|c),$$

where $\pi_c(I_{i_1,i_2}) =$ Average of $\pi_c$ on $I_{i_1,i_2}$.

Let $L_{G,\bar{G}} =$ Number of non-zero $\pi_c$ on $G$ & $\bar{G}$. 
Risk Analysis: Background

Using Hellinger loss \( H^2(p, q) = \int \left( \sqrt{p(x)} - \sqrt{q(x)} \right)^2 dx \),
let risk in estimating \( f \) by \( \hat{f} \) be \( Risk(f, \hat{f}) = (1/N)E[H^2(f, \hat{f})] \).

Conduct search for \( \hat{M} \) over discretized model class \( \mathcal{M}^*_N(\beta) \),
with \( \beta = 1/2 \).

Assumptions:

- \( sup_{c,c',x} g(x|c)/g(x|c') < \infty \)
- \( L_D \leq \mathcal{O}(\log N) \)
Risk Analysis: Results

Let $\mathcal{F}$ be the class of all horizon models. Then for each $f \in \mathcal{F}$,

$$\text{Risk}(f, f_{\hat{\mathcal{M}}}) \leq L'_{G, \bar{G}} \times \mathcal{O} \left( (N^{-1} \log N)^{1/2} \right);$$

furthermore,

$$\sup_{f \in \mathcal{F}} \text{Risk}(f, f_{\hat{\mathcal{M}}}) \leq \mathcal{O} \left( (N^{-1} \log^3 N)^{1/2} \right).$$
Intuition

In the standard (univariate) horizon model:

- can cover a Lipschitz boundary $\partial G$ with $O(m)$ blocks of side-length $O(1/m)$;

- yields an $L_1$ error $\sim O(1/m)$, which must be balanced with a penalty $\sim O(m \log N/N)$.

$\Rightarrow$ Optimal $m \sim (N/\log N)^{1/2}$.

Criteria that $L_D \leq O(\log N)$ insures identifiability for any possible mixture models at this scale.
Extracting a Segmentation

Given $\hat{\mathcal{M}} = \{\hat{\mathcal{P}}, \hat{c}(\hat{\mathcal{P}}), \hat{\pi}(\hat{\mathcal{P}})\}$ a crude segmentation is provide by reporting

- Sub-regions $R \in \hat{\mathcal{P}}$, and
- Labels $\hat{c}(R)$.

But this approach can be too inclined towards declaring labels of coarser granularity e.g., contamination model. 
$\Rightarrow$ Suggests post-processing.
Refinement of Segmentation

Define

\[ \pi_c \equiv \sum_{c' \in \delta_D(c)} \pi_{c'} \quad \forall c \in T_{Cat}, \]

\[ \alpha(c), \delta(c) \]

be immediate ancestor and descendents of \( c \),

\[ \epsilon_1, \epsilon_2 \in [0, 1]. \]

Let \( \hat{c}_{i_1,i_2} \) be the first label \( c \), starting at \( \hat{c}(R(I_{i_1,i_2})) \) for which

(i) \( \hat{\pi}_c(R(i_1,i_2)) > \epsilon_1 \),

(ii) \( \hat{\pi}_c(R(i_1,i_2)) / \hat{\pi}_{\alpha(c)}(R(i_1,i_2)) > \epsilon_2 \),

(iii) the same cannot be said of any \( c' \in \delta(c) \).
Redundant Implementation

Recursive partitioning advantageous and common, but well-known in producing results that are

- ‘blocky’
- non-invariant to translation

Solution: Redundant/Cycle-Spinning Implementation.

- Combine models $\hat{M}^{(s_r, s_c)}$ over shifts $(s_r, s_c)$
- Requires only $O(N \log N)$ comparison stages.

(Details similar to those in wavelet literature.)
Empirical Evaluation

*Cartographic Generalization:* Data reduction and filtering of geographic information to a given scale and theme.

- Generally complex and time-consuming; often subjective.
- Our MS-MG image segmentation framework can be used to produce maps simultaneously sensitive to
  - multiple spatial resolutions,
  - multiple categorical granularities,
  - user priorities.
USFS interested in timber inventory and habitat management in the national forests of the western US.

- **Timber**: *Conifer*, such as ponderosa pine, Jeffrey pine, red/white fir, and mixtures thereof.

- **Habitat Management**: *Hardwood*, such as habitats for the spotted owl.

- **Difficulty**: *Brush*, such as chaparral, is difficult to distinguish from hardwood and considered a ‘nuisance’ class.
Simulation: Pure Sub-Regions
Pure Sub-Regions: Results
Simulation: Mixed Sub-Regions
Mixed Sub-Regions: Results
Sierra Data

- vegetation
  - forest
    - conifer
  - nonforest
    - hardwood brush grass
Sierra Data: Results

AsVh, scale 1x1

AsVh, scale 2x2

AsVh, scale 4x4

AsVh, scale 8x8

AsVh, scale 16x16

AsVh, scale 64x64
## Sierra Data: Changes Across Scale.

### Percentage of Each Landcover, By Scale

<table>
<thead>
<tr>
<th></th>
<th>1x1</th>
<th>2x2</th>
<th>4x4</th>
<th>8x8</th>
<th>16x16</th>
<th>32x32</th>
<th>64x64</th>
<th>128x128</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vegetation</td>
<td>0.00</td>
<td>0.11</td>
<td>0.19</td>
<td>0.28</td>
<td>0.37</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
</tr>
<tr>
<td>Forest</td>
<td>0.00</td>
<td>0.07</td>
<td>0.07</td>
<td>0.09</td>
<td>0.08</td>
<td>0.08</td>
<td>0.08</td>
<td>0.08</td>
</tr>
<tr>
<td>Conifer</td>
<td>0.42</td>
<td>0.40</td>
<td>0.39</td>
<td>0.38</td>
<td>0.36</td>
<td>0.36</td>
<td>0.36</td>
<td>0.36</td>
</tr>
<tr>
<td>Hardwood</td>
<td>0.24</td>
<td>0.12</td>
<td>0.10</td>
<td>0.09</td>
<td>0.08</td>
<td>0.08</td>
<td>0.08</td>
<td>0.08</td>
</tr>
<tr>
<td>Non-forest</td>
<td>0.00</td>
<td>0.02</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Brush</td>
<td>0.31</td>
<td>0.27</td>
<td>0.23</td>
<td>0.16</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>Grass</td>
<td>0.03</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

INRIA, 11 July 2005 – p. 38/40
Discussion

Mixlet framework is intended to be an initial, prototype solution to a complex problem.

Two current directions of our work are

- development of a formal Bayesian analogue
- extension to temporal-indexed image segmentation
Bayesian Analogue

Write mixlet models as $\mathcal{M} = (\mathcal{L}, \pi)$, where $\mathcal{L} = (\mathcal{P}, c(\mathcal{P}))$.

- Model $\Pr(\mathcal{L})$ using adaptations of the spatial random tree grammars of Pollak, Bouman, and colleagues.

- Choose a segmentation according to $\hat{\mathcal{L}} = \arg\min_{\mathcal{L}} \Pr(\mathcal{L}|x)$.

(Joint work with Huang, Gopal, and Ju.)