A cognitive perspective on aerial image interpretation

Michael E. Hodgson
University of South Carolina

Amy L. Griffin
University of New South Wales Canberra
### Motivations / context

**USGS Land Use/Cover Classification System**

#### Level I

<table>
<thead>
<tr>
<th>Level I</th>
<th>Level II</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Urban or Built-up Land</td>
<td>11 Residential</td>
</tr>
<tr>
<td></td>
<td>12 Commercial and Services</td>
</tr>
<tr>
<td></td>
<td>13 Industrial</td>
</tr>
<tr>
<td></td>
<td>14 Transportation, Communications, and Utilities</td>
</tr>
<tr>
<td></td>
<td>15 Industrial and Commercial Complexes</td>
</tr>
<tr>
<td></td>
<td>16 Mixed Urban or Built-up Land</td>
</tr>
<tr>
<td></td>
<td>17 Other Urban or Built-up Land</td>
</tr>
<tr>
<td>2 Agricultural Land</td>
<td>21 Cropland and Pasture</td>
</tr>
<tr>
<td></td>
<td>22 Orchards, Groves, Vineyards, Nurseries, and Ornamental Horticultural Areas</td>
</tr>
</tbody>
</table>
We want to know what information from the image is necessary to accurately/efficiently identify areas.
Sample Research Questions

Cognitive Process

Classical Remote Sensing Interpretation Key

A manufacturing facility may contain raw materials, storage of finished products, shipping lines, etc.

A high school will contain a football stadium, parking lot for student cars, etc.

- Is the analyst working from general to specific? (or vice-versa)
- What sub-objects are identified first, second, etc?
Research Goals

- **Pragmatic Goal:** Design algorithms that …
  - Producing the same results as humans
  - Producing the same results by the same process

- **Science Goal:** Understanding the cognitive process
  - Ultimately for a pragmatic goal (designing algorithms)
  - Pedagogical (e.g., training students, analysts)
– Our ‘imagery’ is of a **natural scene** rather than a **generalization** in graphic form.

(1) Our scene is similar to what humans process continually, except our scene is **viewed from above** and at **distant** objects.

(2) The details of the image (generalized out in a map) may contain important information.
Task-oriented versus free viewing

We have a specific task & we want to understand task-oriented or task-directed cognitive processes.
Why eye-tracking?

• Helps to understand **why** a performance is good or poor:
  – Did the viewer use **contextual** information (that information surrounding the target)?
  – Do viewers need to find one piece of information **before** another can be useful?
  – To what extent do different image characteristics alter/influence **visual search** (bottom-up vision)?
Visual interpretation of aerial imagery is not simple

Do visual search and cognitive processes vary with...

.. experience?
.. gender?
.. age?
.. spatial resolution?
.. spectral band/composition (e.g. panchromatic, natural color, CIR)?
.. geography?
.. classification system (USGS, NWI, etc.)?
.. classification level (i.e., specificity)?

.. ... all of these are factors that could be examined.
Small Exploratory Study

12 remote sensing experts @ AAG meeting in 2013
(university faculty/PhD students teaching remote sensing)
24 panchromatic photos in piedmont South Carolina
(all same spatial resolution and size of geographic area)

General instruction:
You should try to identify the land use/land cover target as quickly but as accurately as you can.

Task:
What Anderson level II class does the target location have?
Experimental setup and details

• Tobii x120 eye-tracker recording at 120 Hz
• Conducted in a hotel meeting room (not ideal!)
• Within-subjects design; stimuli Latin squares presentation of stimuli
• Independent variables:
  – Land use/land cover classes
  – Experience teaching air photo interpretation
• Dependent variables:
  – Answers (verbalized)
  – Answer time (seconds)
  – Eye tracking metrics
• Also asked what strategies they used to identify targets (at end of the experiment).
Some Preliminary Results

- Fast and Accurate
- Slow and less accurate
- Fast and inaccurate
Next steps with our data

• We are still trying to figure out the most appropriate parameterization for the Tobii fixation filter (velocity based).
  – Does generally a better job of identifying fixations than other options.

• To what extent is the **context** surrounding the target important for correct identification of the LULC class?
  – Is context more important for ‘difficult-to-identify’ classes?

• What **distinguishes/differentiates good from poor performance**, especially for ‘difficult-to-identify’ classes?
  – Viewing strategy?
  – Do ‘correct’ performances replicate what used to be taught in classification keys?
  – What critical image details do inaccurate performances miss?

• Is the **order of fixations** (i.e. what image sub-objects are viewed and when) related to effective and efficient performances?