Harvesting Named Geographic Features from Raster Maps

Craig Knoblock and Yao-Yi Chiang
University of Southern California & Geoseemble Technologies
Introduction

- Raster maps are a rich source of geospatial data:
  - Easily accessible
  - Many different types of information
  - Often contains information that cannot be found elsewhere
Challenges

• Maps have lots of useful information, but...
  – They have overlapping features
  – There is limited access to the meta-data
  – Often only available in raster format
• How do we register, extract, and recognize the features in a raster map
Extracting Named Road Vector Data

- Raster Maps
  - Map Registration
  - Road Vectorization
  - Text Recognition
  - Text
  - Text/Road Association
  - Named Road Vector Data
Automatic Map Registration

• Exploit the pattern of intersections found on a map and compare to a road vector dataset
Road Vectorization

Extracting road layers

Raster road layer

Extracting accurate road intersections and tracing road lines

Road vector data
### Extracting text layers

<table>
<thead>
<tr>
<th>Name</th>
<th>Lat./Long.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hillsdale</td>
<td>38.68456, -90.2873</td>
</tr>
<tr>
<td>Curtis Av</td>
<td>38.68653, -90.2907</td>
</tr>
<tr>
<td>Archie Moore Av</td>
<td>38.68484, -90.2901</td>
</tr>
<tr>
<td>Oakdale</td>
<td>38.68538, -90.28511</td>
</tr>
</tbody>
</table>

- **Commercial OCR Product**
- **Run OCR on strings**
- **Build database**

- **Identified text strings**
- **Rotate individual strings to horizontal**
Text/Road Association

Red crosses indicate the start/end point of a line segment

Raw road vector data

Identifying street segments

<table>
<thead>
<tr>
<th>Name</th>
<th>Lat./Long.</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAGE BLVD</td>
<td>38.52356, -90.1283</td>
</tr>
<tr>
<td>Billups AV</td>
<td>38.52483, -90.1290</td>
</tr>
<tr>
<td>Cozens</td>
<td>38.52424, -90.1291</td>
</tr>
<tr>
<td>Evans Av</td>
<td>38.52267, -90.1281</td>
</tr>
<tr>
<td>...</td>
<td>Recognized text</td>
</tr>
</tbody>
</table>

Determine the association between text and street segments using
1. Text and street positions,
2. Text and street Orientations,
3. Road topology
Example Map
Resulting Vector Layer Displayed in ArcGIS
Discussion

• Developed a general approach to registering, extracting, and recognizing features in maps
  – Applies to a wide variety of maps
  – Supports automatic registration of maps
  – Build geospatial layers from poor quality scanned maps

• Applications
  – Annotating imagery
    • Extracted text labels can be used to label imagery
  – Creating and updating maps
    • Provides the ability to create and update maps for regions where we don’t have good map data
  – Constructing gazetteers
    • Construct geographic name databases
    • Critical for linking text (e.g., news, blogs, tweets, etc) to geographic locations
      – See demonstration of Geosemble’s GeoXray at booth 103
More Information

• See our papers:
  – [http://www.isi.edu/~knoblock](http://www.isi.edu/~knoblock)

• Send me email:
  – knoblock@isi.edu

• Come to the Geosemble booth
  – Booth 103