CS 5320 Project

Raster Map Image Analysis

Project Leader: Thomas C. Henderson

Final Project: Due: 13 December 2012

Use the best approach that you can find to classify the following semantic categories from the q<…> maps:

1. Waterways (lakes, rivers, streams, etc.)
2. Roads (various subcategories)
3. Iso-contour lines
4. Vegetation cover
5. Text
6. Political Lines
7. Urban Areas (including buildings, etc.)
8. Various textures (e.g., marsh land)

Write this up as a lab report, and handin the results to cs5320 PF by 5pm on the 13 December 2012.

Week 11: November 6 - 20 2012 P4

Extract and fill in iso-contours from the raster maps named q<…>. Given one of these maps, extract the iso-contours as best as possible using techniques you've learned or things you can find in the literature. This means finding the best connected components comparable to what a human would produce. This includes the following steps:

1. Extract most likely iso-contour pixels. Try method considered in the course, including, the use the color histograms, texture, linear features, etc. combined with various learning methods (e.g., K-means).

2. Connect up reasonable sets of connected components to form closed curves. Try the Delaunay triangulation. [For a set of points P, the Delaunay triangulation, DT(P), has the property that no point p in P is inside any circumcircle of any triangle. This is the dual of the Voronoi diagram which is a set of regions for P where a resgion is a set of points closest to a point in P. Look at paper: "Contour Line Recognition From Scanned Topographic Maps," Salvatore and Guitton, WSCG 2004]
3. Find methods to delete edges from DT(P) that are bad (i.e., do not connect an iso-contour).
4. Measure the performance qualitatively on the q images.

Write up a lab report describing your work. You should handin the PDF by Tuesday 20 November 2012.

Week 6: September 25 - 27 2012 P3

Use k-means to segment the following classes from a map: (1) roads, (2) water, (3) iso-contours, (4) political lines, (5) vegetation. Do the following:

1. use the color histogram only (experiment with window size)
2. use texture parameters
3. test on the 10 sub-images given from the large map (200x200 sub-images).
4. Given the ground truth provided for the sub-images, calculate precision and recall statistics.

Write up a lab report describing your work. You should handin the PDF by Thursday 18 Oct 2012.

Weeks 3&4: September 4-13 2012 P2

Develop a method for the segmentation of the linear features (e.g., roads, rivers, political boundaries and iso-contours) using edge detection techniques. Write up a lab report describing your work. You should email the PDF to me on Thursday 20 Sept 2012.

Week 2: August 28-30 2012 P1

Develop a method for the segmentation of the true map area from the map surround (e.g., white or black areas, the legend, etc.). Write up a lab report describing your work.
You should email the PDF to me on Thursday 30 Aug 2012.

Week 1: August 21-23 2012

Goal

The semantic interpretation of raster map images is a very difficult problem. A major problem is that there is a dearth of work relating high-level models of documents, drawings or graphics to the various levels of analysis. The goal of raster map image analysis is to interpret the contents of an image, such as that shown here (part of a USGS digital map image - see f34086a1.tif in the class data directory):

Several systems have been developed along these lines, although none works well in an automated fashion. Various problems arise; e.g., when text touches other classes or when there is noise in the scanned image or when the thresholds of the image-analysis codes support only a few classes of drawings.

The goal of this project is the extraction of semantic content from raster map images which may be exploited in the anaylsis of aerial imagery. The major map features which need to be extracted include: roads and road insections, texture, text, and miscellaneous symbols.

Where appropriate these may be vectorized, and the image coordinates may be recorded as well. Metadata may be provided including the type of map (e.g., USGS DRG), any specific color encoding of features (e.g., USGS maps may use from 6 to 13 specific colors), geo-spatial information, and names of features may be known and sought in the image (e.g., rivers, lakes, roads, towns, etc.).

The specific goals here are to segment:

- Background
  - Knowledgebase for color use in known map types
  - Color usage probes: interactive and color neighborhood analysis
  - Train on CADRG, DRG, scanned maps and other sources to better understand distributions

- Roads and Road Intersections
  - Road segmentation
  - Road type determination
  - Vectorization
  - Subpixel road tracking
  - Double line roads
  - Road name and/or number
  - Intersections
  - Geo-reference for intersections
- Texture
  - Knowledgebase for contraints on texture appearance based on map type
  - Long-term analysis to discover map textures and populate the knowledgebase (e.g., 2-D statistics, background of map legend
- Text
  - Higher reliability of text extraction
  - Recognition of characters
  - Recognition of words
  - Match to metadata
- miscellaneous Features
  - Highway Markers
    - Interstate Markers
    - US Markers
    - State Highway Markers
  - Other (e.g., noise pixels)

Map feature characteristics may also be classified according to whether or not they correspond to a visible feature in the image. Using this criterion, we have the following sets:

Map features visible in an aerial image

Roads and Road Intersections

Water

Buildings

Various textures: lava, marshland, desert, orchards, etc.

Map Features which provide meta-data

Text

Lattitude, Longitude

Political Boundaries

Scale

The algorithms should eventually apply to a wide range of map types: USGS DRG, NGA, and other miscellaneous types. A dataset of maps has been collected and these are described below.

Tests are run to determine:

- Time cost per pixel
- Recall rate (number of relevant items retrieved divided by the number of relevant items)
- Precision rate (number of relevant items retrieved divided by the number of retrieved items)
- Overall quality of segmentation and interpretation.

Schedule

Week 1: Define Project

Week 2: Develop basic map analysis tools

Week 3: Explore linear and point features in maps

Week 4: Develop line segment model

Week 5: Explore texture in maps

Week 6: Define performance measures

Develop benchmark dataset and ground truth

Week 7: Implement first cut segmentation system

Week 8: Fall Break

Week 9: Measure performance of map analysis

Week 10: Compare segmentation methods

Week 11: Explore classification methods

Week 12: Explore complete system sensitivity analysis

Week 13: Explore further classification methods

```
Week 14:  Measure system performance; optimize

Week 15:  Final results and report
```

Bibliography