Automatic Segmentation of Semantic Classes in Raster Map Images

Thomas C. Henderson, Trevor Linton, Sergey Potupchik and Andrei Ostanin

School of Computing, University of Utah, Salt Lake City, UT 84112 USA E-mail: tch@cs.utah.edu

Abstract

The automatic classification of semantic classes (background, vegetation, roads, water, political boundaries, iso-contours) in raster map images still poses significant challenges. We describe and compare the results of three unsupervised classification algorithms: (1) k-means, (2) graph theoretic (GT), and (3) expectation maximization (EM). These are applied to USGS raster map images, and performance is measured in terms of the recall and precision as well as the cluster quality on a set of map images for which the ground truth is available. Across the six classes studied here, k-means achieves good clusters and an average of 78% recall and 70% precision; GT clustering achieves good clusters and 83% recall with 74% precision. Finally, EM forms very good clusters and has an average 86% recall and 71% precision.

Keywords: Graphics Recognition, Raster Map Images, Segmentation.

1 Introduction

Digital maps contain a wealth of information which can be used for a variety of applications, including the analysis of cultural features, topographical terrain shape, land use classes, transportation networks, or maps can be registered (conflated) with aerial images in order to localize and identify photo imagery structures. Unfortunately, raster map images are typically encoded in such a way that semantic features are difficult to extract due to noise, error or overlapping features. Semantic features of interest include roads, road intersections, water regions, vegetation, political boundaries, and iso-elevation contours. This is still a difficult problem, although various techniques have been proposed in the past[1, 2, 6]. We have worked on road segmentation and road intersection detection [4, 5].

Our goal is to achieve a semantic segmentation of an arbitrary raster map image through the use of unsupervised classification algorithms. An example USGS map sub-image is shown in Figure 1. We are interested in six basic classes:

- Background
- Vegetation
- Roads
- Water
- Political Lines
- Iso-contours

Figure 2 shows the ground truth for these classes for the map in Figure 1.



Figure 1: Example USGS Map Sub-image (200x200).

2 Method

The ground truth was determined using a knowledge-based analysis of a set of sub-images (200x200 pixels) taken from ten USGS map images. (Appendix A shows the ten test images taken from the USGS maps.) These maps use six colors (*black*, *white*, *blue*, *red*, *brown*, *green*), and are given as indexed images (i.e., the colors have indexes 0, 1, 2, 3, 4, 5). The classification analysis process is shown in Figure 3. The index histogram is based on a $w \times w$ window at each pixel. The cluster centers are the representative histogram for a class, and the covariance matrix gives the variation between the colors for that class. These models are found by using a subset of *n* samples from the index histogram image. The number of classes may be pre-defined (as with *k*-means) or determined automatically by the method (e.g., GT). Thus, the parameters of study across the three algorithms are *w*, the histogram window size, *n*, the number of samples used to construct the model, and *k*, the number of classes sought.

The quality measures for the class models are defined in terms of:

- the cluster inter-center distances where, in general, a greater value is better, and
- the distances of points in the cluster from the center where a smaller value is better.

As for the quality of the classification result, recall and precision are defined as:

$$recall = \frac{|relevant \cap retrieved|}{|relevant|}$$

$$precision = \frac{|relevant \cap retrieved|}{|retrieved|}$$

where *relevant* is the set of ground truth pixels in a class and *retrieved* is the set of pixels segmented into that class by the algorithm. The general layout of the classification process is:



Figure 2: Classes from Example Image.

```
for each test image
  for each w in Window_sizes
    for each n in Sample_sizes
      for each k in Number_of_classes
        Obtain class centers (means and covariances)
        Compute class quality
        Compute recall and precision
        end
    end
end
```

Classification Test Process

```
Compute statistics over all test images
```

Algorithm:

The algorithms under study include k-means, GT and EM. k-means initially selects k random centers, then alternates between assigning points (i.e., histogram vectors) to the nearest center and calculating the centers as the means of the points in the cluster. The graph theoretic method forms an affinity measure between all sample points (e.g., $exp^{-|p_1-p_2|}$), then obtains the eigenvalues and eigenvectors of that matrix; finally, the eigenvectors serve to classify pixels in each class. The EM algorithm alternates between the expectation calculation step and the maximization step to determine the set of classes. See [3] for more details on these three methods.

The centers and covariances are found for each classification algorithm by computing the mean of the sample points segmented into a class, and the covariance of those points. Although these are produced directly by k-means and EM, this is done after the fact for GT based on the set of points in the sample.

The map image is classified by simply labeling each pixel according to the closest center to the pixel's index



Figure 3: Segmentation Analysis Process.

color histogram. Note that the method cannot know which, if any, of its discovered classes correspond to ground truth classes. Therefore, we determine the recall and precision by mapping each discovered class to the nearest (Euclidean distance) ground truth class mean histogram vector.

3 Data

Here we give the results of the Algorithm Classification Test on the three algorithms. The possible values for the parameters were:

• *k*-means:

$$w \in \{1, 3, 5\}$$

 $n \in \{1000, 2000, 3000\}$
 $k \in \{6, 8, 10\}$

• Expectation Maximization (EM)

$$w \in \{1, 3, 5\}$$
$$n \in \{1000, 2000, 3000\}$$
$$k \in \{6, 8, 10\}$$

• Graph Theoretic

```
w \in \{1, 3, 5\}n \in \{25, 50, 75\}s \in \{0.1, 10, 20\}
```

These w values correspond to a single pixel (w = 1) up to a window that almost always includes background with any linear feature. The values of n range from about 25% of linear features in an average 200 × 200 sub-image, up to the full number of linear features in a typical sub-image. Of course, there is no guarantee that pixels in a linear feature will be selected as samples. The number of classes of interest is six; however, not

w	n	k	mean recall	mean precision	recall + precision
1	2000	8	0.78	0.70	1.48
1	3000	10	0.78	0.70	1.48
1	2000	10	0.77	0.70	1.47
1	1000	10	0.78	0.69	1.47
1	3000	6	0.79	0.67	1.46
1	1000	6	0.80	0.66	1.46
1	3000	8	0.78	0.68	1.46
1	2000	6	0.80	0.66	1.46
1	1000	8	0.77	0.68	1.45
3	1000	10	0.59	0.54	1.13
3	2000	10	0.58	0.53	1.12
3	2000	8	0.59	0.52	1.11
3	3000	10	0.58	0.53	1.11
3	1000	8	0.59	0.52	1.11
3	3000	6	0.60	0.49	1.09
3	1000	6	0.60	0.49	1.09
3	2000	6	0.60	0.49	1.09
3	3000	8	0.58	0.51	1.09
5	2000	10	0.49	0.45	0.93
5	3000	6	0.52	0.42	0.93
5	1000	10	0.49	0.45	0.93
5	2000	8	0.49	0.44	0.93
5	3000	10	0.48	0.44	0.93
5	1000	6	0.50	0.42	0.92
5	1000	8	0.49	0.43	0.92
5	2000	6	0.50	0.41	0.91
5	3000	8	0.48	0.42	0.91

Table 1: *k*-means Ranked Parameter Combinations (first 3 values per row), followed by average recall (over all classes all images), average precision, and sum of average recall and precision.

all classes may be present in a sub-image; moreover, pixels at the boundary of two classes actually represent a different class (e.g., vegetation-water boundary). Finally, the *s* value is a distance scaling measure in the graph theoretic method which controls the scale of the affinity.

There are 27 combinations of w, n, and k/s values. Tables 1 through 3 give the parameters of the top performing combinations and the recall and precision values averaged over all classes and all images. Figure 1 shows an example raster map image of size 200×200 , while Figure 2 shows the ground truth for this image. Figure 4 shows the classes found by *k*-means; Figure 5 shows the graph theoretic classes, and Figure 6 shows the EM classes.

4 Conclusions and Future Work

The results show that the three clustering methods perform well for unsupervised raster map image classification. Moreover, the optimal parameters all have the window size set to 1×1 (a single pixel). However, the models developed do a little better than simply classifying each pixel based on its color which achieves recall of 80% and precision of 72% for a sum of 1.52; this is worse than graph theoretical and EM, but better than k means.

w	n	s	mean recall	mean precision	recall + precision
1	50.0000	0.1000	0.8282	0.7414	1.5697
1	25.0000	0.1000	0.8282	0.7414	1.5696
1	75.0000	0.1000	0.8282	0.7414	1.5696
3	25.0000	0.1000	0.8203	0.6895	1.5098
3	50.0000	0.1000	0.7786	0.6798	1.4585
3	75.0000	0.1000	0.7511	0.6765	1.4276
1	75.0000	20.0000	1.0000	0.2662	1.2662
5	25.0000	0.1000	0.6680	0.5798	1.2478
1	75.0000	10.0000	1.0000	0.2330	1.2330
5	25.0000	10.0000	0.6439	0.5890	1.2329
5	75.0000	10.0000	0.6375	0.5921	1.2295
1	50.0000	20.0000	1.0000	0.2228	1.2228
5	50.0000	10.0000	0.6343	0.5821	1.2164
1	50.0000	10.0000	1.0000	0.1998	1.1998
3	75.0000	20.0000	1.0000	0.1908	1.1908
1	25.0000	10.0000	1.0000	0.1897	1.1897
1	25.0000	20.0000	1.0000	0.1803	1.1803
3	50.0000	20.0000	1.0000	0.1793	1.1793
3	25.0000	20.0000	1.0000	0.1788	1.1788
5	50.0000	0.1000	0.6006	0.5613	1.1619
5	50.0000	20.0000	0.5122	0.6280	1.1402
5	75.0000	20.0000	0.4758	0.6354	1.1112
3	75.0000	10.0000	0.4547	0.6468	1.1016
5	75.0000	0.1000	0.5483	0.5504	1.0987
3	50.0000	10.0000	0.4402	0.6480	1.0881
3	25.0000	10.0000	0.4144	0.6662	1.0807
5	25.0000	20.0000	0.4384	0.6330	1.0713

Table 2: Graph Theoretic Ranked Parameter Combinations (first 3 values per row), followed by average recall (over all classes all images), average precision, and sum of average recall and precision.

w	n	k	mean recall	mean precision	recall + precision
1	3000	8	0.8644	0.7054	1.5697
1	2000	8	0.8681	0.6850	1.5532
1	3000	10	0.8515	0.7011	1.5526
1	1000	10	0.8482	0.7023	1.5505
1	2000	10	0.8503	0.7001	1.5504
1	1000	8	0.8549	0.6897	1.5446
1	1000	6	0.8710	0.6719	1.5429
1	2000	6	0.8784	0.6637	1.5421
1	3000	6	0.8764	0.6628	1.5392
3	2000	8	0.9550	0.2114	1.1664
3	3000	10	0.9519	0.2123	1.1642
3	2000	10	0.9518	0.2120	1.1638
3	1000	8	0.9540	0.2093	1.1633
3	3000	8	0.9568	0.2063	1.1630
3	1000	6	0.9577	0.2034	1.1611
3	1000	10	0.9509	0.2081	1.1589
3	3000	6	0.9599	0.1986	1.1584
3	2000	6	0.9599	0.1948	1.1547
5	1000	6	1.0000	0.0171	1.0171
5	1000	8	1.0000	0.0171	1.0171
5	2000	6	1.0000	0.0171	1.0171
5	3000	6	1.0000	0.0171	1.0171
5	3000	8	1.0000	0.0171	1.0171
5	2000	8	0.9716	0.0176	0.9892
5	3000	10	0.9541	0.0187	0.9728
5	1000	10	0.9455	0.0186	0.9641
5	2000	10	0.9408	0.0188	0.9596

Table 3: Expectation Maximization (EM) Ranked Parameter Combinations (first 3 values per row), followed by average recall (over all classes all images), average precision, and sum of average recall and precision.



Figure 4: k-means Segmentation Results.

The fact that a small number of samples can be used is also good; the graph theoretic method must calculate the eigenvalues of an $n \times n$ affinity matrix, and thus, the lower n, the better.

Of course, these are relatively simple raster map images with only six colors. It is necessary to study these methods on map images with more colors. This will increase the length of the histogram vectors unless some form of color clustering is performed first to reduce the number of color classes. This may require conversion to a color representation with a reasonable distance metric between colors (i.e., where various types of blue are close in the metric space).

Another issue worthy of study is a more informed method to select samples. It may be worthwhile to ensure that samples represent the variety of classes in the image (as opposed to the standard sampling goal of proportional representation of the sampled population). It may be possible to use edge detection to distinguish class boundary pixels or texture parameters to determine classes expressed as textures. Of course, edge and texture information may be included in the feature vector (in addition to the color histogram).

Finally, all these classification methods have a variety of possibilities in algorithm implementation. Initialization methods, re-starting empty classes, thresholds, and distance measures all offer a number of options which should be studied.

5 Acknowledgment

This work was sponsored in part by AFOSR grant FA9550-08-C-0005 in cooperation with IAVO Corporation. We would like to thank Eric Lester and Brad Grinstead for their support.

A Test Images

Figure 7 shows the ten test images used in the study.



Figure 5: GT Segmentation Results.

References

- [1] E. Ageenko and A. Podlasov. On the Restoration of Semantic Features in Raster Topographic Images. *IADIS International Journal on COmputer Science and Information Systems*, 1(1):101–114, 2006.
- [2] Y.-Y. Chiang and C.A. Knoblock. Classification of Line and Character Pixels on Raster Maps Using Discrete Cosine Transformation Coefficients and Support Vector Machines. In *Proceedings of the 18th International Conference on Pattern Recognition*, August 2006.
- [3] D. Forsyth and J. Ponce. Computer Vision. Prentice Hall, Upper Saddle River, NJ, 2003.
- [4] T. Henderson and T. Linton. Raster Map Image Analysis. In Proceedings of the 10th International Conference on Document Analysis and Recognition, Barcelona, Spain, July 2009.
- [5] Trevor Linton. Semantic Feature Analysis in Raster Maps. Master's thesis, University of Utah, Salt Lake City, Utah, June 2009.
- [6] A. Podlasov, E. Ageenko, and P. Fränti. Morphological Reconstruction of Semantic Layers in Map Images. *Journal of Electronic Imaging*, 15(1):013016–1–013016–10, Jan-Mar 2006.



Figure 6: EM Segmentation Results.



Figure 7: Test Images.