2009 Ninth International Conference on Intelligent Systems Design and Applications

A Comparative Study of Clustering Methods for Urban Areas Segmentation from High Resolution Remote Sensing Image

Safaa M. Bedawi

Department of Electrical and Computer Engineering, University of Waterloo

sbedawi@pami.uawterloo.ca

Mohamed S. Kamel

Department of Electrical and Computer Engineering, University of Waterloo

mkamel@pami.uwaterloo.ca

Abstract

This paper focuses on evaluating and comparing a number of clustering methods used in color image segmentation of high resolution remote sensing images. Despite the enormous progress in the analysis of remote sensing imagery over the past three decades, there is a lack of guidance on how to select an image segmentation method suitable for the image type and size. Clustering has been widely used as a segmentation approach therefore, choosing an appropriate clustering method is very critical to achieve better results. In this paper we compare five clustering methods that have been suggested for segmentation of images. We focus on segmentation of urban areas in high resolution remote sensing images. Effective clustering extracts regions which correspond to land uses in urban areas. Ground truth images are used to evaluate the performance of clustering methods. The comparison shows that the average accuracy of road extraction is above 75%. The results show the potential of clustering high resolution aerial images starting from the three RGB bands only. The comparison gives some guidance and tradeoffs involved in using each.

1. Introduction

Aerial and satellite images (geospatial images) are increasingly valuable sources of information in diverse fields such as cartography, urban planning, environment monitoring, civil and military intelligence, etc. Remote sensing (RS) images contain visual information about various land use features. Manual annotation of geospatial images covering even a relatively small area of the earth is a tedious task.

With the advent of high resolution (HR) RS imagery, 41 cm have become commercially available by the launching of the Geo Eye-1 satellite in 2008. New challenges did arise for the classification of land use/cover in urban areas. Although there is no exact definition for urbanization, most of the previous work characterizes these areas using the density of buildings. HR remote sensing images contain complicated spectral and texture characters of heterogeneous urban scenes which make the accurate discrimination of distinct thematic classes a difficult task. Compared to coarser-resolution images, the ground materials that are present in the imaged scene can be better appreciated in HR imagery. They may include concrete, asphalt, metal, water, grass, trees, bare soil, etc. The presence of such different material responses gives rise to additional problems in terms of information extraction. A finer resolution yields a decrease in the number of mixed pixels, increasing the discrimination accuracy. On the other hand, the higher the resolution, the larger the number of subclasses based on the spatial distribution of material responses, which affects the class discrimination accuracy (internal class variability) [1]. It seems evident that HR images create additional challenges in terms of information extraction and classification as a result, existing approaches are not suitable to HRRS imagery [2].

Research on segmenting this kind of images is necessary to improve the processing ability of RS images. The segmentation quality has a direct influence on the subsequent image analysis and understanding spaces. Image segmentation techniques [3] automatically group neighboring pixels into non-overlapping meaningful regions based on similarity criteria of pixel properties such as color, texture, shape and size. These techniques can be divided into several categories: pixel based [4], region based [5], edge

978-0-7695-3872-3/09 $26.00 © 2009 IEEE
DOI 10.1109/ISDA.2009.109

based [6], hybrid segmentation [6] and clustering-based segmentation [8]. Although image segmentation has been heavily studied in image processing and computer vision fields, no single method is suitable for all types of data. Segmentation algorithms have only recently started receiving emphasis in RS image analysis. Examples of image segmentation in the RS literature can be found in [5][6][9]. However, they are not applied to very high resolution images. Other studies have compared the performance of some techniques, but they were applied to coarse/medium resolution images [1] [10].

This study aims to identify an appropriate clustering method to apply to HRRS images to cluster pixels into regions corresponding to land uses in urban areas. Accordingly, five clustering based segmentation techniques are evaluated and compared. Two of them are popular clustering based image segmentation techniques namely k-means [11] and mean-shift [7] methods. The other three methods, Gaussian mixture model [11], spectral clustering [8][14] and affinity propagation [15], are state-of-the-art techniques that have proven to result in sound performance for other types of images. Clusters are mapped to ground truth classes as described in [11]. We evaluate the algorithms empirically [16] by measuring the quality of segmentation compared with a reference segmentation.

The rest of paper is organized as follows. Section 2 briefly describes the clustering methods used in the comparative study. The experimental results for color aerial images are given in section 3. Finally the conclusions are drawn in section 4.

2. Clustering Methods

Clustering methods are all based on a measure of similarity of data points. A common approach is to cluster data points by iteratively calculating the similarity or some other measurements until termination conditions are satisfied. In the following, a brief overview is given for the five clustering methods used in our comparison.

2.1. K-Means method

K-means algorithm [11] is one of the most popular and simplest clustering methods. The algorithm is composed of the following steps:

Input: Data points $x_1 \dots x_n$; k: number of clusters
1. Select k random points to be the initial clusters centers.
2. Assign each point to the cluster that has the closest center.
3. When all points have been assigned, recalculate the positions of the k centers.
4. Repeat Steps 2 and 3 until the centers are stable.

The result is a set of clusters that are as compact and well separated as possible. Although it can be proved that the procedure will always terminate, the k-means algorithm does not necessarily find the most optimal configuration, corresponding to the global objective function minimum.

2.2. Gaussian mixture model

Gaussian Mixture Model (GMM) [12] is a type of density model which comprises a number of component functions, usually Gaussian. Mixture models are fit to data using the expectation maximization (EM) algorithm. Like k-means clustering, GMM uses an iterative algorithm that converges to a local optimum. The algorithm is composed of the following steps:

Input: Data points $x_1 \dots x_n$; k: number of clusters.; M: number of random sample to present each cluster;
1. Build GMM parameter (mixing weight, mean vector and covariance matrix).
2. E-step: Expected clusters are computed for all data points.
3. M-step: Maximum likelihood posterior probability is computed given the clusters member distribution.
4. Iterate steps 2 and 3 until the likelihood function reaches a local minimum value or stopping criterion is reached.

2.3. Mean-Shift method

Mean-Shift[1] [7] searches for the local maximal density points and then groups all the data to the clusters defined by these maximal density points. :

Input: Data points $x_1,\dots,x_n$; spatial bandwidth H_s; range bandwidth H_r; minimum segment area S.
1. Associate a mean shift point M (x_i) with each data point.
2. Compute mean shift vector $M_v(x_i)$ using a kernel density estimation function (the Parzen window function [7]).
3. Update the mean shift point M (x_i) until $M_v(x_i)$ less than some threshold
4. Merge points whose mean vectors are closer than

[1] http://www.caip.rutgers.edu/riul/research/code/EDISON/index.html

h_s in the spatial domain and h_r in the range domain to produce homogeneous color regions.
5. Optionally, eliminate small region smaller than S.

Since Mean Shift-based clustering begins at each point rather than from an initial guess like k-mean. The algorithm does not rely on a priori knowledge of the number of clusters present. The quality of the output can be controlled by the free parameters, which provides the ability to obtain more meaningful results through user interaction.There are mainly three free parameters: H_s, H_r, S. H_r, determines the color smoothness of the resulting segments. H_s determines the resolution in selecting the local maximal density points and S constrains the resulting segments.

2.4. Spectral clustering

Recently, a family of spectral clustering (SC) algorithms was proposed in the literature. SC [14] can be briefly described with the following steps:

Input: Data points $x_1,\dots x_n$; k: number of clusters
1. Construct n×n similarity matrix $S_{ij} = \exp(-\|x_i-x_j\|^2/2\sigma^2)$, σ is a scaling parameter.
2. Compute the Laplacian matrix L.
3. Compute the first k eigenvectors of L.
4. Compute the normalized matrix U.
5. Use k-means algorithm to cluster n rows of U into k groups.

When the data size is large, SC encounters a quadratic resource bottleneck in computing pair-wise similarity between n data points, and in storing such a large matrix. Moreover, the algorithm requires considerable time and memory to find and store the first k eigenvectors. In order to reduce the computational cost, different ways of approximating the dense similarity matrix are proposed such as SC using Nyström approximation [13] SC using t-nearest neighbor [15][2].

2.5. Affinity propagation

Affinity propagation[3] (AP) is a new clustering method [15], which is slightly different from common clustering methods. It takes pair-wise similarities between data points as input, where the similarity s(i,k) indicates how well the data point with index k is suited to be the exemplar for data point i. Two types of messages are passed between each pair of data points recursively until a good set of clusters emerges. The first type is the responsibility message which shows how much the source date point prefers the target data point as its exemplar. The second one is the availability message which indicates how much the source data point would like to be the exemplar of the target data point. These two types of messages influence each other so the algorithm updates them iteratively until convergence or after a predetermined number of iterations. A brief description of the algorithm step is as follows:

Input: similarity matrix ($s_{ik} = -\| x_i-x_k\|^2$); p: preference for data points be a cluster center
For each pair of data two types of messages passes iteratively
1. Construct the responsibility message r(i,k) (the availabilities are initialized to zero a(i,k) =0).
2. Construct the availability message a(i,k).
3. Update both messages till converge or stop condition reached.
5. Combine both messages to determine the centre of each point.

AP has the advantage that it can determine the cluster number automatically based on a priori setting of how preferably each data point is an exemplar. However, AP has some limitations: it is hard to know what value of parameter preference can yield an optimal clustering solution and the quadratic memory used in computing the affinity matrix.

3. Experimental results

The goal here is to compare the performance of the previous clustering methods, and to investigate their abilities to segment land-use classes in urban areas. The comparison gives some guidance and tradeoffs involved in using each method (performance, computational time, parameter tuning, etc). The software implementation of these five methods is publicly available. For k-means and GMM the Matlab implementation is used. We are targeting basic objects, which are usually found in the map such as roads, buildings, and objects that help in their extraction like green parts and shadows.

3.1. Dataset

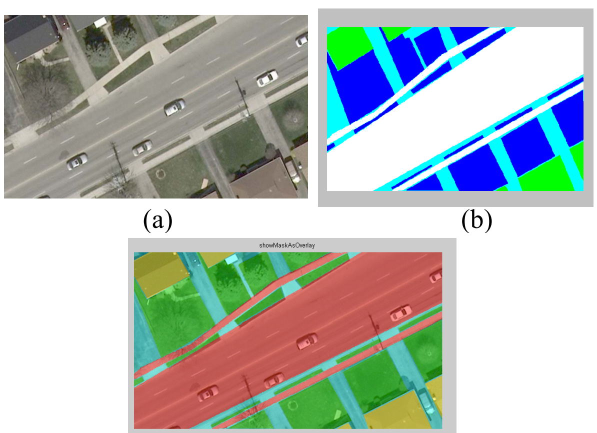
The study area is the city of Kitchener-Waterloo (K-W), Canada. The data was provided by the University Map Library at the University of Waterloo [17] as ortho-rectified aerial images taken in April 2006 at 12 cm spatial resolution by a digital color airborne camera

[2] http://www.cs.ucsb.edu/~wychen/sc.html
[3] http://www.psi.toronto.edu/affinitypropagation/

with 8 bit radiometric resolution. We cropped a set of twenty test images of size 666*372 from the original image. The cropped test images were chosen for high density urban parts which are highly corrupted by noise. Samples of the test image are shown in Figure. 1.

Figure 1. Samples of test images that are corrupted by noise

Test images were manually segmented into four land use types (roads, buildings, green area and other). Other represents pixilation which is either difficult to interpret or does not correspond to the objects of interests like building entrance with a very small parking area alongside the road, swimming pools and other small objects in the image. A sample test image is shown in Fig. 2 with its ground truth.



(a) (b)

(c)

Figure 2. Sample of the images and its ground truth: (a) original image, (b) ground truth and (c) ground truth overlaying the original image (right).

3.2. Experiment setup

To extract the basic objects, the number of clusters is defined empirically to be five clusters. It was chosen by minimize the error between the clustered image and the ground truth images. Although four to seven spectral clusters work well for most of the test images, five clusters have been selected as it gives the best average accuracy for the entire set of the test images.

Throughout this paper we compare the average performance of different clustering methods. We look at the average over 40 multiple runs for each method and consider the standard deviation. The average is reported for three methods: k-means, GMM and SC which either have random sample selection (GMM, SC-Nyström) or have random initialization (k-means and SC). The average execution time of the 40 runs are also compared.

To make a fair comparison, we compare the average performance when images are clustered into the same number of segments. As for k-mean, GMM and SC the number of clusters is a required input. AP has the advantage of getting the optimal number of clusters and it is also able to cluster the images into a specific number of clusters. However, this is not the case for the mean-shift as three free parameters which need tuning to reach to the required number of clusters.

For each clustering method there are some free parameters need to be tuned in order to assess the best average performance provided by each one of them over the whole set of the test images. The tuning of each is set as follows: For GMM, we carried out different trials with different values for the number of samples randomly chosen to present points in each cluster. The value was set to 1000 point for all images. For the mean-shift, different combinations of the free parameters were tried. Experiments show that S affects the performance much more than H_r and H_s do. The appropriate number of clusters found in our case is 20-100 with an average value of 45 clusters. For the SC-Nyström several trails were done for the smoothing parameter σ and the number of samples randomly chosen from the test image. Experiments show that 150-300 randomly chosen points are sufficient to capture the five clusters in our case. The value of σ was empirically set to 10 which is the smallest possible value to get a solution from eigenvectors. For the t-NN, two free parameters needs tuning, t and σ. We used the self tuning σ as in [14] so no tuning is required. We tried several values for t in the interval [5, 200]. The value was set to 50 where we had the best average performance over all images. For The k-means and AP need no/minimal tuning.

3.3. Results

We compare the average performance of different clustering methods that have been suggested for segmentation of images. We focus on segmentation of urban areas in HR aerial images and we report the tradeoff in sing each of such images. The original image pixel's RGB values are used as spectral feature. The average is taken over 40 for each method. The standard deviation of the error is around 0.1 and 0.2 for all methods. Table 1 shows the results of applying K-means, GMM, SC-Nyström and mean-shift over the 20 test set images. The CPU time, in seconds (for a Intel core 2 Duo T5550 @ 1.83 GHz Processors with 2 MB cache and 3 GB RAM) for all the methods is also given in Table 1 for comparison.

The aforementioned methods are applied to the full size images.

Table 1: Comparison of clustering accuracies and execution time for k-means, GMM, SC-Nyström and Mean-shift methods applied on original size aerial images.

	Roads	Green areas	Buildings	Others	Time (s)
K-means	0.684± 0.101	0.591± 0.063	0.319± 0.103	0.407± 0.055	6.4
GMM	0.722± 0.080	0.717± 0.052	0.427± 0.100	0.455± 0.077	131.1
SC	0.710± 0.085	0.639± 0.079	0.317± 0.086	0.439± 0.057	100.5
Mean-shift	0.775± 0.133	0.238± 0.0845	0.364± 0.95	0.239± 0.083	21.3

Both AP and SC require comparing all possible data points for composing the affinity matrix. For instance, a given image of 666×372 pixels has 247,752 points, and the size of the generated affinity matrix will be $247{,}752 \times 247{,}752$. This requires a great cost of computation and storage. The storage required is not practical to apply. However by using the Nyström approximation we were able to apply it for the original image size with a poor execution time. It is 16 slower than k-means which has the fastest performance in our experiment. To include AP and SC- t-NN in our comparison the images had to be resized to 256×256 to fit the memory requirement. Table 2 shows the average performance for AP, SC and k-means. SC-Nyström and k-means were applied for the same image size for clarification.

Table 2: Comparison of clustering accuracies and execution time for AP, SC-t-NN, SC-Nyström and k-means, applied on resized aerial images.

	Roads	Green areas	Buildings	Others	Time (s)
AP	0.6843 ±0.129	0.5801 ±.056	0.3507 ±0.095	0.4275 ±0.067	3516.8
SC- t-NN	0.6356 ±0.130	0.5731 ±0.100	0.3708 ±0.095	0.3859 ±0.041	62.536
SC-Nyström	0.6752 ±0.098	0.5737 ±0.073	0.3820 ±0.109	0.4217 ±0.075	73.322
K-means	0.6674 ±0.110	0.5676 ±0.094	0.3517 ±0.114	0.3901 ±0.051	1.6436

The comparisons in Table 1 and Table 2 show that all the five clustering methods give good result for some land uses such as roads and green area. The results are also comparable to each other in accuracy [10] specially in the case of the resized images. However the execution time for the k-mean is the fastest one. Higher accuracy is obtained in the expense of execution time (AP and SC). The accuracy is reduced by resizing the test images as resizing loses details.

The result shows the potential of the clustering aerial images starting from the three RGB bands only. In the experiment we could achieve an average rate of 66% of extracting road areas even in the resized images. The best accuracy for road land use is obtained using the mean-shift. However, in the case of mean-shift method the clustered images had 45 clusters on average, hence it's not fair to compare it's output to the four land uses. GMM provides the best results in extraction both in road areas and green areas. It should be also noted that the fact that images used have been captured in April for the K-W areas in Canada explains the relatively lower value of the extracted green areas as in this time of the year, the land uses are mixed between bare land and green. Taking images at this time of the year is of special importance to the city of Waterloo as it used in the urban planning process. It's however apparent that using RGB values as the only features to extract buildings is not enough. It even has the lower accuracy in extraction (30%) than the mixed land use (others). Buildings have a large internal variation i.e. roofs behave differently to the variations of the sun's incidence angle and to the orientation of its faces.

4. Conclusion

In this paper, a comparison of the performance of several clusters methods for the segmentation of HR aerial images was done. In terms of accuracy the result shows that mean-shift have the highest result for the extraction of road areas (77 %). However, its output cannot be compared to the other land uses and it requires a careful tuning. GMM has the highest performance for extracting the road and green areas, 72% and 71% respectively. The increase in accuracy yields a significant increase in run time. Considering k-means as the reference, on the average mean-shift, GMM, SC-Nyström are 3 times, 20 times, 15 times slower than k-means respectively. For the resized images case, the AP has the slightly better accuracy. However, it is too slow compared to k-means. AP can determine the optimal number given that preferences are provided correctly.

The next step in this research is to investigate the effect of adding texture and shape descriptor to raise the extraction accuracy of building and to differentiate between objects with similar spectral signatures such as roads and parking lot.

5. References

[1] P. Carleer, O. Debeir, and E. Wolff, “Assessment of very high spatial resolution satellite image segmentations”, *Photogrammetric Engineering & Remote Sensing*, 71(11), 2005, pp.1285–1294.

[2] F. Dell'Acqua, P. Gamba, A. Ferrari, J.A. Palmason, J.A. Benediktsson, K. Arnason, “Exploiting spectral and spatial information in hyperspectral urban data with high resolution”, *Geoscience and Remote Sensing Letters*, IEEE, 1(4), 2004, pp. 322-326.

[3] R. M. Haralick, and L. G. Shapiro,: *Computer and Robot Vision. Addison-Wesley*, Reading, MA, 1992.

[4] J.B. Mena, and J.A Malpica, “Color image segmentation using the Dempster sharer theory of evidence for the fusion of texture”, *Pattern Recognition. International Arch. Photogrammet. Remote Sensing*, 34, Part 3/W8. 2003.

[5] C. Evans, R. Jones, I. Svalbe, and M. Berman,” Segmenting Multispectral Landsat TM Images into Field Units”. *IEEE Trans. Geosci. Remote Sens.* 40, 2002, pp. 1054-1064.

[6] A. Rydberg, and G. Borgefors,” Integrated Method for Boundary Delineation of Agricultural Fields in Multispectral Satellite Images”, *IEEE Trans. Geosci. Remote Sens.*, 39(11), 2001, pp.2514-2520.

[7] D. Comaniciu, and P. Meer,” Mean shift: A robust approach toward feature space analysis”, *IEEE Trans. Pattern Anal. Mach. Intell.*, 24 , 2002, pp. 603-619.

[8] J. Shi, and J. Malik,, “Normalized Cuts and Image Segmentation’, *IEEE Trans. Pattern Anal. Mach. Intell.*, 22, 2000, pp. 888-905.

[9] A. Sarkar, M.K. Biswas, B. Kartikeyan, V. Kumar, K.L. Majumder, and D.K. Pal, “A MRF Model-Based Segmentation Approach to Classification for Multispectral Imagery”, *IEEE Trans. Geosci. Remote Sens.*, 40, 2002, pp. 102-113.

[10] L. Wen, X. Chen, and P. Guo, “A Comparative Study on Clustering Algorithms for Multispectral Remote Sensing Image Recognition”, In *5th Int. Symp. Neural Networks 2008: IN ANN. F. Sun,* J. Zhang, Y. Tan, J. Cao, and W. Yu, Eds. LNCS, vol. 5263, 2008, pp. 610-617 Springer, Berlin, Heidelberg.

[11] B. Larsen, and C. Aone, ‘Fast and Effective Text Mining Using Linear-time Document Clustering”, *KDD-99*, San Diego, California, 1999,pp. 16-22..

[12] P. Guo, and H. Lu, “A Study on Bayesian Probabilistic Image Automatic Segmentation”, *Acta Optica Sinica*, 22, 2002, pp.1479-1483.

[13] C. Fowlkes, S. Belongie, F. Chung, and J. Malik,” Spectral Grouping Using the Nyström Method”, *IEEE Trans. Pattern Anal. Mach. Intell.* 26, 2004, pp. 214-225.

[14] Y. Song, W.-Y.Chen, H. Bai, C.-J.Lin, and E.Y. Chang, E.Y. In *LNAI*, vol. 5212, 2008, pp. 374-389.

[15] B.J. Frey, and D. Dueck, “Clustering by Passing Messages between Data Points”, *Science*, 315, 2007, pp. 972–976.

[16] Y.J. Zhang, “A Survey on Evaluation Methods for Image Segmentation”, *Pattern Recognition* 29, 1996, pp. 1335-1346.

[17] Tri-Cities and Surrounding Communities Orthomosaics 2006 [computer file]. Waterloo, Ontario: The Regional Municipality of Waterloo 2006.