Comment: good explanation of k-means color algorithm to perform clustering on satellite images.
Nothing related to computer vision or deep learing: the clusters are then labelled "manually" as far as I understood.
Distinction between pixel based clustering and object based clustering.

Available online at www.elixirpublishers.com (Elixir International Journal)

Computer Science and Engineering

Elixir Comp. Sci. & Engg. 63 (2013) 18671-18675

Elixir ISSN: 2229-712X

Satellite Imagery Land Cover Classification using K-Means Clustering Algorithm Computer Vision for Environmental Information Extraction

Babawuro Usman

Department of Computer Science, Faculty of Computing and Mathematical Sciences, Kano University of Science and Technology, Wudil, Kano State, Nigeria.

ARTICLE INFO

Article history:

Received: 10 August 2013;
Received in revised form: 2 October 2013;
Accepted: 21 October 2013;

Keywords

K-means algorithm,
Clustering,
Satellite Imagery,
Classification.

ABSTRACT

Segmentation and classification of high resolution satellite imagery is a challenging problem due to the fact that it is no longer meaningful to carry out this task on a pixel-by-pixel basis. The fine spatial resolution implies that each object is an aggregation of a number of pixels in close spatial proximity, and accurate classification requires that this aspect be subtly considered. K-means clustering algorithm is a better method of classifying high resolution satellite imagery. The extracted regions are classified using a minimum distance decision rule. Several regions are selected as training samples for region classification. Each region is compared to the training samples and is assigned to its closest class. The procedure significantly reduces the mixed pixel problem suffered by most pixel based methods. In this paper, we used K-means clustering algorithm to classify satellite imagery into specific objects within it for cadastral and environmental planning purposes, thereby eliminating the above mentioned problems and getting better classification accuracy with the overall performance for accuracy percentage as 88.889% and Kappa values as 0.835.

Introduction

Satellite imagery is an indispensable tool in scientific research and environmental planning, with applications in numerous fields. One of the applications is in Image classification. Image classification approaches have evolved over the years. This development has been driven by the need for higher accuracies in the classified results coupled with the emergence of high resolution satellite imageries, such as Quick Bird and IKONOS which pose a greater challenge to many image classification methods [1]. The purpose of image classification is to label the pixels in the image with meaningful information of the real world for better and useful information extraction. Through classification of satellite imagery, thematic maps bearing the information such as cadastral information, land cover type, vegetation type, etc. could be obtained [2]. Image, classification methods may be grouped into two main categories depending on the image primitive used, viz pixel based or object based methods. Pixel based methods classify individual pixels without taking into account any spatial information of the pixel. Only the spectral patterns are used. On the other hand, object based methods attempt to group pixels into objects by an image segmentation process based on a chosen similarity, e.g., texture, color, intensity and then use the spectral, spatial and contextual information inherent in these objects to classify the whole image [1]. It has emerged as a superior way of doing image classification. One of its strength is the ability to extract real world objects, proper in shape and accurate in classification. It eliminates the mixed pixel problem suffered by most pixel based methods. This is because the image is classified on an object level and usually more information is used. Object based methods are also able to handle high resolution satellite imagery which aggravates the classification process for most pixel based methods [1]. Classification techniques include conventional statistical algorithms, such as discriminate analysis and the maximum likelihood classification, which allocate each image pixel to the land cover class in which it has the highest probability of membership [4]. One of the major disadvantages of these classifiers is that they are not distribution free [5]. In the same vain, traditional pixel-based classification methods have difficulty with high resolution satellite imagery, resulting in a "salt and pepper" appearance.

inform. catastali

In this paper, an attempt has been made to practical classify high resolution satellite imagery into accurate spatial groups with the following classes using the object based approach for cadastral, environmental studies and management. The classes are farmlands, bare lands, built-up areas, and others. In this study, the classification of the imagery has been done using color k-means clustering algorithm, where the imagery was classified into various classes with a view to determine the most optimum clusters based on apriori knowledge of the imagery, and then the land cover classification was performed. The aim has been to identify and classify farmlands for statutory environmental functions. Initially the imagery has been georectified to assume the planer surface that could be needed for environmental quantitative image analysis [5]. The paper is organized as follows: In section 2, we describe the related work of the paper while in Section 3, we describe the color segmentation scheme and the implementation process. In Section 4, we give experimental results and an evaluation method was presented in Section 5. Finally in Section 6, we draw a concise conclusion.

clustering

classification

Related Work

Lonesome M. M. [6], used a region-based approach for doing image classification. The main goal was to develop an alternative procedure for an object-based image classification. The procedure significantly reduced the mixed pixel problem suffered by most pixel based methods. Wen C. et al. [7], presented a Satellite image classification method using color and

Tele:
E-mail addresses: wuroabu@yahoo.com

texture features. Afroz S.M. et al. [8], presented a methodology to derive meaningful area-wide spatial information for city development and management from high resolution satellite imagery. According to Bardsley J. M. et al. [9], they presented image classification using spectral based method, quadratic discriminant analysis, where the classification of a particular pixel does not depend on that of its neighbors and the spatially based techniques, where such may be considered at times. Mayank T. [10], provided an insight into the application of feed-forward neural networks in the area of satellite image classification. The different image classification methods were compared using the satellite images by Aykut A. et al. [11]. Out of their work, the maximum likelihood method was found more applicable and reliable for their type of satellite image classification. Harish K. E. et al. [12], presented it from a slightly different perspective by using Particle Swarm Optimization to extract some cadastral features such as buildings and roads. Theirs is a land cover mapping by using swarm computing techniques. Balasubramanian S. et al. [13], approached it from another direction by using two k-means clustering algorithms with Laplacian of Gaussian, LoG, coupled with Prewitt filter. Ashwini T. et al. [14], for segmentation and classification of Satellite images, implemented two different algorithms, the k-means algorithm and back propagation algorithm of artificial neural network. Meher S.K. et al. [15], carried out multispectral imagery classification using wavelet based features. They noted that the wavelet transform provided a precise and unifying framework for the analysis and characterization of a signal at different scales unlike when Gabor transform was employed, which has the demerit of having its output bank filters not being orthogonal and this could result into having correlations between texture features. Many classification algorithms are considered vis-à-vis the nature of our imagery. In the end, the imagery was accurately classified into four environmental classes using the robust k-means clustering algorithm.

Methodology/Implementation

A high resolution satellite imagery, (400 x 400) illustrating various types of land use and land cover has been used as the test image for classification as shown in figure 3.2.

Satellite Imagery dataset: A Quick Bird Satellite imagery with a 2.4m resolution acquired in 2005, over a relatively flat landscape has been used in this study. This, imagery, 593X533, has a total number of pixel vectors, N, 316069. It is composed of a residential matrix textured with farmland patches of varying sizes and shapes which are excellent features with cadastral and environmental values. The three land-use classes dominating the scene are residential, agriculture and commercial. The satellite imagery was first geo-rectified to remove some errors and to make it assume a 2D plane surface making it suitable for environmental image assessment and analysis. Details could be seen in [5].

Satellite Imagery Classification using K-Means Algorithm: Clustering is the process by which discrete objects with similar characteristics can be assigned to groups. Clustering is used to group together like species, survey results, or satellite image data, among others. Clustering algorithm has been used as a tool to analyze varieties of data. According to [16], this concept has been researched by many clustering practitioners, indicating how useful it is in data analysis. It is the unsupervised classification of patterns derived from observations, data items and or feature vectors into groups or clusters. To carry out a clustering exercise, the following components of a typical clustering task have been considered, which include the pattern representation, definition of pattern proximity measure appropriate to certain data domain and clustering or grouping as shown in figure 3.1.

kmeans explanation.

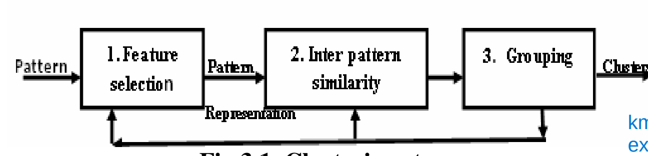


Fig 3.1, Clustering stages

The algorithm composed of the following steps:

1. Place K points into the space represented by the objects that are being clustered. These points represent initial group centroids.
2. Assign each object to the group that has the closest centroid.
3. When all objects have been assigned, recalculate the positions of the K centroids.

Repeat steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

This algorithm aims at minimizing the objective function of a squared error function. The objective function is

$$J = \sum_{j-1}^{k} \sum_{i-1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 , \qquad (3.1)$$

where $\left\| x_i^{(j)} - c_j \right\|^2$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster centre c_j, which is an indicator of the distance of the n data points from their respective cluster centres.

Color-Based Segmentation using K-Means Clustering

The algorithm of color segmentation belongs to the general class of region growing type of segmentation scheme. If a color profile is given in form of its three component profiles: red, green and blue and two markers have been defined; the markers are labeled with values. Then we find and choose a similarity measure between a point and its neighboring marked region. In this case, we defined it as the color difference between the point and its neighbor already in the marker. We simply use a trivial color difference. If *(rz, gz, bz)* and *(ry, gy, by)* are the values of two colored pixels, *z* and *y*, then the color difference is:

$$Max(|rx - ry|, |gx - gy|, |bx - by|) \qquad (3.2)$$

But this is a coarse color measure in reality. In the case of our satellite imagery a CIE L* a* b* color scale has been used. This is a uniform scale color and a standard for colors to be compared with [17]. Here L* axis represents the intensity variation from top to bottom, 100 to 0 values, which represents a perfect intensity diffuser; a* the color variation along red- green axis: positive a* is red and negative a* is green, while b* represents the color variation along blue–yellow axis of the scale with positive b* yellow and negative b* blue. In this paper, the environmental features are the objects of interest from the imagery. k-means clustering treats each object as having a location in space. It finds partitions such that objects within each cluster are as close to each other as possible, and as far as possible from objects in other clusters. Three different colors are selected based on the color and nature of our imagery for the partition and a distance metric to quantify how close two objects are to each other has been determined. This is because distance is known to be a measure of "closeness". For every object in our input imagery, k-means returns an index corresponding to a cluster. Label every pixel in the image with its cluster index. The cluster center output from k-means is determined. Using pixel labels, we separate objects in the satellite imagery by color, which resulted in images as shown in fig.3.3, fig.3.4, fig.3.5, fig.3.6, fig.3.7 and fig.3.8.

Fig. 3.2, Original Satellite imagery

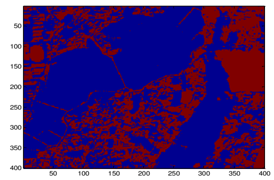


Fig. 3.3, Imagery labeled with two cluster indices

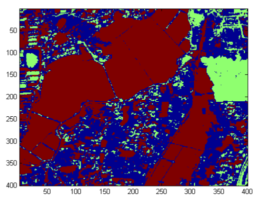


Fig. 3.4, Imagery labeled with three cluster indices

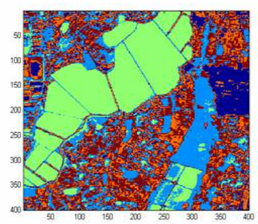


Fig. 3.5, Imagery with five clusters

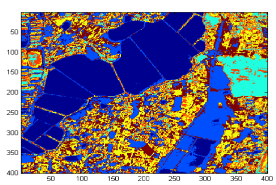


Fig. 3.6, Imagery with six clusters

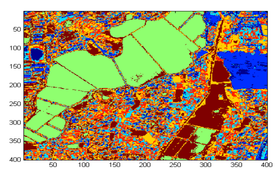


Fig. 3.7, Imagery with seven clusters

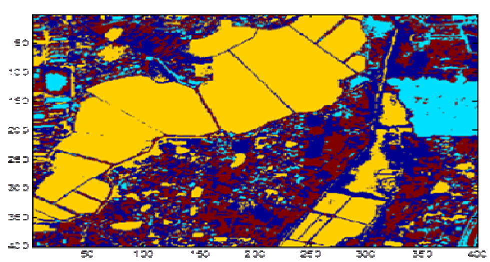
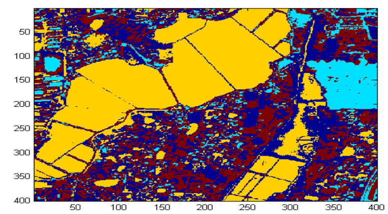


Fig 3.8, Classified imagery containing the farmlands in yellow and other three clusters

The outcome of the color segmentation with k means clustering technique is a parametric approach of image classification. Though our satellite imagery has high inter and intra class similarity but the classification was successful. The farmlands and other features with their general boundaries have been clearly classified as seen in fig 3.8, and the result is relatively excellent. The only thing to bear in mind is the correct number of clusters that the imagery should contain. For our imagery, there are four clusters, as practically determined. Although many literatures claimed that one could accurately map land cover from satellite imagery data, in practice, it is an extremely difficult task [18]. There are many instances in which land cover classification developed successfully at one site, fails in another.

Result

The procedure has been tested on a number of images. The procedure significantly reduces the mixed pixel problem suffered by most pixel based methods. In the classified image, farmland is shown in yellow, bare land in light blue, built-up in ox-blood color while others in dark blue, fig. 4.1. The classification result is influenced by the number of samples used either for segmentation or classification. Samples selected for spectral grouping have an effect on the image segmentation result. When fewer samples are used for the spectral grouping, this causes under-segmentation as shown in fig 3.3. Conversely, when too many samples are used, an over-segmented result is produced as shown in fig 3.7. The under-segmentation is easily solved by selecting more samples especially in the regions where the problem occurs.



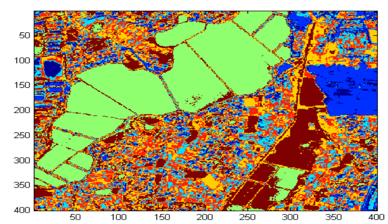


Fig.4.1, Imagery classified into four distinct classes

Farmlands — Built up

Bare lands — Others

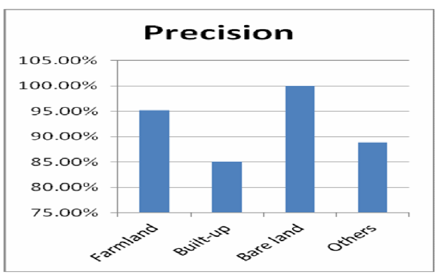
Fig.4.2, Imagery classified into seven classes

Performance Evaluation of the Algorithm:

Accuracy assessment or classification accuracy is another area that is continuing to receive increased attention when using Satellite imagery [2]. A classification is not complete until its accuracy is assessed [20]. This explains the importance of accuracy assessment of the classification result. Accuracy assessment is a general term for comparing the classification result to the ground truth, in order to determine the accuracy of the classification process [19]. Generally, classification algorithms evaluation is conducted by the Confusion Matrix. The Confusion Matrix is composed of the number cm_{ij} of elements from the class i classified in the class j. The data produced in a classification scheme are the counts of the correct and incorrect classifications from each class and this information is displayed in a Confusion Matrix, which is a form of contingency table showing the differences between the true and predicted classes for a set of labeled data [21]. Hence, our evaluation has been based on comparison with the visual original imagery and then tabulation of Confusion Matrix or rather Precision-Recall Values as in [22]. A matrix representing classification accuracy test (%) for the classification has been neatly constructed as shown in Table 5.1. The Kappa coefficient of the confusion matrix was calculated too to evaluate the agreement between the result and the reference imagery. It is used as an index of accuracy for the classification. It has been recommended as stated by [23] as a suitable accuracy measure in thematic classification for representing the whole Confusion Matrix. This is because it takes all the elements into consideration, rather than just the diagonal elements which occur when calculating the overall classification accuracy. Overall classification accuracy for a single class is determined by dividing the number of correctly classified samples by the total number of samples. Table 5.1 shows Confusion matrix accuracy, with the overall performance for accuracy percentage as 88.889% and Kappa values of 0.835.

Table 5.1 Confusion Matrix/ Precision-Recall values for satellite imagery classification

	True Classes					
Classes/ Classifier Results	Classification overall / Total	Farm land	Built-up	Bare land	Others	**Producer Accuracy (Precision)**
Farmland	21	20	0	0	1	95.238%
Built-up	47	1	40	4	2	85.106%
Bare land	4	0	0	4	0	100%
Others	45	2	3	0	40	88.889%
Truth Overall	117	23	43	8	43	
User Accuracy (Recall)		86. 957%	93. 023%	50%	93. 023%	
Overall Accuracy (OA):	88.889%					
Kappa	0.835					



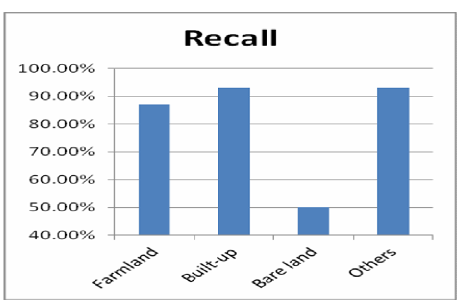


Fig.5.1 Precision-Recall percentages for the classes

Conclusion

Good results have been obtained by using this procedure. The main strength of this approach is the reduction of the mixed pixel problem suffered by most pixel based methods. The result of the classification is much cleaner so post processing is avoided. The procedure is also interactive and has few tuning parameters. On the whole, this approach has demonstrated the strength of classifying satellite images on an object level with better clarity and higher accuracy. The classified information can be used for land use/cover management, and for planning [3] as well as for cadastral purposes.

Acknowledgment

Xu Bing from C.S.U, who secured the Satellite imagery for the research, has been acknowledged as well.

References

[1]. B. Usman, Image Processing Algorithms for Automated Cadastral Feature Extraction and Delineation using Satellite Imagery. PhD Thesis, Central South University, 2012.

[2]. B. Usman, Z. Beiji, Satellite Imagery Quality Evaluation using Image Quality Metrics for Quantitative Cadastral Analysis. International Journal of Computer Applications in Engineering Sciences, 2011, 1(1):391~ 395.

[3].T. Seemann, Digital Image Processing Local Segmentation. PhD Thesis, Monash University, Australia, 2002.

[4].D.V. Rao, S. Patil, N.A. Babu,V. Muthukumar, Implementation and Evaluation of Image Processing Algorithms on Reconfigurable Architecture using C-based Hardware Descriptive Languages. International Journal of Theoretical and Applied Computer Sciences, 2006, 1(1): 9~ 34.

[5]. B. Usman, Z. Beiji, X. Bing, High Resolution Satellite Imagery Rectification using Bi-linear Interpolation Method for Geometric Data Extraction. ISDEA2012, Second International Conference on Intelligent System Design and Engineering Application, 2012:1430~1434.

[6]. M. Lonesome, A Region Based Approach to Image Classification, Photogrammetry, Earth Observation Systems, Information Extraction. Applied Geoinformatics for Society and Environment-Stuttgart University of Applied Sciences, 2009

[7].W. Cao, T.Q. Peng, and B.-C. Li, A Remote Sensing Image Classification Method Using Color and Texture Feature, (Eds.): ISNN 2004, LNCS 3173, pp. 965–970, 2004.© Springer-Verlag Berlin Heidelberg, 2004

[8].A.S. Mohammed and S. Rusthum, Object-Oriented Image Processing of an high resolution satellite imagery with perspectives for urban growth, planning and development, International Journal of Image Processing, volume (2) issue (3), 2011.

[9].J. M. Bardsley, M. Wilde, C. Gotschalk, M.S. Lorang, MATLAB Software for Supervised Classification in Remotely Sensing and Image Processing. Journal of Statistical Software, 2010, 55(2):4.

[10].M.Toshniwal, Satellite Image Classification using Neural Networks. Third International Conference: Sciences of Electronic, Technologies of Information and Telecommunications, 2005.

[11].A. Akgün, A.H. Eronat, N. Türk, Comparing Different Satellite Image Classification Methods. International Society for Photogrammetry and Remote Sensing Journal, ISPRS, July 2004, 35:1091~ 1097.

[12].E. H. Kundra, V.K. Panchal, K. Singh, H. Kaura, S. Arora, Extraction of Satellite Image using Particle Swarm Optimization. International Journal of Engineering, IJE, October 2010, 4(1): 86 ~ 92.

[13]B. Subbiah, S. Christopher, Image Classification through integrated K- Means algorithm. IJCSI International Journal of Computer Science Issues, March 2012, 9,2(2):518 ~ 524.

[14].A. T. Sapkal, C. Bokhare, N. Z. Tarapore, Satellite Image Classification using the Back Propagation Algorithm of Artificial Neural Network. Geomatrix Conference, 2009**.**

[15].S.K. Meher, B.U. Shankar, A. Ghosh, Multispectral Remote Sensing Image Classification using Wavelet Based Features. StudFuzz 210, 2007, 3~34.

[16]. A.K. Jain, M.N. Murty, P.J. Flynn, Data Clustering: A Review. ACM Computing Surveys, September, 1999, 31, (3).
[17].Hunter Lab, Application Notes: CIE L*a*b* Color Scale, Insight on Color, 2008, 8 (7).
[18].S.M. Dejong, F. D. Van der Meer, Remote Sensing and Digital Image Processing, Kluwer Academic Publishers, 2004, 5.
[19].O. Sharmay, D. Miocz, F. Antony, Polygon Feature Extraction from Satellite Imagery based on Color Image Segmentation and Medial Axis. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2008, 37: 235~240.
[20].A.P Bradley, The Use of the Area Under the ROC Curve in The Evaluation of Machine Learning Online draft paper. Cooperative Research Centre for Sensor Signal and Information Processing, The University of Queensland, 1997.
[21].J. R. Landis, G.G. Koch, The Measurement of Observer Agreement for Categorical Data. Biometrics, 2012, 33(1):159~174.
[22].P. Gong, P.J. Howarth, Land-use classification of SPOT HRV data using a cover frequency method. International Journal of Remote Sensing, 1992, 13(8):1459~ 1471.
[23].Robert C. Weih, Norman D. Riggan, A Comparison of Pixel-based versus Object based Land Use/Land Cover Classification Methodologies.