Unsupervised Classification of Remote Sensing Images using K-means Algorithm

N.Vishwanath

Assistant Professor
Department of Electronics and Communication Engineering,
St.Peters Engineering College, Dhoola pally;
Ranga Reddy (Dt); Telangana, India.

B.Ramesh

Assistant Professor
Department of Electronics and Communication Engineering,
St.Peters Engineering College, Dhoola pally;
Ranga Reddy (Dt); Telangana, India.

Sreenivasa Rao P

Assistant Professor
Department of Electronics and Communication Engineering,
St.Peters Engineering College, Dhoola pally;
Ranga Reddy (Dt); Telangana, India.

Abstract-- Image classification techniques are used in the field of remote sensing to cluster pixels in order to represent land cover features. Land cover could be classified into forested, urban, agricultural and other type of land. In unsupervised classification, the user identifies the number of clusters to generate and the bands which has to be used. Using this information clusters are generated from the image. In this paper, we propose a clustering algorithm which can achieve the robustness required in unsupervised classification of remote sensing images. The experimental results prove that the desired algorithm is very effective in producing desired clusters of the given data sets as well as diagnosis. The algorithm is very much useful in classification as well as extraction of regions in satellite images.

Keywords: Clustering algorithm, Image Classification, Image Segmentation

I.   INTRODUCTION

Clustering is the process by which a set of objects are grouped together in such a way that the objects in the same group are more similar to each other than to the objects in the different group. Clustering can be used in varied applications like Marketing in order to find customers with similar behavior given a large database of customer data. In biology we can use it to classify plant and animals based on their features. We can use clustering in libraries to arrange the books in ordering. Insurance companies can identify groups of motor insurance holders with a high average claim cost and to thereby identify frauds. The applications which are developed based on clustering are infinite. The main aim of this paper is provide a segmentation algorithm which can be used for image classification. We have combined k-means clustering algorithm with edge detection algorithms like LoG filter and Prewitt filter to achieve this. Image classification is done in two steps

    a)   Segmentation step
    b)   Classification step

1.1 K-Means Algorithm

K means clustering is an unsupervised algorithm that tries to cluster data based on their similarity. Initially the algorithm assigns each data to a cluster randomly and finds the centroid of each cluster. Then the algorithm executes the following two steps in iteration

- Reassign data points to the cluster whose centroid is closest.
- Calculate new centroid of each cluster.

These two steps keep on executing till the intra class variation cannot be reduced any further by a great deal. The intra cluster variation is calculated as the sum of the Euclidean distance between the data points and their respective cluster centroids.

1.2 Laplacian of Gaussian (LoG) Filter

The Laplacian is a 2-D isotropic measure of the 2nd spatial derivative of an image. TheLaplacian of an image highlights regions of rapid intensity change and is therefore often used for edge detection. Before applying the Laplacian to an image we smooth it by applying a Gaussian smoothing filter to the image.That is why it is named as Laplacian of Gaussian filter. The Gaussian operator uses a Gaussian kernel to blur images and remove noise from it. In 2-D Gaussian operator has the distribution

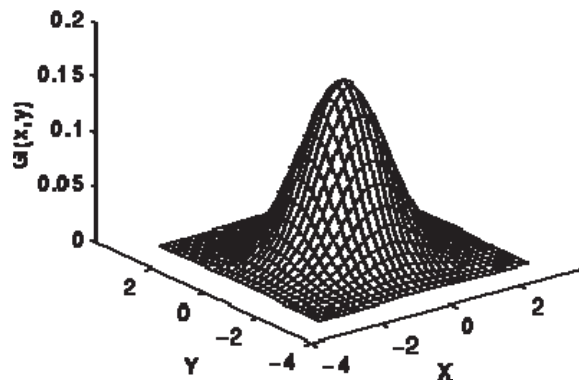$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

Fig 1 Gaussian distribution

The Gaussian filter uses this distribution as a point-spread function. The Laplacian L(x, y) of an image with pixel values I(x, y) is given by

$$L(x, y) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}$$

Combining Laplacian and Gaussian gives the function as

$$LoG(x, y) = -\frac{1}{\pi\sigma^4}\left[1 - \frac{x^2 + y^2}{2\sigma^2}\right] e^{-\frac{x^2+y^2}{2\sigma^2}}$$
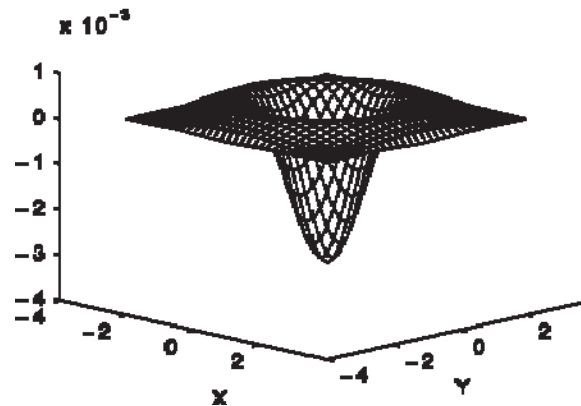
Fig. 2 LoG Function. The x and y axes are marked instandard deviations

The kernel size should be selected large enough to accommodate the negative values.

1.3 Prewitt Filter

Prewitt operator is used for edge detection in images. It detects two types of edges – horizontal edges and vertical edges. Edges are calculated by using difference between corresponding pixel intensities of an image. All the masks that are used for edge detection are also known as derivative masks. The Prewitt filter is similar to the Sobel in that it uses two 3 x 3 kernels,one for changes in the horizontal direction, and one for changes in the vertical direction. The two kernels are convolved with the original image to calculate the approximations of the derivatives.

II.   LITERATURE REVIEW

Bentley et al in his 1980 research proposed an algorithmic paradigm that would yield a number of algorithms and data structures for multidimensional problems like classification of images. Moore et al. in 1999 used multiresolution kd-trees to reduce the cost of EM based clustering of data. Ordonez integrated k-means clustering with a relational DBMS (Data Base Management System) by using 3 SQL implementations: 1) a straightforward translation of K-Means computations into SQL, 2) an optimized version based on improved data organization, efficientindexing, sufficient statistics, and rewritten queries, and 3) an incremental version that uses the optimized version as a building block with fast convergence and automated reseeding. Zhang et al. in 2006 incorporated a parallel strategy in clustering method. For enhancing the efficiency of parallel k-means they used dynamic load balance. Ding et al. in 2007 proposed a new improved K-Means algorithm with penalized term. They first calculated a prior probability distribution based on classic K-means algorithm. Then the new objective function for the penalized K-means algorithm was introduced by minimizing this function with genetic algorithm.

There have been a lot of research and proposition on the different versions of k-means algorithm. Zhang et al. proposed a method to remove the sensitivity to the initial centroid and instability. Jian Zhu used genetic algorithm to optimize the k value and improve the clustering performance. Chiang et. al. in 2011 presented an efficient algorithm for reducing the computation time of k-means and k-means based clustering algorithms and termed it as pattern reduction(PR). . The proposed algorithm worked by compressing and removing at each iteration patterns that were unlikely to change their membership thereafter. Not only was this proposed algorithm simple and easy to implement, but it could also be applied to many other iterative clustering algorithms such as kernel-based and population-based clustering algorithms.

III. PROPOSED ALGORITHM

The basic principle of the proposed algorithm was to integrate the K-Means algorithm with LoG filter and Prewitt filter as follows.

Algorithm: 1

Step 1: Read the RGB Image available for classification.

Step 2: Convert available image from RGB color space to L*a*b* color

Space Step 3: Classify the colors in 'a*b*' space using K-means clustering algorithm

Step 4: Label every pixel of the image using the results from K-means algorithm

Step 5: Create images that segment the image by color.

Step 6: Segment the nuclei of the image into a separate image

Step 7: The Laplacian of Gaussian filter finds edges by looking for zero crossings.
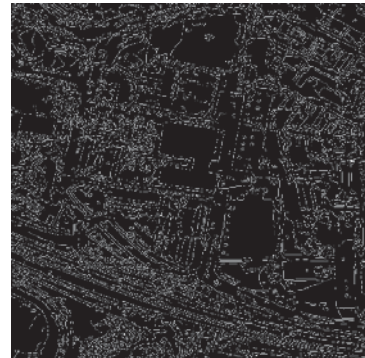
Algorithm: 2

Step 1: Do the Step 1 to 6 proposed in the algorithm 1

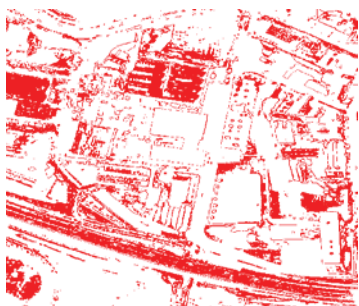Step 2: Prewitt Filter finds edges by looking for zero crossings.

IV. EXPERIMENTAL RESULTS

Original Image

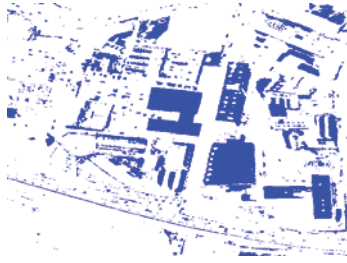Prewitt Edge Image

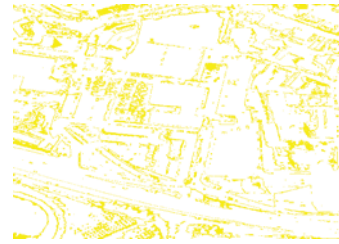First Cluster

Second Cluster

Third Cluster   Fourth Cluster

V.   CONCLUSIONS

In this paper, we have combined K Means algorithm with Laplacian of Gaussian algorithm and compared that with the combination of K Means algorithm with Prewitt Filter. The integrated novel clustering algorithms for image classification are tested with different images including satellite images. We found that the latter is performing well compared to the earlier. These algorithms are robust and very effective in producing clusters.

REFERENCES

[1]   J.L. Bentley. Multidimensional Divide-and-Conquer, Vol 23 Issue 4:214-229, April 1980, *Communications of the ACM*.
[2]   Andrew Moore. Very Fast EM-based Mixture Model Clustering using multi-resolution KD-trees, pp. 543-549, April 1999, *Proceedings on Advances in Neural Information Processing Systems*.
[3]   Jian Zhu, Hanshi Wang. An improved K-means clustering algorithm, pp. 190-192, April 2010, *The 2nd IEEE International Conference on Information Management and Engineering (ICIME)*.
[4]   Zhe Zhang, Junxi Zhang, HuifengXue. Improved K-Means Clustering Algorithm, Vol 5:169-172, May 2008, *IEEE Congress on Image and Signal Processing*.
[5]   Carlos Odonez. Integrating K-Means Clustering with a Relational DBMS using SQL, Vol 18:188-201, February 2006, *IEEE Transactions on Knowledge and Data Engineering.*
[6]   Yufang Zhang, ZhongyangXiong, Jiali Mao, Ling Ou. The Study of Parallel K-Means Algorithm, Vol 2:5868-5871, 2006, *6th World Congress on Intelligent Control and Automation*.
[7]   Zejin Ding, Jian Yu, Yanqing Zhang. A New Improved K-Means Algorithm with Penalized Term, pp. 313, Nov 2007, *IEEE International Conference on Granular Computing.*