River extraction from satellite image

Ms. Venu Shah[1], Ms. Archana Choudhary[2] and Prof. Kavita Tewari[3]

[1]Sr. Lecturer, Electronics and telecommunication department
Atharva College of Engineering, Mumbai, India

[2]Lecturer, Electronics and telecommunication department
Atharva College of Engineering, Mumbai, India

[3]Head, Electronics department,VESIT,Chembur, Mumbai, India

Abstract

Satellite image processing plays a cr ucial role for the research developments in many fields of study including Astronomy, Remote Sensing, GIS, Agriculture Monitoring and Disaster Management. The remote sensing images are utilized in many of the researches with the aim of predicting natural disasters so that essential precautions can be taken to protect the environment. Besides the other, the water resource analysis plays a vital role in these researches. Traditionally, lots of methods are utilized for the analysis and determine some resources like water which are becoming extinct in nature. In this work, the methods like edge detection, thresholding, image erosion and other color and feature extraction algorithms are presented to extract water content (river). The algorithms used here includes, K means clustering algorithm, Hill Climbing Algorithm, Color histogram and image thresholding. Here, the condition of river like normal, drought or flood is also predicted by visual inspection of the processed satellite image.

Keywords: *Binary Thresholding, Hill Climbing Algorithm and K means clustering algorithm.*

1. Introduction

Content extraction particularly from satellite is gaining great importance as it addresses to retrieving specific problem like extraction of vegetation, road, railways, water bodies etc. from the image. Water is becoming extinct in some areas while in the rest it is causing trouble like floods. The salient image regions are often useful for applications like image segmentation, adaptive compression and region-based image retrieval and obtained by mapping the pixels into various feature spaces, where upon they are subjected to various grouping algorithms. In this paper, the algorithms explained are used to predict condition of the rivers such as drought, normal or flooded and thus helps to predict the amount of water content in an area.

2. Scheme1: River image processing using Color Using Color histogram, k means clustering and Hill climbing Algorithm

2.1 K means Clustering

In this method of image segmentation we use K means method to divide the image into a number of clusters. But before considering all this it is important to understand how a basic K means algorithm works K-means (Macqueen, 1967) is one of the simplest unsupervised learning algorithms that solve the well known clustering problem and is used to cluster observations into groups of related observations without any prior knowledge of those relationships.

The algorithm clusters observations into k groups, where k is provided as an input parameter. It then assigns each observation to clusters based upon the observation's proximity to the mean of the cluster. The cluster's mean is then recomputed and the process begins again.

2.1.1 Algorithm Steps

1. To classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define k centroids, one for each cluster.
2. These centroids should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other.
3. The next step is to take each point belonging to a given data set and associate it to the nearest centroid.
4. Continue the process until no point is pending. At this point we need to re-calculate k new centroids as centers of the clusters resulting from the previous step.

5. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroids and a loop is generated.
6. As a result of this loop we may notice that the k centroids change their location step by step until no more changes are done. In other words centroids do not move any more.

2.1.2 Objective Function

This algorithm aims at minimizing a squared error function (objective function) given by:

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

,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, is an indicator of the distance of the n data points from their respective cluster centres.

One of the main disadvantages to k-means is the fact that it requires the number of clusters as an input to the algorithm. The algorithm is not capable of determining the appropriate number of clusters and depends upon the user to identify this in advance. For example, if you had a group of people that were easily clustered based upon gender, calling the k-means algorithm with k=3 would force the people into three clusters, when k=2 would provide a m ore natural fit. Similarly, if a group of individuals were easily clustered based upon home state and you called the k-means algorithm with k=20, the results might be too generalized to be effective.

For this reason, it's often a good idea to experiment with different values of k to identify the value that best suits your data. Suppose that we have n sample feature vectors x_1, x_2, ..., x_n all from the same class, and we know that they fall into k compact clusters, $k < n$. Let m_i be the mean of the vectors in cluster i. If the clusters are well separated, we can use a minimum-distance classifier to separate them. That is, we can say that x is in cluster i if $\| x - m_i \|$ is the minimum of all the k distances. This suggests the following procedure for finding the k means:

Make initial guesses for the means m_1, m_2, ..., m_k

Until there are no changes in any mean
1. Use the estimated means to classify the samples into clusters.
2. For i from 1 to k, replace mi with the mean of all of the samples for cluster i

Else end.

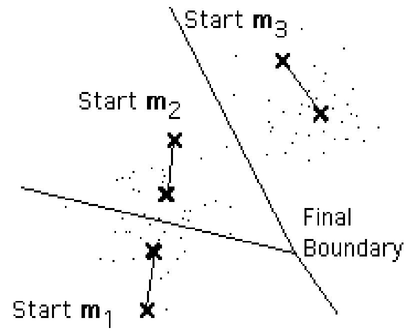


Figure1: K means algorithm

Here is an example showing how the means m_1 and m_2 move into the centers of two clusters.

2.2 Hill climbing algorithm

Hill climbing is good for finding a local optimum (a good solution that lies relatively near the initial solution) but it is not guaranteed to find the best possible solution (the global optimum) out of all possible solutions (the search space).

The relative simplicity of the algorithm makes it a popular first choice amongst optimizing algorithms. It is used widely in artificial intelligence, for reaching a g oal state from a starting node. Choice of next node and starting node can be varied to give a list of related algorithms. Hill climbing can often produce a better result than other algorithms when the amount of time available to perform a search is limited, such as with real-time systems. Hill climbing technique can be mathematically represented as follows:

Hill climbing algorithm attempts to maximize (or minimize) a target function $f(\mathbf{x})$, where $\mathbf{X}$ is a vector of continuous and/or discrete values. At each iteration, hill climbing will adjust a single element in $\mathbf{X}$ and determine whether the change improves the value of $f(\mathbf{x})$. (Note that this differs from gradient descent methods, which adjust all of the values in $\mathbf{X}$, at each iteration according to the gradient of the hill.

With hill climbing, any change that improves $f(\mathbf{x})$ is accepted, and the process continues until no change can be found to improve the value of $f(\mathbf{x})$. $\mathbf{X}$ is then said to be "locally optimal".

In discrete vector spaces, each possible value for $\mathbf{X}$ may be visualized as a vertex in a graph. Hill climbing will follow the graph from vertex to vertex, always locally increasing (or decreasing) the value of $f(\mathbf{x})$,until a local maximum (or local minimum) x_m is reached.

2.3 Color histogram

In image retrieval systems color histogram is the most commonly used feature. The main reason is that it is independent of image size and orientation. Also it is one of the most straight-forward features utilized by humans for visual recognition and discrimination. Statistically, it denotes the joint probability of the intensities of the three color channels. Once the image is segmented, from each region the color histogram is extracted. The major statistical data that are extracted are histogram mean, standard deviation, and median for each color channel i.e. Red, Green, and Blue. So totally 3× 3 = 9 features per segment are obtained. All the segments need not be considered, but only segments that are dominant may be considered, because this would speed up the calculation and may not significantly affect the end result.

2.4 Procedure

Following iterations are performed to get the color based image retrieval using hill climbing segmentation:
1) Converting input image from RGB to CIE Lab
2) Generate CIE Lab 3D Color Histogram
3) Searching the initial seeds from 3D color histogram
4) Obtain 3 initial seeds for K-means clustering
5) Applying K-means clustering.
l

3. Scheme2: River extraction by using grayscale and Thresholding for segmentation and image erosion

3.1 Thresholding

It is a simple and effective means for obtaining a segmentation of images in which different structures have contrasting intensities or other quantifiable features. The input to a thresholding operation is typically a grayscale or color image. In the simplest implementation, the output is a binary image representing the segmentation. Black pixels correspond to background and white pixels correspond to foreground (or vice versa). In simple implementations, the segmentation is determined by a single parameter known as the intensity threshold. In a single pass, each pixel in the image is compared with this threshold. If the pixel's intensity is higher than the threshold, the pixel is set to, say, white in the output. If it is less than the threshold, it is set to black.

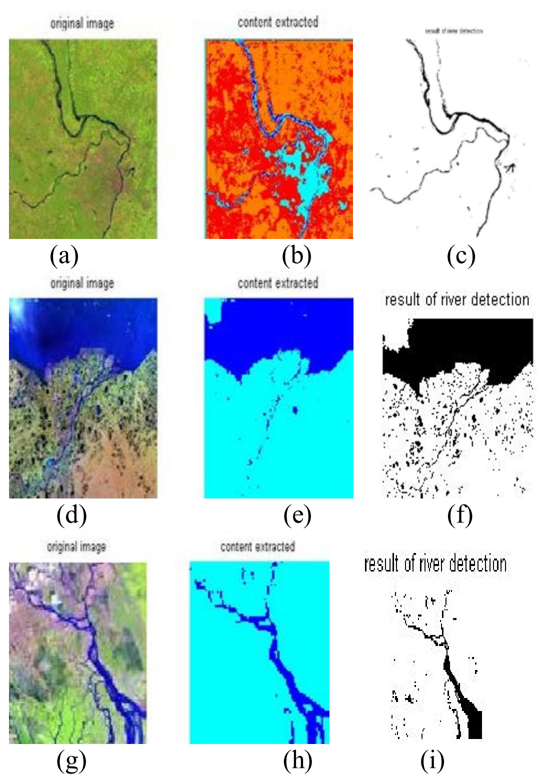
It converts the grayscale image I to a binary image. The output image BW replaces all pixels in the input image with luminance greater than level with the value 1 (white) and replaces all other pixels with the value 0 (black). Specify level in the range [0, 1]. This range is relative to the signal levels possible for the image's class.

3.2 Program code

The code for which is as follows:

```
clc;
clear all;
a=imread('Belgrade.jpg');
figure(1),imshow(a),title('original satellite image of Belgrade river ');
x= rgb2gray(a);
b=im2bw(a,.35);
c=(~b);
d=imfill(c,'holes');
s=immultiply(x,b);
x1=double(b);
mask = adapthisteq(x1);
se = strel('square',5);
marker = imerode(mask,se);
obr = imreconstruct(marker,mask);
figure(3),imshow(obr,[]),title('result of river detection');
```

4. Results of river extraction from the satellite image [17], [18], [19], [20] and [21]



(a) (b) (c)
(d) (e) (f)
(g) (h) (i)

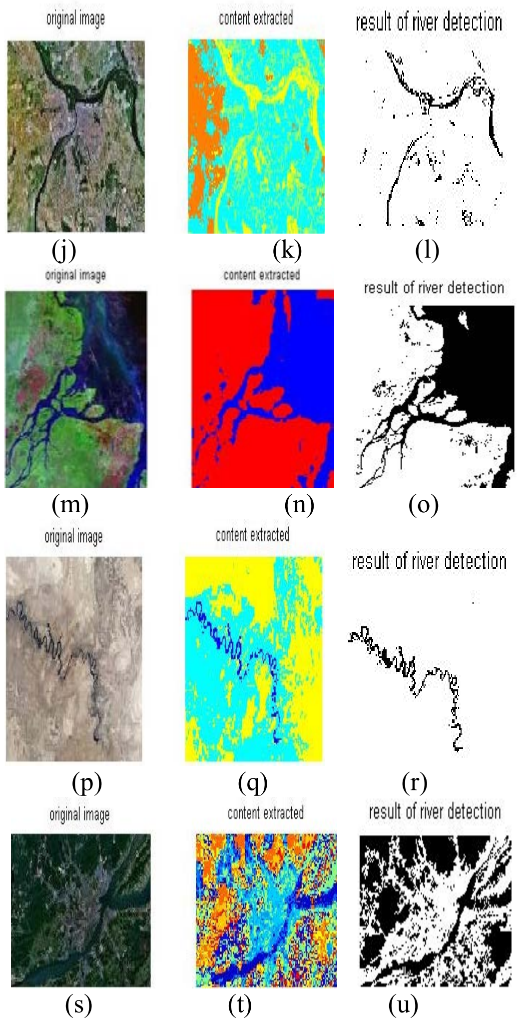


Figure2: (a), (d), (g), (j), (m), (p) and (s) are original input images of rivers(viz. Missisippi river, Colville River delta, ganges-brahmaputra River delta, Belgrade River, Amazon River , Boteti river Botswana and Montmorency River respectively) from satellite.

Figure2: (b), (e), (h), (k), (n), (q) and (t) are the results obtained from k-means algorithm along with color histogram and Hill climbing algorithm.

Figure2: (c), (f), (i), (l), (o), (r) and (u) are the results obtained from thresholding, followed by image erosion and edge detection.

Note: For figure 2: (c), (f), (i), (l), (o) and (r) shows results obtained at the specified threshold level in table 1and shows that a very low contrast image (eg. figure 2(s)) does not give satisfactory result (figure 2(u)) even if the threshold is reduced.

Table 1: impact of threshold level and number of histogram of bins

Figure No	**Name of the image**	**Scheme 1: Number of histogram bins**	**Scheme 2: Value of the Threshold level**
2(a)	Missisippi river	13	0.32
2(d)	Colville River delta	7	0.2
2(g)	Ganges-brahmaputra River delta,	11	0.35
2(j)	Belgrade River	12	0.15
2(m)	Amazon River	10	0.28
2(p)	Boteti river Botswana	14	0.40
2(s)	Montmorency River	25	0.15

Above table 1 shows that for a very high contrast image less number of bins gives good result while for a very low contrast image scheme 2 cannot be used and it also requires large number of bins to uniquely identify the content from the image.

Below table 2 shows that for a very low contrast image the status of the river is not predictable and scheme 1 gives better result while for a high contrast image, scheme 1and 2 both gives sufficient results.

Table2: Condition of the river

Figure No	Name of the image	Condition of the river
2(a)	Missisippi river	Normal
2(d)	Colville River delta	Flooded
2(g)	Ganges-brahmaputra River delta,	Normal
2(j)	Belgrade River	Normal
2(m)	Amazon River	Flooded
2(p)	Boteti river Botswana	Drought
2(s)	Montmorency River	Unpredictable

5. Conclusion

Thus, the proposed Hill climbing algorithm along with k means clustering algorithm works well and helps in extracting color of the desired information from the image while edge detection and threshold level also helps to extract fine details from the satellite image. Both these schemes can be used to predict river condition as drought, flooded or normal.

By varying threshold level in k m eans clustering a distinguished color that helps to distinctly identify the desired information.

By varying threshold level and then applying image erosion and edge detection fine details can be extracted from the image but this gives gray level.

6. References

[1]. M. Inaba, N. Katoh, and H. Imai, ªApplications of Weighted Voronoi Diagrams and Randomization to Variance-Based k-clustering, Proc. 10th Ann. ACM Symp. Computational Geometry, pp. 332-339, June 1994.

[2]. A.K. Jain and R.C. Dubes, Algorithms for Clustering Data. Englewood Cliffs, N.J.: Prentice Hall, 1988.

[3]. S. Arora, P. Raghavan, and S. Rao, "Approximation Schemes for Euclidean k-median and Related Problems", Proc. 30th Ann. ACM Symp. Theory of Computing, pp. 106-113, May 1998.

[4]. S. Kolliopoulos and S. Rao, ªA Nearly Linear-Time Approximation Scheme for the Euclidean k-median Problem, Proc. Seventh Ann. European Symp. Algorithms, J. Nesetril, ed., pp. 362-371, July 1999.

[5]. P.K. Agarwal and C.M. Procopiuc, ªExact and Approximation Algorithms for Clustering,º Proc. Ninth Ann. ACM-SIAM Symposium on Discrete Algorithms, pp. 658-667, Jan. 1998.

[6]. An Efficient k - Means Clustering Algorithm: Analysis and Implementation by Tapas Kanungo, David M. Mount,Nathan S. Netanyahu, and Christine D. Piatko, Ruth Silverman, and Angela Y. Wu, IEEE transactions on pa ttern analysis and machine intelligence, VOL. 24, NO. 7, July 2002.

[7]. A pplication of a hill-climbing algorithm to exact and approximate inference in credal networks,, Andres Cano, Manuel Gomez, Seraf Moral, 4th International Symposium on Imprecise Probabilities and Their Applications, Pittsburgh, Pennsylvania, 2005.

[8]. http://www.wikepedia.com/ Hill climbing algorithm.html

[9]. Russell, Stuart J.; Norvig, Peter (2003), Artificial Intelligence: A Modern Approach (2nd ed.), Upper Saddle River, New Jersey: Prentice Hall, pp. 111–114, ISBN 0-13-790395-2, http://aima.cs.berkeley.edu/

[10]. M.J.Swain and D.H.Ballard, "Color Indexing," International Journal of Computer Vision, 1991, Vol. 7, No.1, pp.11-32.

[11]. G. Pass and R. Zabih, "Histogram Refinement for Content-Based Image Retrieval," IEEE Workshop on A pplications of Computer Vision, pp. 96-102, December 1996.

[12]. C. Y. Wu, X.Y. Tai and J.Y. Zhao, "Image Retrieval Based on Color Feature," Computer applications, 2004, Vol. 24, No. 6, pp. 135-137.

[13]. X.J. Zhang, Research on color feature based image retrieval, Master degree thesis, Liaoning technical university, 2005.

[14]. Gonzalez, Rafael C. & Woods, Richard E. (2002). Thresholding In Digital Image Processing, pp. 595–611. Pearson Education, ISBN 81-7808-629-8.

[15]. Shapiro, Linda G. & Stockman, George C. (2002). "Computer Vision". Prentice Hall, ISBN 0-13-030796-3.

[16]. Mehmet Sezgin and Bulent Sankur, Survey over image thresholding techniques and quantitative performance evaluation, Journal of Electronic Imaging 13(1), 146–165 (January 2004).

[17]. Rajiv Kumar Nath and Deb, "Water-Body Area Extraction from High Resolution Satellite Images-An Introduction, Review, and Comparison", International Journal of Image Processing (IJIP), Vol.3, No.6, pp.353-372,2009.

[18]. Alaguraja, Yuvaraj, Sekar, Muthuveerran and Manivel, "Remote Sensing and GIS Approach for the Water Pollution and Management In Tiruchirappli Taluk, Tamil Nadu, India ", International Journal of Environmental Sciences, Vol.1, No.1, pp.66-70,2010

[19]. Andreoli Remi, Li Jiren and Yesou Herve, "Flood Extent Prediction from Lake Heights and Water Level Estimation from Flood Extents Using River Gauges, Elevation Models And Envisat Data", In proceedings of Envisat Symposium, Switzerland, 2007

[20]. R. Kalaivani and P. Thangaraj, "An Efficient Mechanism to Predict the Stage of River using Artificial Neural Network in Satellite Images" , European Journal of Scientific Research ISSN 1450-216X Vol.50 No.3 (2011), pp.385-398.

[21]. 'Automatic Extraction of Rivers in Tropical Rain Forests from JERS-1 SAR Images using Spectral and Spatial Information', by Takako Sakurai-Amano, Shogo Onuki, Mikio Takagi, Dept. of Applied Electronics, Science Univ. of Tokyo, 2641 Yamazaki, Noda, Chiba, Japan 278-8510.

Acknowledgments

We are very thankful to the Electronics and telecommunications department of Vivekanand Education Society's College of Information and Technology and AET's Atharva College of engineering for proving us with the necessary facilities to complete our research in the field

CBIR. We also thank our family, friends for supporting us throughout our research.

Biography

Ms. Venu Shah, BE(sEXTC), pursuing ME(EXTC) from VESIT and has over nine years of teaching experience and is working as Senior lecturer in Electronics and telecommunication department at Atharva college of engineering Three papers ‘Image fusion based on genetic algorithm’ and ‘Genetic algorithm and its application to image fusion’ and ‘image enhancement techniques’ were published in national conference at Nagpur, Bhopal and Indore, India in the year 2010. The areas of interest are wireless communications, computer communications and image processing. Two papers have been accepted in international journals for publish this July; they are also in the area of image processing.

Ms. Archana Choudhary, BE(EXTC), pursuing ME(EXTC) from VESIT and h as over eight years of teaching experience and is working as lecturer in Electronics and telecommunication department at Atharva college of engineering. Two papers were published in national conference, India in the year 2010. The areas of interest are wireless communications, digital communication and image processing.

Prof. Kavita Tewari, working as Assistant Professor and Head(electronics) department, VESIT since last 15 years and has interest in the field of Digital signal processing and Digital image processing. She has guided many ME and BE level projects and has number of papers published on her name.