ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 11, November 2013

An Approach Using Imbalanced Data Classification for Accurate Detection of Roads

Neetha Joseph[1], Arun kumar M N[2]

PG Scholar, Department of Computer Science and Engineering, FISAT, Ernakulam, India [1]

Assistant Professor, Department of Computer Science and Engineering, FISAT, Ernakulam, India [2]

Abstract: Reliably extracting information from satellite and aerial imagery is a difficult problem with many practical applications. Extraction of objects such roads, buildings etc from high resolution satellite imagery is an important task in urban planning, military applications etc. This is a difficult task because of occlusions, shadows, and other non-road objects. This paper proposes a method for road detection. It also highlights the importance of the imbalanced data classification in detecting the road in a complex scenario. First, the spectral angle of multi spectral input image is calculated and then dog filtering is applied. Then this image is subjected to morphological operations and area based filtering. At this stage, there may be some noises like buildings, rivers etc. To remove these noises, imbalanced data classification can be used.

Keywords: DOG filtering, spectral angle, Morphological operations, imbalanced data classification

I. INTRODUCTION

Recognition of roads is very important in the field of civilian and military applications which includes emergency planning systems for evacuation and fire response, navigation or location aware systems etc. Automated road extraction can save both time and labor to build and update the road spatial database in such applications. Fully automated algorithms for recognize roads with greater accuracy is not currently available. Higher resolution images provides greater accuracy but it has greater computational and complexity due to noise.

The main problem of these methods is the difficulty to provide the best parameters for a given image. Given a new image, the user often has to try different combinations before obtaining a satisfactory result. The computation time is also often a problem. As the resolution of optical sensors increases, size of images increases as well, thus increasing the computation time necessary to process a scene.

II. PROPOSED METHOD

The paper proposes a method for road detection in satellite images in a complex scenario. The input images are multi spectral images. Three or more spectral bands contains in a multispectral image.

A. Spectral Angle

The color information can be used to distinguish road from its surroundings [1]. The color information are available in spectral bands and it can be interpreted as a vector. First step is to convert the vector image to scalar image using spectral angle. The spectral angle is calculated using Eq. (1).

$$A = cos^{-1}(\sum_{b=1}^{n_b} r(b).c(b)/\sqrt{\sum_{b=1}^{n_b} r(b)^2.\sum_{b=1}^{n_b} c(b)^2}) \quad (1)$$

b represents spectral band, r represents reference pixel and c represents current pixel.

The resultant image contains roads in darker color and then it is converted to a binary image based on a threshold value.

B. DOG Filtering

DOG filtering [2] is applied to the given image using Eq. (2)

$$g_\sigma(y) = \frac{-y}{\sqrt{2\pi\sigma^3}} e^{\frac{y^2}{2\sigma^2}} \quad (2)$$

σ represents standard deviation of the Gaussian function. The resultant binary image may contain irrelevant information. So this image is subjected to morphological operations.

C. Morphological Operations

Morphological operations are used to apply changes in the properties of shapes of the areas of image. These operations are defined on sets. Different morphological operations like cleaning, majority, filling are applied on the given image in order to remove noise. Cleaning operation remove isolated pixels ie,1's surrounded by 0's. Majority operation set a pixel to 1 if five or more pixels in its 3-by-3 neighborhood are 1's. Filling operation fill isolated interior pixels ie,0's surrounded by 1's. At this stage there may exist some discontinuity between the thinned roads due to noise. Therefore dilation and erosion is applied to the binary image.

Dilation is an operation that thickens objects in a binary image. Dilation fills the small holes and connects the disjoint objects in an image. It is controlled by the structuring element. The shape and extent of thickness is specified by

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 11, November 2013

structuring element. The shape can be line, ball, disk, diamond, arbitrary, periodic line, rectangle and octagon. The operation performed in following manner. The structuring element slides over the image. There is no change if the origin of the structuring element coincides with the white pixel in the image and moves to next pixel. If it coincides with a black pixel in the image, it makes all pixels from the image covered by the structuring element to back color.

Erosion shrinks objects in a binary image. It is also controlled by the structuring element. In this operation, the structuring element slides over the image. There is no change if the origin of the structuring element coincides with the white pixel in the image and moves to next pixel. If it coincides with a black pixel in the image, it at least one of the 'black' pixels in the structuring element falls over a white pixel in the image, then change the 'black' pixel in the image to a white.

D. Area Based Filtering

Small regions that do not belong to any road can be removed by area based filtering [3]. The binary image pixels are inverted so that road pixels are represented by 1 and non-road pixels is represented by 0. The region whose area is less than a threshold is then removed from the image. But in case of very complex scenarios, there may be some noise and artifacts present in resultant images. These images are the subjected to imbalanced data classification.

E. Imbalanced Data Classification

In imbalanced data classification, at least one class is under represent relative to others [4]. Because of this unequal class distribution, the performance of the existing classifiers tends to be biased towards the majority class. The existing classification algorithms are having poor performance on imbalanced data sets.

In this step, first extract road features and building features from each image in the data set. Since the road class is under-representative to building class, these features should be submitted to sampling technique.

Sampling approaches are applied on the data to change the class distribution of data and make it balanced. The two types of sampling approaches are : Undersampling and Oversampling. Undersampling method removes examples from the majority class to make the data set balanced. This approach works in two steps. In the first step, minority samples are clustered and then identify the majority class samples that are out of minority classification regions. This improves minority prediction rate, in the second step, majority samples are randomly selected in individual clusters and this enhances majority prediction rate.

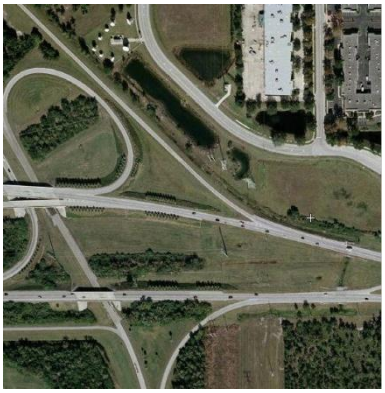
Oversampling is a sampling approach which balances the data set by replicating the examples of minority class. There is no loss of data as in this technique. But it may lead to overfitting and can introduce an additional computational cost if the data set is already fairly large but imbalanced.

Synthetic Minority Oversampling Technique (SMOTE) is an oversampling approach. In this method, the minority class is oversampled by creating synthetic examples rather than by oversampling with replacement. In SMOTE [5], minority class is oversampled by taking each minority class sample and introducing synthetic examples along the line segments joining any/all of the k minority class nearest neighbors.

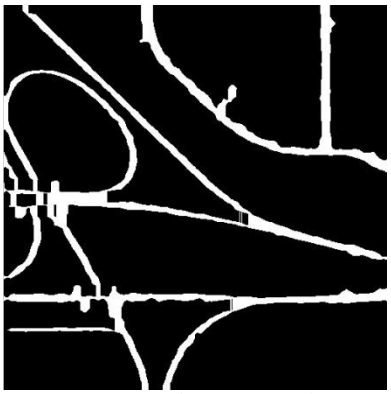
After applying smote algorithm, the balanced feature set are given to C4.5 classifier, decision tree, SVM classifier and display those features corresponding to road label. Finally select the classifier having greater accuracy.

III. **RESULTS**

Figure 2 and figure 3 shows results of some of the images. All the roads are detected in the results of first image, but some noises are present in the results of second image. Therefore second image should be subjected to imbalanced data classification.



(a) Input Image

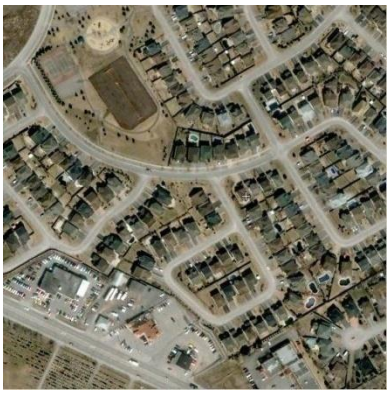


(b) Image after area based filtering

Fig.2. Results of first image

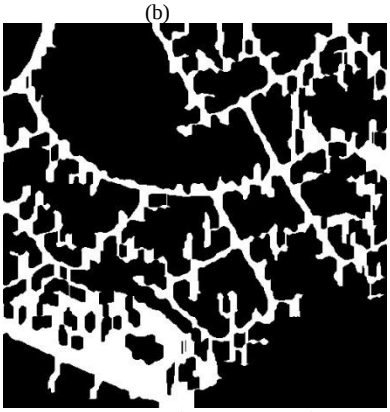
ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 11, November 2013



(a) Input image

(b)



(c) Image after area based filtering

Fig.3. Results of second image

IV. CONCLUSION

Road network extraction is an important, but challenging, problem to solve. There are many civilian, commercial, and military applications for this problem including the determination of existence of roads after a natural or man-made disaster. The use of morphological operations provides results with greater accuracy for sub urban images. This method is limited to a specific type of image and a particular geographic area. In order to improve the accuracy of the results for highly complex images, imbalanced image classification method can be performed.

REFERENCES

[1] Emmanuel Christophe and Jordi Inglada, "Robust Road Extraction for High Resolution Satellite Images", ICIP, 2007.

[2] Qin Li, Lei Zhang, Jane You, and David Zhang., "Dark Line Detection with Line Width Extraction", ICIP,2008

[3] D. Chaudhuri, N. K. Kushwaha, and A. Samal, "Semi-Automated Road Detection Using High Resolution Satellite Images by Directional Morphological Enhancement and Segmentation Techniques," , IEEE Journal Of Selected Topics In Applied Earth Observations And Remote Sensing, Vol. 5, No. 5, October 2012.

[4] Arun Kumar M.N and H S Sheshadri. ," On the Classification of Imbalanced Datasets", International Journal of Computer Applications 44(8):1-7, April 2012.

[5] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, W. Philip Kegelmeyer, "SMOTE: Synthetic Minority Over-Sampling Technique", Journal of Artificial Intelligence Research, 2002.