IOSR Journal of Engineering (IOSR JEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
PP 16-20

Comparison between Random Forest and Decision Tree Algorithm to Assessing The Vegetation Area from Satellite Images

P.Sathya[1], Dr.V. Baby Deepa[2]
Department of computer science, Research Scholar, Government Arts College, Karur, Tamilnadu.
Department of computer science, Assistant Professor, Government Arts College, Karur, Tamilnadu.

Abstract: *The large damage after effect of a natural tragedy is not a simple task and in fact it is a inconvenient process to scope the real consequences with accuracy. This paper deals with estimating the disastrous impact regarding the loss of the vegetation. The random forest and decision tree classifiers are used to extract the green areas before and after the cyclone to determine the unite loss. The random forest approach is compared with the decision tree algorithms and from the experimental results the random forest algorithm is performed and produced high accuracy.*
Keywords*: Classifier, segmentation, classification, accuracy.*

I. Introduction

Pixel based classification and object based classifications are used to land classifications in remote sensing. These two techniques are mostly used to classify land surfaces and in this paper the object based method is used. Most of the researchers used random forest technique to produce high quality with maximum accuracy.

The random forest technique is commonly called non-parametric algorithm effective of handling large amount of datasets and the random forest produce different decision tree results and then evaluate the average to apply the prediction. The accuracy level is better by the random forest technique uses the conclusion of large amount of prediction created by the trained classifiers.

Natural adversities change the very nature of the environment acutely and dangers like cyclone, tsunami, flood, landslide and soil erosion causes great damage to the environment. The above mentioned factors can seriously affect the human lives as they constantly intercede in such locality and tragically lose their maintenance. This leaves the research support to accurately predict and level the danger associated with these natural calamities and gets the humans informed.

II. Image Segmantation

The image segmentation is the important and first step of image classification. Segmentation is used to plainly narrate the shape and measurement of the image that are being used in classification. The input image acquire is originallyneed to pre-processed to be classified with the perfect accuracy. In order to manage, extract, and determine data from the satellite images it is quite authoritative to develop automated image processing techniques. Here in this section a new approach is proposed to extract and find the region of interest in the satellite image.

Practically all images related projects has the ability to process a subset of the pixels in the input image and trusting upon the errand, the selected pixels might be a reasonably irregular area of the information or a normal sub-picture of the info picture arranged. In this way the ROI is fundamentally named as a subset of a picture bare for a particular reason. The edges present in the images are first exposed using canny method. The Canny edge detector is an excellent technique which utilizes multistage algorithm to determine a broad variety of edges in a noisy image.

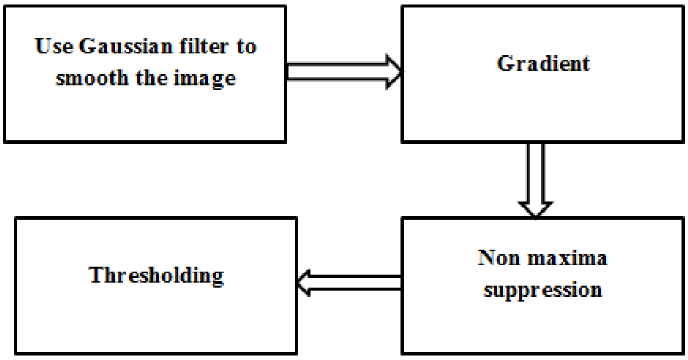

Figure 1: Work flow of canny detector

Smoothening the image with Gaussian filter is used to eliminate the noise and the formula for that is
$S[i, j] = G[i, j, \varsigma] \times I[i, j]$
Where $I[i, j]$ is the input image, $G[i, j, \varsigma]$ is Gaussian filter, ς is spread of Gaussian, $S[i, j]$ is the smoothened output image.
Gradient of the smoothened image is calculated by
$M[i, j] = (S[i, j+1] - S[i, j] + S[i+1, j+1] - S[i+1, j]) / 2$
$N[i, j] = (S[i, j] - S[i+1, j] + S[i, j+1] - S[i+1, j+1]) / 2$

Where $M[i, j]$ is gradient in x direction, $N[i, j]$ gradient in y direction. After detecting the edges, the multi-resolution algorithm is realistic to find the types and separate the objects present in the input satellite image. Most of the research groups [Trimble, 2010], [Liu and Xia,2010], [Myint et al. 2011]used three main input parameters in multi-resolution segmentation algorithm namely, shape/color, compactness/smoothness and scale.
The table 1 showcases the user input values provided and the number of objects acquired at each level of segmentation.

Segmentation	User Scale	Shape/Color	Compactness/ Smoothness	Objects identified
1	10	0.1/0.9	0.5/0.5	412
2	20	0.1/0.9	0.5/0.5	2317
3	50	0.1/0.9	0.5/0.5	7871
4	60	0.1/0.9	0.5/0.5	17991

Table 1: Image segmentation parameters used

I. EXTRACTING IMAGE FEATURES

The finest and optimum way to select the features existent in the input images attained using multi-resolution segmentation algorithm is arrived by making use of the preceding knowledge about the image and essentially from the end user's perception knowledge. Also the existing feature selection algorithms [Genuer et al, 2010], [Yu et al., 2006] can be engaged to cater to the need of the problem and address animportant solution. The type of features extracted from the objects is enumerated in the table 2.

Features	Description of the feature
Mean	Provides the mean values of the object for a particular input image.
Standard deviation	Provides the SD values of the object for a particular input image.
Mean difference to neighbors	Provides the difference between the mean value of the object for a particular input and its super object.
Mean of sub-objects	Provides SD of the mean value of sub-object to its super objects.

Table 2: The object features used in classification

The pre-processing works are initially carried out and the canny operator applied images are shown in the figure 1.

Figure 1: Pre-processing and segmentation

The classifiers consume the radial basis function (RBF) and the results of the unclassified pixels of the two algorithms are showcased in the table3. Since RBF achieved better than linear and quadratic functions, the random forestworks and outperformed the other algorithm.

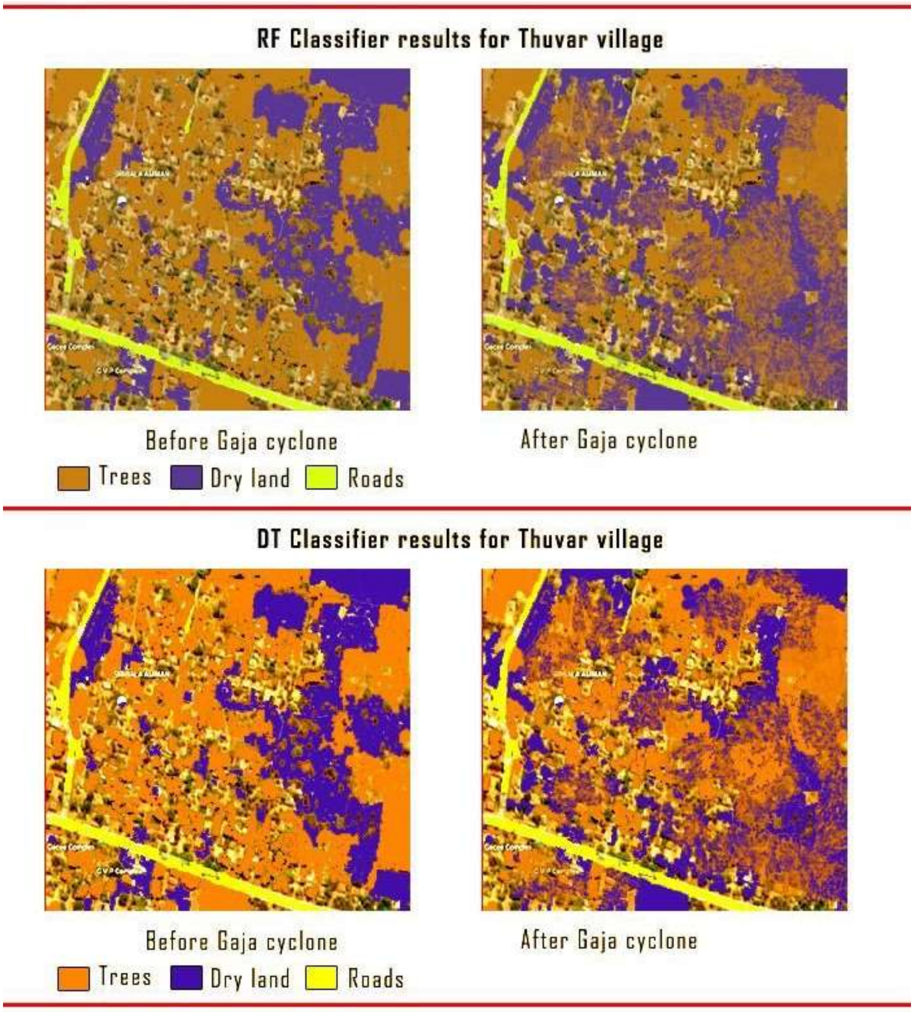

Figure 2: Results related to Thuvar village

The figures 1, 2 and 3 showcased the classification results of the locationmeasured for the training and from the results the accuracy and the correctness are calculated using evaluation metrics and the unclassified pixels are initially compared and showcased in the table 3.

Overall data (Thuvar village image)		
Algorithm	**Unclassified pixels**	**Mixed pixels**
Random Forest	76129	12387
Decision Tree	81009	11090

Table 3: Summary of unclassified and mixed pixels

II. ACCURACY ASSESSMENT

The accuracy assessment of the work is performed using the confusion matrix and kappa statistics. The accuracy levels are assimilated by calculating the ratio of the user accuracy with the producer accuracy. The overall performances are calculated by the following formula

$$\textbf{User Accuracy Ua} = \frac{qii}{\sum_{j=1}^{N} qij}$$

$$\textbf{Producer Accuracy Pa} = \frac{1}{N} \sum_{i=1}^{N} Ua$$

$$\textbf{Overall Accuracy N} = \frac{1}{N_O} \sum_{i=1}^{N} qii$$

Here, qij is the matrix which declares that how many samples belongs to class *i* and classified into class *j*. No is the total number of input images in the research. The kappa coefficient a popular statistical measure often engaged for merging and categorizing the similar classes.

$$\textbf{Kappa Coefficient k} = \frac{\textbf{Observed Accuracy - } \varepsilon}{\textbf{1 - } \varepsilon}$$

$$\varepsilon = \frac{\sum_{i=1}^{N} (\textit{Classified pixels} - \textit{Actual classified pixels}}{\textbf{Total Pixels}^2}$$

The overall accuracies are calculated for pixel based and object based approaches and the values are plotted as shown in the figure 3.

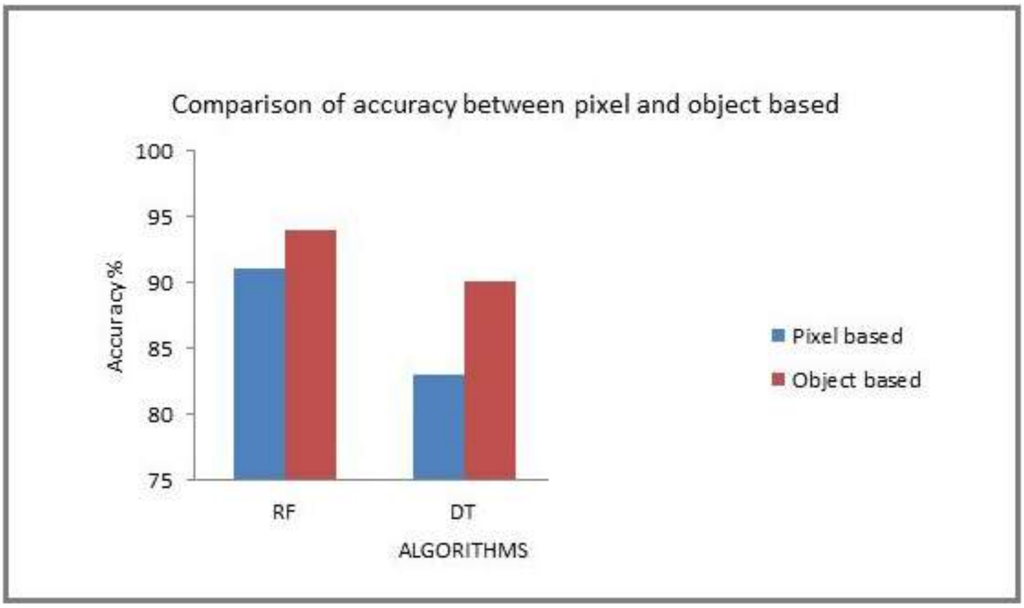

Figure 3: Accuracy comparison of pixel based and object based classification

The accuracy assessment between the pixel and object based is supported out using repetitive k-fold validation test and the results obviously showcased that object based classification is greater in the two algorithms and the Decision tree which employed RBF kernel outscored the other algorithm with respect to the accuracy levels. The authentic loss of trees due to the cyclone is perfectlyexposed by the DT classifier.

The investigates and evaluations are directed using MATLAB and as predictable both the pixel and object based methods progressed well in spite of the fact that the object based method slightly performed better than the pixel based method and offered clear and more universal visuals when compared with pixel based. When the complete accuracy levels are compared, there is almost 2 to 4 percent increase in the accuracy levels of the object based approach and there by the optimum approach that can be employed in these conditions can be object based approach.

The random forest algorithm almost matched the SVM on certain occasions with respect to classification accuracies but the decision tree algorithm protected behind both SVM and RF algorithms.

III. Conclusion

Classification of satellite images using pixel and object based is carried out on two machine learning algorithms and from the results the accuracies of two algorithms are almost the same when employed pixel based technique but when the algorithms engaged object based method, the DT and RF displayed lot of development with respect to the accuracy. The DT classifier reached the highest accuracy when compared with the random forest algorithms and when it consumed object based approach, it performed the best and with RBF kernel the Dt outscored the other algorithm by a reasonable margin.

References

[1]. BREIMAN, L. (2001). Random Forests. *Machine Learning*, 45, 5–32.
[2]. Brenning.A, Benchmarking classifiers to optimally integrate terrain analysis and multispectral remote sensing in automatic rock glacier detection, Remote sensing environment, 113(1),239-247, 2009.
[3]. Genuer.R, Poggi, and Tuleau- variable selection using random forst, pattern recognition letters, 31, pp,2225-2236.
[4]. Yu.Q, Gong.P, Clinton.N,Biging.G, 2006, Object base3d detailed vegetation classification with airborne high spatial resolution remote sensing imagery, Photogrammetric engineering and remote sensing, 72,pp.799-811, 2006.
[5]. TRIMBLE (2010A). eCognition® Developer 8.64.0 reference book. Available at:. http://www.definiens.com/ .
[6]. Liu.D, Xia.F. 2010, Assessing object based classification :advantages and likitations. Remote sensing letters,1,pp.187-194.
[7]. Myint.S, Gober.P, Brazel.A,Clarke.S and Weng.Q 2011, Per-pixel vs object based classification of urban land cover extraction using high spatial resolution imagery, *Remote Sensing of Environment*, 115, pp. 1145-11.