HSV Color Histogram based Land Cover Classification of Remotely Sensed Data using different Classifiers

[1] M. Christy Rama, [2] D. S. Mahendran, [3] T. C. Raja Kumar
[1] Research Scholar, Manonmaniam Sundaranar University, Abishekpatti, Tirunelveli-627 012, Tamil Nadu, India.
[2] Principal, Aditanar College of Arts and Science, Tiruchendur, Tamil Nadu, India
[3]Associate Professor, Department of Computer Science, St. Xavier's College, Tirunelveli, Tamil Nadu, India.
[1] christyrama@gmail.com, [2] dsmahe65@yahoo.com, [3] rajazion@yahoo.co.in

Abstract

Land cover classification of remotely sensed images is the basis for many environmental and socio economic applications. Feature extraction plays a vital role in multispectral remote sensing image classification before classifying the image. In this paper, color features are extracted using HSV color histogram model to classify the land cover. The extracted features are trained and tested by supervised classifiers such as K Nearest neighbor, Decision Tree, RUSBoost, Random Forest, Support vector machine and Naive Bayes classifiers. The performance of the classifiers is compared based on the metrics accuracy and kappa coefficient. An IRS LISS IV orthorectified dataset is chosen as the input image for this experiment. It is observed that the random forest classifier outperformed when compared to other classifiers.

1. Introduction

Land cover refers the natural features and artificial constructions covering the land surface. The geospatial phenomena are varying over time and the land cover data has to be up-date periodically. Up-to-date land cover knowledge is an important factor for the various planning authorities with responsibilities for the management of territory [1] [2]. Planners and land managers need accurate information to address land cover problems. Remote sensing is an important tool in wide areas of environmental research and planning. In particular the classification of spectral images has been a successful application for deriving land cover maps, assessing deforestation, real time fire detection, estimating crop acreage and production, monitoring of environmental pollution etc [3-6]. The overall objective of land cover image classification method is to automatically categorize all pixels in an image into different land cover classes. The classification algorithms are very essential for the success of the land cover classification process. A large number of classification algorithms have been developed and applied for classifying remotely sensed data [7]. The classification accuracy depends on the classifier(s) chosen. Some of the most commonly used supervised parametric and non-parametric classification approaches are Baysian Classifier, Decision trees, Support vector machines, K Nearest Neighbor Classifier, RUSBoost Classifier and Random Forest classifier etc [8, 9].

2. Frame Work

A generic framework of classification methodology using supervised parametric and non parametric machine learning-based classifiers is given in Figure 1. This methodology is performed in two phases: a training phase and a testing phase. In the training phase, a classifier is trained using a set of historical (or training) data. Color feature vector extracted from the HSV color space model of training samples is used to train the various classifiers such as Baysian Classifier, Decision trees, Support vector machines, K

Nearest Neighbor Classifier, RUSBoost Classifier and Random Forest classifier. The performance of the trained classifier is evaluated using a set of unobserved test signals in the testing phase. The classifier is loaded with the particular features selected in the training phase and features extracted from the test signals as shown in Figure-1. The classifiers return the class labels based on the prior learning of training samples and the classification accuracy depends on the selection of features. A feature vector of 81 bins generated by the HSV color model based Histogram is used as a feature vector to classify the image.

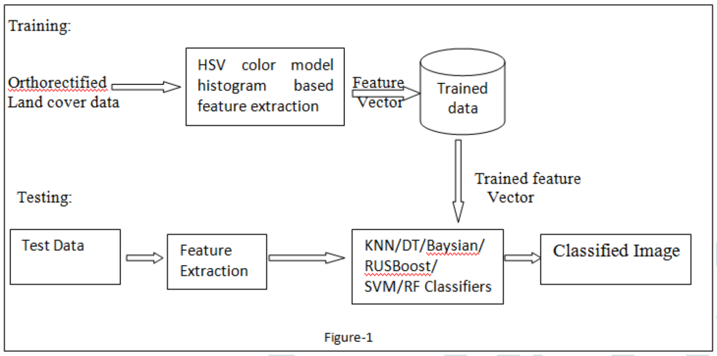

Fig I Architectural model of the land cover classification method

3. Classifiers

The most commonly used supervised parametric and non parametric classifiers applied in this classification work are discussed here.

3.1 k- Nearest Neighbor

A k Nearest Neighbor (KNN) is a simple non parametric classifier that takes a decision about the test data by simply considering the k-nearest points in the training set. The classifier counts the members from each class in this set of k-nearest neighbors to test the input and classifies the test signal in the class having the highest members. The Euclidean distance is normally used as a distance metric to obtain the set of the nearest neighbors. However, kNN has several shortcomings such as its performance highly depends on the selection of k and pooling nearest neighbors from training data that contain overlapping classes is considered unsuitable [10,11].

3.2 RUSBoost

RUSBoost eliminates the data distribution imbalances between the classes and improve the classification performance of the weak classifiers. RUSBoost is a hybrid data sampling and boosting algorithm. RUS (Random under Sampling) in RUSBoost method refers to random data extraction which means the method randomly deletes data from the training data set until the required balanced class distribution is achieved. RUSBoost is also considered as a kind of advanced data sampling technique. Boosting can be performed by re-weighting and re-sampling. Re-sampling is done according to the weights given to samples included in the training data set. RUSBoost method has both advantages and disadvantages. Data extraction in data distribution until the desired balance is achieved is a disadvantage where as the reduction of the time required for model training is an advantage [12, 13].

3.3 Decision Tree Classifier

A decision tree classifier has a simple form that can be compactly stored and efficiently classifies new data. DT classifiers can perform automatic feature selection and complexity reduction and their tree structure provides easily understandable and interpretable information regarding the predictive or generalization capability of the classification. To construct a classification tree by heuristic approach it must be assumed that a data set consisting of feature vectors and their corresponding class labels are available. The features are identified based on the specific knowledge that we have from the problem in hand. Decision tree techniques have been used successfully for a wide spectrum of classification problems in various fields. They are computationally efficient and flexible and also have an intuitive simplicity. They therefore have substantial advantages in remote sensing applications. One of the simplest alternatives to traditional classification systems is decision tree classification. The basis of this approach is establishing a set of binary rules that are applied sequentially to discriminate between different target categories. Those rules include thresholds on spectral bands but also on auxiliary information such as soil maps, slope or digital elevation model, and therefore are very flexible to different types of input data [14].

3.4 SVM Classifier

Support Vector Machine (SVM) is a supervised nonparametric statistical learning method. The SVM aims to identify a hyper-plane that separates training samples into predefined number of classes [14]. In the simplest form, SVMs are binary classifiers that assigns the given test sample to one of the two possible classes. The SVM algorithm is extended to nonlinearly separable classes by mapping samples in the feature space to a higher dimensional feature space using a kernel function. SVMs are particularly appealing in remote sensing field due to their ability to successfully handle small training datasets often producing higher classification accuracy than traditional methods [15, 16].

3.5 Random Forest Classifier

Random Forest is based on tree classifiers. Random Forest grows many classification trees. To classify a new feature vector, the input vector is classified with each of trees in the forest. Each tree makes a classification and one can say that the tree "votes" for that class. The forest chooses the classification having the most votes over all the trees in the forest. Random Forest provides unexcelled accuracy among current algorithms, efficient implementation on large data sets, and an easily saved structure for future use of pre-generated trees. Random forest (RF) is becoming increasingly popular in ALS remote sensing [17, 18].

3.6 Naive Bayes Classifier

The Naive Bayes (NB) classifier is a collection of independent classifiers based on Bayesian probability. The assumption is that the relation with the response classes is independent for every predictor. Therefore, instead of searching for the Maximum likelihood (ML) classifier hyper planes, a separate classifier is deducted for each predictor variable. Otherwise, NB is very similar to ML, in the sense that it is based on adjusting Gaussian kernels to the training data. The final hard classification in NB is based on selecting the classes with higher posterior probability. Only few studies have tested the ability of NB classification using remote sensing data [19].

4. HSV color space Histogram based Color features

Color features are normally represented by the color histogram. HSV color space histogram is used in this work since it is more perceptually uniform than other color spaces. The color conversion from RGB color space is HSV color space is performed before extracting features. The H, S and V values of HSV color space are combined together as given below (Eqn 1) and a feature vector of 81 bins is generated [20].

$$H=\begin{cases} 0 & h \in [340, 20] \\ 1 & h \in [20, 50] \\ 2 & h \in [50, 75] \\ 3 & h \in [75,140] \\ 4 & h \in [140,160] \\ 5 & h \in [160,195] \\ 6 & h \in [195, 285] \\ 7 & h \in [285,305] \\ 8 & h \in [305, 340] \end{cases} \qquad S=\begin{cases} 0 & s \in [0, 0.2] \\ 1 & s \in [0.2, 0.65] \\ 2 & s \in [0.65, 1] \end{cases} \qquad V=\begin{cases} 0 & v \in [0, 0.2] \\ 1 & v \in [0.2, 0.7] \\ 2 & v \in [0.7, 1] \end{cases} \tag{1}$$

5. Study area and Dataset

The test is experimented on a remotely sensed IRS (Resouresat2 satellite) LISS-IV (sensor) orthorectified image supplied by National Remote Sensing Centre (NRSC), Hyderabad, Government of India. The image has been taken in January 2012 and is of size 552X414 pixels. LISS-IV image has a spatial resolution of 5.8m. For study purpose, bands 2, 3 and 4 of LISS-IV data (Green, red and near IR respectively) are combined together to form a RGB image. The remotely sensed image covers the area in and around Nagercoil city located in kanyakumari district in the state of Tamil Nadu in India. The ground truth of the study area has been taken from ENVI.

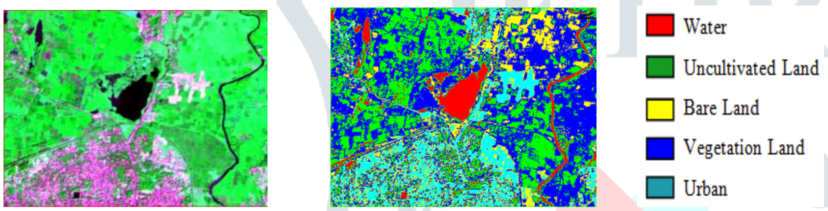

Fig-II IRS RGB image and its Labeled Ground Truth

6. Performance Metrics

The classification based on HSV color model based histogram analyzed using various supervised classifiers such as Baysian Classifier, Decision trees, Support vector machines, K Nearest Neighbor Classifier, RUSBoost Classifier and Random Forest classifiers. The performance of these classification methods are analyzed using the metrics accuracy and Kappa coefficient and are described in the following tables (Table-I & Table II). The Comparison between all these classifications methods on the IRS dataset is also represented in the form of graphs (Fig-III & Fig IV). From the experiments it is proved that the Random Forest classifier yields higher classification accuracy and outperforms other classifiers taken for study based on various metrics.

Table-1 classification accuracies of various supervised classifiers

CLASS	KNN	Baysian	SVM	RUSBoost	Decision Tree	Random Forest
Water	0.941176471	0.902439024	0.960784314	0.970588235	0.970588235	0.970588235
Uncultivated land	0.823529412	0.658536585	0.901960784	0.882352941	0.970588235	0.970588235
Ground	0.81372549	0.658536585	0.794117647	0.803921569	0.941176471	0.990196078
Vegetation land	0.941176471	0.853658537	0.921568627	0.87254902	0.911764706	0.990196078
Urban	0.970588235	0.707317073	0.931372549	0.960784314	0.970588235	1
Average Accuracy	0.898039216	0.756097561	0.901960784	0.898039216	0.952941176	0.984313725

Table-2 Kappa coefficient of various supervised classifiers

CLASSIFIER	KNN	Baysian	SVM	RUSBoost	Decision Tree	Random Forest
KAPPA	0.646116894	0.254002911	0.665880503	0.654597551	0.843016545	0.947184466

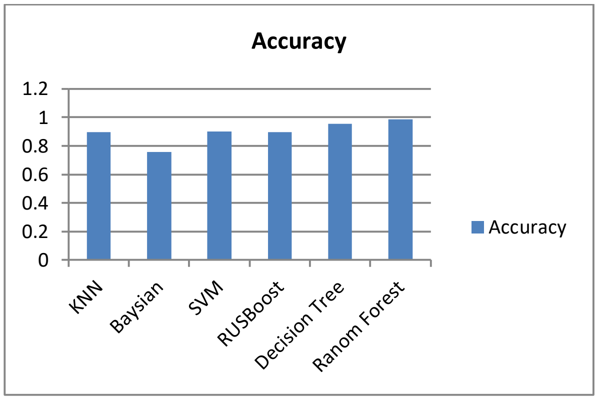

Fig-III comparison of classification accuracies of various classifiers

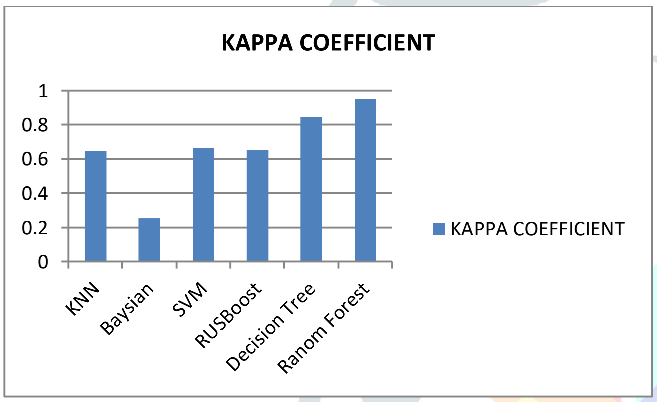

Fig-IV comparison of Kappa coefficient various classifiers

7. Conclusion

In this research we analyzed the classification of IRS dataset using various supervised parametric and non parametric classifiers. The classification task is modeled as a 81-feature, 5-class classification problem and it is observed that from Table 1 & 2 that the performance of parametric Random Forest classifier yields better result when compared to all other classifiers in terms of overall accuracy and kappa coefficient.

References

[1] A.R.S Marçal, J.S Borges, J.A. Gomes and J.F Pinto Da Costa, "Land cover update by supervised classification of segmented Aster images",International journal of remote sensing, 26(7), pp.1347-362,2005

[2] Hamid Reza Matinfar and Majid Shadman Roodposhti, "Decision tree land Use/ Land cover Change detection of Khoram Abad City using Landsat Imagery and Ancillary SRTM Data", Annals of Biological Research, Vol.3, No.8, pp.4045- 4053, 2012.

[3] S.L Powell, W.B.Cohen, Z.Yang, J.D Pierce and M.Alberti , "Quantification of impervious surface in the Snohomish water resources inventory area of Western Washington from 1972-2006", Remote Sens. Environ., *Vol.112, pp.1895-1908, 2008.*

[4] P.K.K Joshi, P.S.Roy, S.Singh, S.Agrawal and D.Yadav, " Vegetation cover mapping in India using multi-temporal IRS Wide Field Sensor (WiFS) data", *Remote Sens. Environ., Vol.103*, pp.190–202, 2006.

[5] P.Potapov , M. C.Hansen, S.V.Stehman, T.R.Loveland and K.Pittman, "Combining MODIS and Landsat imagery to estimate and map boreal forest cover loss", Remote Sens. *Environ., Vol.112*, pp.3708–3719, 2008.

[6] P.E.Dennison and D.A.Roberts, "Daytime fire detection using airborne hyperspectral data", Remote Sens.Environ., *Vol.113*, pp.1646–1657, 2009.

[7] Chiming Chang, Paul-Armand Verhaegen and Joost R.Duflou, “A Comparison of Classifiers for Intelligent Machine Usage Prediction”, IEEE Transaction Machine Learning, 2014.

[8] Sana Ullah Jan and Insoo Koo, “A Novel Feature Selection Scheme and a Diversified-Input SVM-Based Classifier for Sensor Fault Classification” ,Journal of Sensors , 2018 https://doi.org/10.1155/2018/7467418

[9] P. Thanh Noi and M. Kappas, “Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery”, Sensors, vol.18, no. 2, 2018.

[10] H.W.Zhu, O.Basir, “An adaptive fuzzy evidential nearest neighbor formulation for classifying remote sensing images”, IEEE Trans.on Geosci. Remote Sens., vol.*43*, pp.1874–1889, 2005.

[11] J.Yang, Z. Sun, and Y. Chen, “Fault detection using the clustering-kNN rule for gas sensor arrays,” Sensors, vol.6, No.12, 2016.

[12] M. H. Kesikoglu , U. H. Atasever , C. Ozkan and E.Besdok, “ The usage of rusboost boosting method for classification of impervious surfaces” **,** The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XLI-B7, 2016 XXIII ISPRS Congress, July 2016.

[13] C.Seiffert, T.M.Khoshgoftaar, J.V.Hulse and A.Napolitano , “ Rusboost: A hybrid approach to alleviating class imbalance”, IEEE Transactions On Systems Man and Cybernetics”, 40(1), pp.185-197,2010

[14] P.M. Pal and P.M.Mather, “An assessment of the effectiveness of decision tree methods for land cover classification”, Remote Sensing of Environment, Vol.86, pp.554-565, 2003.

[15] G. Mountrakis, and C. Ogole, “Support vector machines in remote sensing: A Review:”, Int. Journal of Photogrammetry and Remote Sensing, vol. 60, pp 247-259,2011.

[16] P. Mantero, G. Moser, S. B. Serpico, “ Partially supervised classification of remote sensing images through – SVM-based probability density estimation”, IEEE Transaction on Geoscience and Remote Sensing, vol. 43, no. 3, pp.559- 570,2005.

[17] B. Lowe and A. D. Kulkarni, “Multispectral image analysis using Random Forest”, International Journal on Soft Computing (IJSC), vol. 6, no.1, pp. 1-14,2015.

[18] Arun D. Kulkarni and Barrett Lowe, “Random Forest Algorithm for Land Cover Classification”, International Journal on Recent and Innovation Trends in Computing and Communication, Vol. 4 Issue: 3, 2016.

[19] J. García-Gutiérrez, F.Martínez-Álvarez, J. C .Riquelme, “ Using Remote Data Mining on LIDAR and Imagery Fusion Data to Develop Land Cover Maps”, Trends in Applied Intelligent Systems. Lecture Notes in Computer Science, 6096, pp.378-387, 2010.

[20] Nishant Shrivastava and Vipin Tyagi, “An efficient technique for retrieval of color images in large databases”, Computers and Electrical Engineering, Elsevier, 2014.