International Journal of Engineering & Technology, 7 (4) (2018) 3070-3073

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet.v7i4.16382
Research paper

Implementation and comparison of classifiers for different hyperspectral dataset based on machine learning algorithms

Kishore M [1] *, S. B. Kulkarni [2]

[1] *Research Scholar, Dept. of EE, VTU-RRC, Belagavi, India*
[2] *Associate Professor and Head, Dept. of CSE, SDMCET, Dharwad, India*
**Corresponding author E-mail:kishore.energy@gmail.com*

Abstract

In these paper better classification accuracy techniques is proposed for flower and land cover hyperspectral dataset. Initially flower dataset is considered in which newly proposed improved particle swarm optimization is implemented and compared with particle swarm optimization and K means algorithm followed by land cover dataset is considered in which proposed random forest algorithm is compared with support vector machine and k means and Navie Bayes classifiers. In both the hyperspectral dataset proposed methods gives good classification results in terms of accuracy.

Keywords:*Hyperspectral Image; Classifiers; Machine learning.*

1. Introduction

Hyperspectral imaging collects and processes dataacross the electromagnetic spectrum, which cannot be achievedby other imaging techniques.The goal of hyperspectral imaging is to obtain the spectrum for each pixel in the image of a scene to find objects and classify. Unlike the human eye, that sees visible light in three bands (red, green, and blue), spectral imaging divides the spectrum into many more bands. This technique of dividing images into bands can be extended beyond the visible. In hyperspectral imaging, the recorded spectra have fine wavelength resolution and cover a wide range of wavelengths. Therefore, Engineers build hyperspectral sensors and processing systems for applications in astronomy, agriculture, biomedical imaging, geosciences, physics, and surveillance. Certain objects leave unique 'fingerprints' in the electromagnetic spectrum known as spectral signatures; these 'fingerprints' enable identification of the unique features that make up a scanned object.

Hyperspectral (HS) imaging has a wide range of applications in terrain classification, mineral detection and exploration, pharmaceutical counterfeiting, environmental monitoring and military surveillance. Instrument limitation and imperfect imaging optics do not allow HS sensors to acquire high-spatial resolution data. On the other hand, a low-spatial resolution will result in many mixed pixels and greatly degrade the detection and recognition performance required in the civil and military fields. Therefore, enhancing the spatial resolution of HS data has become an important area of research. In HSI classification, sufficient training samples are usually crucial to obtain reliable results due to the curse of dimensionality, i.e. Hughes phenomenon. Manifold learning promotes dimensionality reduction of intelligent data analysis. Hyperspectral image processing has been an area of active research owing to the large number of applications associated with it and large datasets, which urge the development of less-complex and more accurate processing mechanisms to deal with the high-dimensional data. Hyperspectral image (HSI), satellite sensors collect imagery simultaneously in hundreds of narrow and contiguously spaced spectral bands, with wavelengths ranging from the visible spectrum to the infrared region (0.4–2.5μm).

The organization of the paper is as follows. Section 1 introduces the importance of use of hyperspectral imaging in the assessment of land covers, followed by some of the existing works in Section 2. The details of proposed methodology have been covered in Section 3 followed by Discussion in Section 4. Results obtained and comparative analysis has been provided in Section 5 with Conclusion in Section 6.

2. Literature review

The effect of urbanization over the years by human beings havehad a negative impact on the sustainability of the environment. With the development of roads, buildings, water parks, commercial units there is a threat to environment because of massive deforestation, extinction of wildlife, reduction of wetlands and yielding lands. The major reason behind this is the change in living standards of human beings over the years without any measures being taken for reducing and eliminating the hazardous practices undertaken. The survey conducted by Vitousek, P.M el.at [1] has highlighted the increasing amount of changes on land cover day by day due to the unethical and harmful life style practices in the environment. These practices not only are harmful to the human beings but have an ill effect on the natural flora and fauna resulting in endangering the existing species. The deterioration of natural resources by the act of humans has had an ill effect on the environment there by creating an imbalance in the ecosystem [2]. In order to overcome the environmental issues knowledge about land use and changes occurring must be kept intact so as to create a safe ambient environment [3].

When we consider an outside environment, there are several factors like illumination factors, background in which the image is present and these affect the quality of images being captured. Image segmentation is carried out mainly to divide the image into multiple regions and further extract such regions, which are of interest. The segmentation process proposed by S L S Abdullah et.al [4] aims at the production of a threshold before segmentation so as to support a uniform illumination effect on the image used. The use of Otsu's method for generation of threshold and Fuzzy c Means for segmentation have still not been able to produce an image of a better quality due to poor illumination efforts and improper backgrounds. The data after segmentation undergoes the process of feature extraction so as to identify the differences in minute patterns by using a multivariate texture model [5]. This model was able to provide four distinct levels of output thereby providing better classification of land covers.
The authors in [8] obtained compared features using different filtering strategies for morphological attribute filters by considering non-increasing attributes. The Attribute Profiles (APs) and Self Dual Attribute Profiles (SDAPs) were obtained by sequentially applying attribute filters on tree-based image representations, such as Min- or Max-trees and Inclusion tree. The main objective of this study was to investigate the effects of filtering rules max, min, direct and subtractive, when considering the non-increasing attributes of moment of inertia and standard deviation. By using a high-spatial resolution dataset, the extracted information from the profiles has been analyzed. This is done by studying the effects on the classification accuracy by using the profiles as additional input features to a Random Forest classifier.
Rebetez et al [9] have studied the problem of domain adaptation in the context of hyperspectral satellite image analysis. Authors proposed a new correlated correspondence algorithm based on network analysis. The algorithm finds a matching between two distributions, which preserves the geometrical and topological information of the corresponding graphs. The performance of the algorithm evaluated on a shadow compensation problem in hyperspectral image analysis has shown to improve to the the land use classification obtained with the compensated data.
Liu, et al. [10] have proposed a neighboring filtering kernel to spatial-spectral kernel sparse representation for enhanced classification of hyperspectral images. The proposed neighboring kernel presented a framework of spatial-spectral kernel sparse representation classification (KSRC) and could measure the spatial similarity by means of neighborhood filtering in the kernel feature space. The authors have shown the presented method to be effective and was able to outperform the existing spatial-spectral kernels. In addition, the proposed spatial-spectral KSRC could open a wide field for future developments in which filtering methods could be easily incorporated.

3. Discussion

Hyperspectral image is a set of data which measure the spectrum of solar radiation reflected by the earth’s surface. The detailed spectral information provides new opportunities to identify objects that exhibit distinguishable spectral characteristics associated with absorption and reflectance features. These images have many applications in agriculture, geology, forestry, landscape, biodiversity conservation, regional planning as well as warfare. The information contained in hyperspectral image allows the characterization, identification, and classification of the land-covers with improved accuracy and robustness. With very narrow spectrum band, hyperspectral image data include ample information, which reflects tiny differences among materials of the earth’s surface. Hyperspectral image is of high input dimension of pixels, small number of labelled samples, and spatial variability of the spectral signature. However, high dimensionality of hyperspectral data, usually coupled with limited reference data available, limits the performances of supervised classification techniques. The commonly used pixel-wise classification lacks information about spatial structures of the image. In order to increase classification performances, integration of spatial information into the classification process is needed. A new IPSO technique and spectral-spatial for feature extraction based on GWLM scheme for hyperspectral image classification is proposed, and classification is done using ANN classifier to generate knowledge base.

4. Proposed methodology

The appropriate process flow of the proposed technique is as follows: An Improved Particle Swarm Optimization for based Artificial Neural Network is proposed for Hyper Spectral Image Classification. Firstly, Hyperspectral image considered is pre-processed which includes resizing, de-noising and color separation.
The resulting image is segmented using IPSO. Spectral-spatial features are extracted from the segmented image by using Gaussian Weighted Local mean operator. Laplacian Eigen values are generated for the extracted features. Then, feature vectors are created using the generated Laplacian Eigen values. Then classification is done using ANN classifier to generate knowledge base. This constitutes the training phase. In the testing stage the Hyperspectral image which is to be tested is classified using ANN classifier to obtain the statistical data for proper image analysis.
Let us consider the hyperspectral image with N number of pixels is represented as.

$$X = \{x_1, x_2, \cdots, x_N\}$$

The spectral vector associated with an image pixel

$$x_i = \{x_{i1}, xi_2, \cdots, x_{id}\}$$

with number of spectral bands and image pixel $i \in B$, where $B = \{1, 2, \cdots, n\}$ is the set of integers indexing the n pixels of x.
Here, first we do pre-processing such as resizing, de-noising, color separation. For the purposes of this paper, the tested images were resized into (1024x1024), and de-noising the HSI (Hyper Spectral Image) used by the median filtering. Image de-noising or filtering is a technique to reduce noises from a corrupted image. In many applications of image processing, the input image may be corrupted by noise and thus may not show the features or colors clearly. Thus, a major task in image processing is to extract information from the noisy images. The median filtered image $g(x, y)$ can be obtained from the median pixel values in a neighborhood of (x, y) in the input hyperactive spectral image $f(x, y)$ as defined by the following formula:

$$g(x, y) = median\left(\sum_{i=-1}^{1} \sum_{j=-1}^{1} f(x-i, y-j)\right) \quad (1)$$

Hyperactive Spectral Image segmentation using Improved Particle Swarm Optimization Algorithm: Particle Swarm Optimization is an algorithm developed by Kennedy and Eberhart, that simulates the social behaviors of bird flocking or fish schooling and the methods by which they find roosting places, foods sources or other suitable habitat. The proposed IPSO algorithm not only improves the standard PSO algorithm but also adds new strategy in order to find the global solution better than PSO algorithm by applying the chaotic sequences for weight parameter. The chaotic sequences are included in the inertia weight factor of the classical PSO to improve the searching capability of the algorithm. The proposed approach automatically determines the optimum number of clusters. The algorithm starts by partitioning the hyperspectral image into a relatively large number of clusters to reduce the effect of initial conditions.
In the basic IPSO technique, suppose that the search space is d-dimensional. Each pixel is called particle, and each particle (i-th particle) is represented by d-dimensional positional vector and described as.

$$X_i = [x_{i1}, x_{i2}, \cdots, x_{id}]$$

The fitness function is calculated using the following formula,

$$f(t) = \sum_{k=1}^{k} m(k) \quad (2)$$

Feature Extraction Using Gaussian Weight Local mean Operator (GWLM):

In this section, GWLM is proposed to extract the spectral and spatial information. In our proposed scheme, GWLM is applied for each pixel x_i of HSI to extract its spatial information x_i^s. Let f is usually a gray value of image, u is any pixel point, v denotes the translation interval, and $B_r(u)$ is a sliding window of size $r \times r$ (r is an odd number) pixels with the centre from the optimal solution of IPSO.

Set the Gaussian weight between the point u and $u+v$ after translation,

$$w(u,v) = \exp\left(-\| f(u+v) - f(u)\|^2 / 2\sigma^2\right) \quad (3)$$

$$W_r(u) = \sum_{v \in B_r(u)} w(u,v) \quad (4)$$

It is the normalization factor. GWLM of u in the neighborhood $B_r(u)$ is given by

$$GWLM(u) = \frac{1}{W_r(u)} \sum_{v \in B_r(u)} w(u,v) f(u+v) \quad (5)$$

The above procedure is followed for the hyperspectral data set of Lillies close up flower and below shown technique is applied for land cover hyperspectral data set of Hyderabad region.

The process of image acquisition is followed by segmentation where the pixels making up the image are assigned suitable labels either in a supervised or unsupervised manner. Selection of the supervised or unsupervised mode mainly depends on the underlying application. The segmentation techniques used in the proposed system are Fuzzy C-means method and Otsu's thresholding method. Otsu's method aims at optimizing the parameters without any prior estimation by following three main discriminating criteria which are Intra-class variance, interclass variance and total variance. The histogram normalization of the image is obtained by the probability of the pixels

$$PI = \frac{h(i)}{\sum_{j=0}^{L-1} h(j)} \text{(3)with } Pi \geq 0$$

$$\sum_{i=0}^{L-1} Pi = 1 \quad (6)$$

Whereh(i)represents occurrence of gray level pixel and L represents the total number of gray levels. In case of binarization, the ratio of interclass variance to total variance is maximized. The Otsu method groups the pixels into two classes, following which the objects are segmented there by performing maximum segmentation.

Random Forest Classifier

Step 1: An initial variable is taken into consideration and split, at every possible point resulting in 2 child nodes at each level, where the left nodes correspond to a positive response and right nodes respond to negative response.

Step 2: The split points are evaluated based goodness-of-split criteria which can be defined as

$$\Delta I(s,t) = i(t) - pL[I(tL)] - pR[i(tR)] \quad (7)$$

here s represents the exact split, pL is defined as the observation proportion at a node t entering the left child node tL,and pR is the proportion observed that is entering the right node tR and is similar topL. Impurity is given by i(tL) and i(tR) of left node and right node.

Step3: The best split designated for each variable is considered and selected based on its maximum impurity reduction capacity. This is repeated in iterations at the root node for rest of the variables.

Step 4: Ranks are allotted for each of the variables based on the impurity reduction, further selecting the variables and the split points such that the root node or parent node is capable of reducing the impurity.

Step 5: Classes are assigned to nodes based on the minimization of mis-classification cost such that cost of mis-classification is as minimal as possible.

5. Results

The following simulation results obtained forhyperspectral flower dataset shown below.

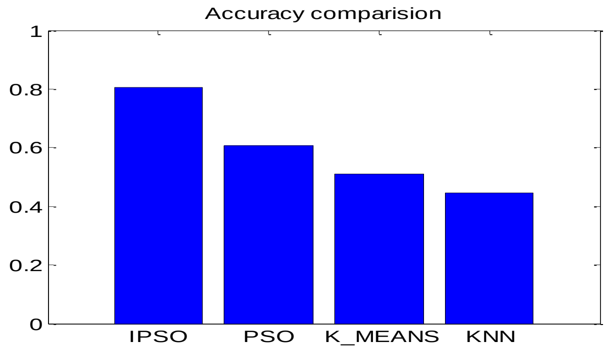

Fig. 1: Overall Accuracy.

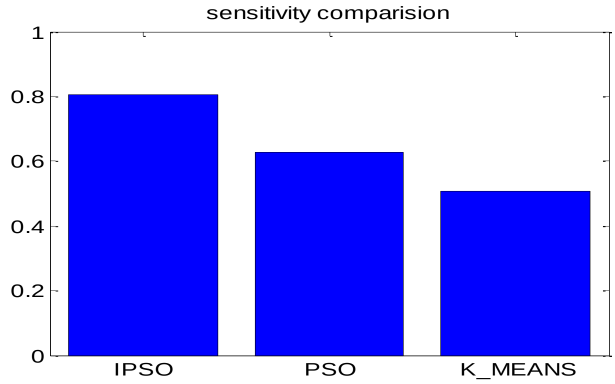

Fig. 2: Sensitivity Comparison.

Classifier Accuracy = 3×4 table

Table 1: Accuracy for Land Cover Dataset

Parameters	NN	SVM	RF	Navie Bayes
LBP	33.333	35	92.5	39.167
Area	33.333	35	91.667	39.167
LBP_Area	33.333	35	92.5	39.167

6. Conclusion

In this paper for the flower hyperspectral dataset new IPSO technique is proposed and it provides better classification accuracy as shown in figure 1 and also sensitivity graph is obtained which can be seen in figure 2. From figure 1 it is clear that the proposed technique provides80% accuracy when compared to others . In the case of Land cover hyperspectral dataset proposed improved Random Forest algorithm provides 92.5% classification accuracy when compared to other classifiers such as support vector machine, K Nearest neighbor and Navie Bayes as shown in Table 1 . In both the considered different hyperspectral dataset newly proposed techniques achieves better classification accuracy.

References

[1] Vitousek, P.M., 1994. Beyond global warming: ecology and global change. Ecology 75, 1861-1876.https://doi.org/10.2307/1941591.

[2] Chandan, M. C, Vinay, S., Bharath H. Aithal, Ramachandra, T.V. 28-30th December 2016. Land use assessment and urban growth monitoring in Hyderabad region, India.

[3] Ramachandra, T. V., Aithal, B. H., Vinay, S., 2013. Land use land Cover dynamics in a rapidly urbanising landscape SCIT Journal, vol: 13: Pages 1-1.

[4] Sharifah Lailee Syed Abdullah, HamirulAiniHambali, Nursuriati Jamil, "Segmentation of Natural Images Using an Improved Thresholding-based Technique", International Symposium on Robotics and Intelligent Sensors 2012, Science Direct.

[5] S. Jenicka, A. Suruliandi, “A Textural Approach for Land Cover Classification of Remotely Sensed Image", International Journal of Computer Applications (0975 – 8887) Volume 92 – No.10, April 2014.

[6] Chandan, M. C, Vinay, S, Bharath H. Aithal, Ramachandra T.V, "Land use assessment and Urban Growth Monitoring in Hyderabad region, India", Conference on Conservation and Sustainable Management of Ecologically Sensitive Regions in Western Ghats, THE 10TH BIENNIAL LAKE CONFERENCE.

[7] V.F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, J.P. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification", ISPRS Journal of Photogrammetry and Remote Sensing, 2012, Science Direct.

[8] Cavallaro, G.; Dalla Mura, M.; Benediktsson, J.A.; Bruzzone, L., "A comparison of self-dual attribute profiles based on different filter rules for classification", in Geoscience and Remote Sensing Symposium (IGARSS), 2014. IEEE International, pp.1265-1268, 2014.

[9] Rebetez, J.; Tuia, D.; Courty, N., "Network-Based Correlated Correspondence for Unsupervised Domain Adaptation of Hyperspectral Satellite Images", in Pattern Recognition (ICPR), 2014.22nd International Conference, pp.3921-3926, 2014.

[10] Jianjun Liu; Zebin Wu; Zhihui Wei; Liang Xiao; Le Sun, "Spatial-Spectral Kernel Sparse Representation for Hyperspectral Image Classification", in Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal, vol.6, no.6, pp.2462-2471, 2013.