Digital Assessment of Spatial Distribution of the Surface Soil Types Using Spatial (Texture) Features with MLC and SVM Approaches

Amol D. Vibhute, Karbhari V. Kale, Rajesh K. Dhumal, Ajay D. Nagne, Suresh C. Mehrotra, Amarsinh B. Varpe, Rupali R. Surase, Dhananjay B. Nalawade and Sandeep V. Gaikwad

Abstract In the present work, the effort has been made to identify and distribute of surface soil types using high spatial resolution multispectral (HSRM) image investigated in two ways. First, multispectral data is classified based on conventional approaches. Second, a method based on gray level co occurrence matrix (GLCM) as spatial objects extraction of the multispectral data is proposed. In this view, various texture parameters of the co-occurrence matrix method were used to highlight and extract the textures in the image. The method was computed on increasing matrix window size starting from original one. The Resourcesat-II Linear Imaging Self Scanning (LISS-IV) sensor multispectral image was used for testing the algorithms of the study area Phulambri Tehsil of Aurangabad region of Maharashtra state, India. The proposed approach was used as an input for Maximum Likelihood Classifier (MLC) and Support Vector Machine (SVM) approaches for identification and distribution of surface soil types and other patterns. The experimental outcomes of the present research were appraised on the basis of classification accuracy of methods. The overall accuracy of classification by MLC and SVM after spatial feature extraction was 92.82 and 97.32% with kappa value of 0.90 and 0.96 respectively. It was found that, the accuracy of the classification has increased after considering spatial features based on co-occurrence matrix. The results were promising to extract the mixed features for classification of soil type objects.

A. D. Vibhute (✉) · K. V. Kale · R. K. Dhumal · A. D. Nagne · S. C. Mehrotra · A. B. Varpe
R. R. Surase · D. B. Nalawade · S. V. Gaikwad
Geospatial Technology Research Laboratory, Department of Computer Science and IT,
Dr. Babasaheb Ambedkar Marathwada University, Aurangabad 431004, Maharashtra, India
e-mail: amolvibhute2011@gmail.com

C. R. Krishna et al. (eds.), *Proceedings of 2nd International Conference on Communication, Computing and Networking*, Lecture Notes in Networks and Systems 46, https://doi.org/10.1007/978-981-13-1217-5_74

Keywords Soil type classification · Multispectral data · Gray level co-occurrence matrix · Spatial features · Soil mapping · Maximum likelihood algorithm · Support vector machine

1 Introduction

Soils are very heterogeneous in nature which is one of the most imperative resources. Soils are assorted product of rock type, landform or topography, vegetation cover with climate. As a result, spectral features of soil or single landscape model cannot suffice to estimate soil features or soil boundary [1, 2]. Detailed catchment scale or farm scale maps of soil variability that offer high spatial resolution are requisite to estimate soil threats [3]. Moreover, the soil and land resource survey is vital in precision farming for better management and planning soil for better food production. However, traditional methods for analysis of soil and its mapping are time consuming, expensive and does not fulfill the spatiotemporal variability [4]. Furthermore, the topographic and cadastral maps as a reference were used by the conventional survey methods and the traditional methods are also time-consuming, formidable and subjective [5]. Consequently, soil features can be extracted from remote sensing datasets with its analysis and mapping. Nevertheless, the digital assessment of surface soil types, spatial distribution and its mapping is somewhat formidable task due to various soil attributes and assorted effect of numerous features of planet surface that can vary spectral and spatial features of soils and compose it non-consistent through the spectrum region [6].

The traditional methods of HSRM image classification consider solely spectral information while neglecting the spatial information. Under this constraint, we have tried to extract the spectral and spatial features of mixed land use and covered area for soil classification like settlements, hilly area, farmland, natural vegetations, etc., with HSRM images.

2 Materials and Methods

2.1 The Test Location, Used Satellite Data and Soil Sampling

The test location is geographically located at 19° 28′ 43.27″–20° 24′ 52.19″N latitude and 75° 13′ 10.75″–75° 30′ 14.87″E longitude which covers 72.70 km^2. The data was used for this study have been acquired from various sources like satellite data along with field data. Global Positioning System (GPS) was used to acquire the ground coordinate points. The base maps were developed using Survey of India (SOI) toposheet of 1:50,000 scales. The proposed approach was applied on Indian Remote Sensing (IRS) Resourcesat-II—P6 satellite imagery of LISS-IV sensor. The spatial resolution of LISS-IV image is 5.8 m with three spectral channels and 23.5 km swath. The imagery was geometrically and radiometrically corrected (orthorectified) by the provider [7]. Total 74 soil samples (0–20 cm depth) were collected and their GPS values were recorded with scenes. The soil sampling sites were selected based on spatial distribution. The field work for collecting soil samples and ground truth points were carried out during the period of February 10 to March 25, 2015 and in between 0800 and 1330 h on light days with clear environment. The soil samples were collected in air-tight container. Every specimen was air dried and passed through 2 mm sieve for laboratory analysis of some physicochemical soil properties. The soil properties were analyzed by standard laboratory methods at “MIT Soil and Water Testing Laboratory”, Aurangabad, Maharashtra, India.

2.2 Experimental Methodology

In this research, image processing operations were performed through Environment for Visualizing Images (EVVI 5.1) image analysis software and ArcGIS 10 software. The methods are discussed as follows.

2.2.1 Spectral and Spatial Feature Information

A reliable approach to deal with pixel-based classification of satellite imagery is to consider its spectral information. The spectral features (information) of image is directly used as the input for the classification algorithms and processed as a feature vector. It is reliable for classification based on spectral reflectance of each object. Unfortunately, it does not consider spatial structures of various objects in the classification [8]. The texture-based spatial features were analyzed for the HSRM satellite imagery. The GLCM method is broadly used for texture analysis and pattern recognition also known as spatial co-occurrence matrix (SCM) which is more scientific depicter of texture and more accurate as compared to other methods. From the given spectral band or its subset, the GLCM method is implemented from the spatial stochastic properties [9, 10]. The co-occurrence matrix is two-dimensional matrix of joint probabilities between pairs of pixel values one with gray level value i and other with gray level value j, separated by a distance $d = (1{,}0)$ (i.e. neighboring pixels in the same row) at an angle of 0° (i.e., horizontally) in a given direction from left to right. The GLCM matrix is a symmetrical matrix which element $p\,(i, j \mid d, \theta)$ contains the second-order statistical probability values [9, 10]. The elements of the GLCM method were computed using Eq. (1).

$$p(i, j) = \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} i \cdot p(i, j), \tag{1}$$

Table 1 GLCM textures and their equations

GLCM texture features	Equations	
Mean	$x = \sum_{i,j} i \cdot p(i, j)$	(2)
Variance	$f_4 = \sum_i \sum_j (i - u)^2 p(i, j)$	(3)
Contrast	$f_2 = \sum_{n=0}^{Ng-1} n^2 \left\{ \sum_{\substack{i=1 \\ \lvert i-j \rvert = n}}^{Ng} \sum_{j=1}^{Ng} p(i, j) \right\}$	(4)
Dissimilarity	$f = \sum_{n=1}^{Ng-1} n \left\{ \sum_{\substack{i=1 \\ \lvert i-j \rvert = n}}^{Ng} \sum_{j=1}^{Ng} p(i, j)^2 \right\}$	(5)

where, $p(i, j)$ is the GLCM, and Ng is the digit of gray levels.

The pairs of pixels of GLCM are computed in four angular directions such as 0° in horizontal, 45° in right diagonal, 90° in vertical and 135° in left diagonal within the instantly neighboring pixels of the image. Hence, to produce the texture image, the GLCM or the texture measures were usually calculated within a moving window. Haralick [9] has extracted 14 textural features from the GLCM which are the significant features. In this study, we have computed four texture measures based on co-occurrence matrix such as; mean, variance, contrast, and dissimilarity. These measures were computed from the co-occurrence matrix using Eqs. 2, 3, 4, and 5 drawn in Table 1.

2.2.2 Proposed Approach

The above texture measures with increasing window sizes of 3×3, 5×5, and 7×7 were considered. The x and y values were used to calculate the co-occurrence matrix. In the calculation of co-occurrence matrix by considering each pixel, 8-neighboring pixels were preferred. The four directions were used, i.e., 0°, 45°, 90°, and 135° separately for the calculation. The grayscale quantization level was used with 64-bit. The 64-bit value was beneficial when the grayscale values of the image are spread over a broad range. After the setting of grayscale values, co-occurrence matrix was computed in four directions. Mean, variance, contrast and dissimilarity features were extracted for further processing. In our experiments, four features were calculated; hence the feature set was four for each of the band. The used data has three bands; thus the final size of feature set was 12. The resulted textures of GLCM method is shown in Fig. 1 for window sizes of (a) 3×3, (b) 5×5, and (c) 7×7 respectively.

For each window size, these all twelve features were used and MLC as well as SVM methods were implemented over it. The window sizes were increased up to

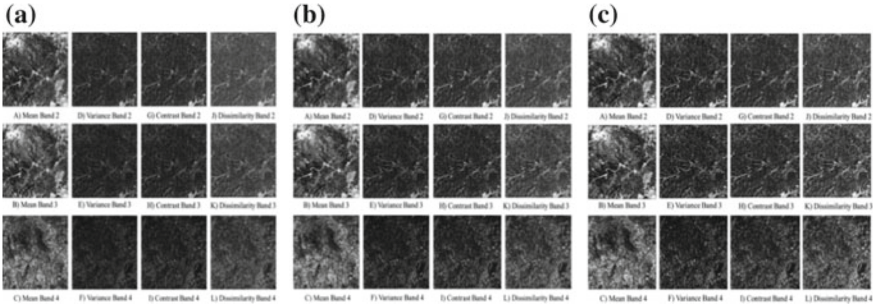

Fig. 1 Spatial features based on GLCM texture measures with window sizes of **a** 3×3, **b** 5×5 and **c** 7×7 respectively

7×7 due to the stability (similarity) of the textural features of window sizes 5×5 and 7×7. The GLCM feature set with increased window size was given as input for MLC and SVM algorithms. The LISS-IV image for MLC and SVM methods were trained accordingly field observations in the combinations with laboratory results, geolocated ground reference data with visual inspection. The ROIs were developed by said reference points which were the support vectors for SVM-based classification and also used for the MLC-based classification to train and test the HSRM image for classification. The available training and testing pixels were 920 and 2203 for said data.

The GLCM feature set with increased window size was given as input for MLC algorithm first. The data scale factor for MLC was one and probability threshold value was 0.1 for all training pixels. The MLC is parametric method based on the Bayes theorem [4, 10–12] is a widely acceptable by the researchers including remote sensing community. The MLC algorithm was implemented by the Eq. 6 [12].

$$p(X/Cj) = \frac{1}{(2\pi)^{n/2}\left|\sum j\right|^{0.5}} \times \exp\left[-\frac{1}{2}(\mathrm{DN} - \mu j)^T \sum_j^{-1} (\mathrm{DN} - \mu j)\right] \tag{6}$$

Additionally, SVM method was also considered for LISS-IV data for comparison of MLC results and for getting better results. The kernel of Gaussian Radial Basis Function (RBF) of SVM method was implemented using Eq. 7 for classification analysis. SVM non-parametric supervised machine learning algorithm [13] was chosen due to its high accuracy for classification of heterogeneous and noisy remote sensing data with less training pixels. The SVM method is originally formulated by Vapnik in 1995 based on statistical learning theory [10, 11, 13, 14].

Gaussian (RBF)

$$k(xi, xj) = \exp(-\gamma \cdot ||xi - xj||^2), \quad \gamma > 0, \tag{7}$$

where, $k(xi, xj)$, xi, xj and γ is respectively the kernel function, training vectors and kernel constant parameter.

The SVM method was implemented with gamma (γ) value 0.010, penalty parameter 100, and classification pyramid level one with reclassification threshold 0.90 and classification probability threshold 0.10.

3 Results and Discussion

According to the laboratory results of soil physicochemical properties and field data with visual inspection one major (black cotton soil or “Regur”) and two minor (Lateritic soil and Sand dunes) soil types were identified and classified with other land features. The black cotton soil includes “vertisol”, “inceptisol” and “entisol”; lateritic soil includes “alfisol” and sand dunes include “Typic Torripsamments”. Five soil classes according to USDA soil taxonomy [15] were detected and classified on the basis of report generated by laboratory analysis of soil physicochemical properties and field investigations. The soil classes were “vertisol”; “inceptisol” and “entisol” of black cotton soil, “alfisol” of lateritic soil, “Typic Torripsamments” of sand dunes, and other land features such as vegetations, water bodies, settlements and boundaries with roads were also classified. The classification maps (Fig. 2 (a) for MLC method and (b) for SVM method) clearly indicate that, black cotton soils [16] have covered most area of the test site followed by vegetations, sand dunes, lateritic soil, settlements, roads and water bodies. As per the laboratory reports of soils, these black soils are deep or heavy and medium or lighter as per its physical properties. The textures of black soils are loamy to clayey with mixed carbonates (mostly $CaCo_3$) and are suitable for cotton cultivation. The organic carbon, organic matter and nitrogen found to be less in this soil and pH values are near about 7–9. The EC values vary from 0.25 and 0.46 d Sm-1 where values were less than 0.36 d Sm-1. The iron contents are good in black soils. The lateritic soil included only the “alfisol” in the studied areas as per USDA soil taxonomy which found to be hilly part and somewhat farming sectors of test site. The pH value of these soils is low and organic matter is high with fine texture. Sand dunes were observed to be more at riverside and hilly rocks due to the spectral structure of sand dunes and rocks. The texture of sand dunes was sandy. Electrical Conductivity (EC) values and organic matter contents are very low in sand dunes. Natural vegetations including agricultural crops were accurately classified and mapped.

First, the classification methods were computed on original preprocessed LISS-IV data (only spectral information) and outcome was evaluated with accuracy. The classified maps derived by MLC and SVM methods with spatial features are illustrated in Fig. 2a, b respectively. The soil classes were classified well with both the classifiers. As black cotton soil was found to be more followed by sand dunes and alfisol soils.

It was observed that, the accuracy was less with both methods on original image (Table 2). Accordingly, the spatial texture measures were pondered and accuracy

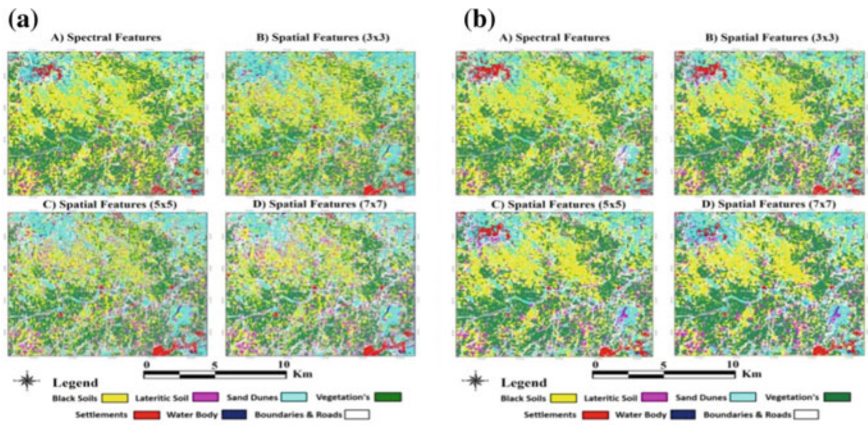

Fig. 2 Result of classification using (a) MLC and (b) SVM approaches respectively on **A** original data, **B** spatial information (3×3 window size), **C** spatial information (5×5 window size) and **D** spatial information (7×7 window size)

Table 2 Accuracy assessment of MLC and SVM algorithms

Methods/ features	Spectral features		Spatial features (3×3)		Spatial features (5×5)		Spatial features (7×7)	
	OA	K	OA	K	OA	K	OA	K
MLC	83.38	0.79	89.83	0.87	92.23	0.90	92.82	0.90
SVM	84.52	0.80	91.51	0.89	94.55	0.93	97.32	0.96

Where, OA-Overall Accuracy and K-Kappa Value

was evaluated of each window sizes and found to be similarity in textures of 5×5 and 7×7 and hence window sizes were finalized up to 7×7. Producer's accuracy, user's accuracy, overall accuracy and kappa statistics were calculated for evaluation of the accuracy [10, 12]. The ground truth points were used for deriving the confusion matrix. The classified values against actual ground observation values at particular location were determined in the confusion matrix. The diagonal values of the confusion matrix depicts the correctly classified features, where as nondiagonal nonzero value demonstrates the misclassification between classified features from related view [17]. The resulted outcome of both MLC and SVM algorithms are drawn in Table 2 with overall accuracy and kappa values.

It was observed that, the spectral confusion were remarkable in between sand dunes and settlements with both the classification methods. The four texture measures were highlighted the individual objects, when increased the window size, because they are related to the size of the object. Larger the window size within the window, the object reflects higher gray values. Consequently, the objects with various sizes were separated each other. The classification methods were implemented on these four texture measures with increasing window sizes and achieved better accuracy

(Table 2) than the original one. The confusion was reduced with spatial information and achieved good results.

The accuracy has increased with increased window sizes. The soils and vegetations were classified more accurately than other classes. The window size 7×7 was given best results than other two window sizes such as 3×3 and 5×5 for four texture measures. In fact, all soil classes were classified accurately except sand dunes and settlements caused by similarity in spatial (texture) and spectral (signature) information.

The class-specific accuracies, black cotton soil, lateritic soil, and vegetations were classified with the higher accuracy excluding sand dunes which were misclassified with settlements. Our aim was to identify and classify soil classes accurately within the farming sectors which were classified well; however, we have also considered soils on hilly area, settlement area. It causes the low accuracy of settlement class within the test site. It can be concluded that, heterogeneous lad features can be correctly classified by using spatial information.

4 Conclusions

The present research highlighted the advantages of GLCM-based spatial features for the classification of five soil classes and other mixed land patterns. The research work introduces the new approach for soil taxonomy using spatial information along with spectral information. Only spectral information is not sufficient for classification of heterogeneous land features. Consequently, we have used HSRM satellite image of LISS-IV with 5.8 m of pixel size and three spectral bands. The MLC and SVM classification approaches were computed for spatial distribution of soil accordingly USDA soil taxonomy and four textures of GLCM were considered with three window sizes of 3×3, 5×5, and 7×7 matrix. It was observed that, sand dunes and settlement area, were not classified accurately when applied the two approaches on spectral as well as spatial information. Nevertheless, when included spatial features in the classification as an input with increased window sizes, the accuracy was also enhanced up to 10–13% more for MLC and SVM approaches respectively. According to the classification maps and accuracies, it is noticed that, spatial features are very imperative for mixed land features especially soils in different platforms. In the final conclusion, SVM method has given much better results than MLC method. The classification results can be useful for better utilization of soil for precision farming practices along with its planning and management.

Acknowledgements The authors would like to acknowledge the UGC-BSR fellowship; UGC-SAP(II)DRS Phase-I and Phase-II for providing infrastructure, DeitY, Government of India, under Visvesvaraya Ph.D. Scheme, DST and also extend our gratitude to DST-FIST program to Department of CSIT, Dr. BAM University, Aurangabad, M.S. India. We would also thankful to Prof. D. T. Bornare and his team for physiochemical analysis of soil specimens at “MIT Soil and Water Testing Laboratory, Aurangabad”, Maharashtra, India.

References

1. K.S. Lee, G.B. Lee, E.J. Tyler, Determination of soil characteristics from thematic mapper data of a cropped organic-inorganic soil landscape. Soil Sci. Soc. Am. J. **52**(4), 1100–1104 (1988)
2. M.L. Manchanda, M. Kudrat, A.K. Tiwari, Soil survey and mapping using remote sensing. Trop. Ecol. **43**(1), 61–74 (2002)
3. F. Garfagnoli, A. Ciampalini, S. Moretti, L. Chiarantini, S. Vettori, Quantitative mapping of clay minerals using airborne imaging spectroscopy: New data on Mugello (Italy) from SIM-GA prototypal sensor. Eur. J. Remote Sens. **46**(1), 1–17 (2013)
4. A.D. Vibhute, B.W. Gawali, Analysis and modeling of agricultural land use using remote sensing and geographic information system: A review. Int. J. Eng. Res. Appl. (IJERA) **3**(3), 081–091 (2013)
5. S.S. Asadi, M.S. Kumar, M.V. Raju, B.V.T.V. Rao, Preparation of soil map using remote sensing and GIS: A case study from Andhra Pradesh. Int. J. Emerg. Trends Eng. Dev. **4**(2), 138–146 (2014)
6. E. Ben-Dor, K. Patkin, A. Banin, A. Karnieli, Mapping of several soil properties using DAIS-7915 hyperspectral scanner data-a case study over clayey soils in Israel. Int. J. Remote Sens. **23**(6), 1043–1062 (2002)
7. http://www.nrsa.gov.in/
8. M. Fauvel, J.A. Benediktsson, J. Chanussot, J.R. Sveinsson, Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles. IEEE Trans. Geosci. Remote Sens. **46**(11), 3804–3814 (2008)
9. R.M. Haralick, K. Shanmugam, I. Dinstein, Textural features for image classification. IEEE Trans. Syst. Man Cybernet. **3**(6), 610–621 (1973)
10. J. Gao, *Digital Analysis of Remotely Sensed Imagery* (McGraw-Hill Professional, New York, 2008)
11. J.A. Richards, X. Jia, *Remote Sensing Digital Image Analysis An Introduction*, 4th edn. (Springer, Berlin, 2006)
12. A.D. Vibhute, R.K. Dhumal, A.D. Nagne, Y.D. Rajendra, K.V. Kale, S.C. Mehrotra, in *Analysis, Classification, and Estimation of Pattern for Land of Aurangabad Region Using High-Resolution Satellite Image*. Proceedings of the Second International Conference on Computer and Communication Technologies (Springer, New Delhi, 2016), pp. 413–427
13. V.N. Vapnik, An overview of statistical learning theory. IEEE Trans. Neural Netw. **10**(5), 988–999 (1999)
14. A.D. Vibhute, K.V. Kale, R.K. Dhumal, S.C. Mehrotra, in *Soil Type Classification and Mapping using Hyperspectral Remote Sensing Data*. IEEE, International Conference on Man and Machine Interfacing (MAMI) (2015), pp. 1–4
15. USDA, *Keys to Soil Taxonomy, United States Department of Agriculture* (Natural Resources Conservation Service, 2014)
16. T. Bhattacharyya, D.K. Pal, C. Mandal, P. Chandran, S.K. Ray, D. Sarkar, A.K. Sahoo et al., Soils of India: Historical perspective, classification and recent advances. Curr. Sci. 1308–1323 (2013)
17. A.D. Vibhute, A.D. Nagne, B.W. Gawali, S.C. Mehrotra, Comparative analysis of different supervised classification techniques for spatial land use/land cover pattern mapping using RS and GIS. Int. J. Sci. Eng. Res. **4**(7), 1938–1946 (2013)