Integration of Heterogeneous Features from Co-registered Hyperspectral and LiDAR Data for Land Cover Classification

Farah Jahan, Jun Zhou, Mohammad Awrangjeb, Yongsheng Gao

Institute for Integrated and Intelligent Systems, Griffith University,
Brisbane, Australia.

farah.jahan@griffithuni.edu.au , {jun.zhou, m.awrangjeb,
yongsheng.gao} @griffith.edu.au

ABSTRACT

Exploiting the multi-source data is an effective but challenging problem for land cover classification. Popular remote sensor data, e.g., hyperspectral (HS) and light detection and ranging (LiDAR), contain complementary information for land cover if they are co-registered. In this paper, we aim to integrate information extracted from these data sources for land cover classification. At first, we propose a novel feature extraction method by calculating the inverse coefficient of variation (ICV) using the Gaussian probability of neighbourhood between every pair of bands in HS data. This is calculated for each band with respect to every other band to form an ICV cube. We reduce the number of planes in the cube by applying PCA on it and spatial features are then extracted for significant principal components. The spectral information from HS data, their ICV responses, and spatial information from ICV responses have complementary information; that is why we fuse them together by layer stacking to generate discriminant features. Secondly, we also derive height and spatial features from LiDAR Digital Elevation Model (DSM), which are later concatenated with the HS derived features. Finally, these features are classified using linear discriminant analysis (LDA) classifier. The classification results prove the effectiveness of the derived features from both data sources.

Keywords: Hyperspectral, LiDAR, ICV, PCA, DSM, LDA.

1. INTRODUCTION

Recently, various sensors are being used to obtain complementary information from the surface of the earth. These sensors can be categorised as passive (e.g., multispectral and hyperspectral) and active (e.g. LiDAR and synthetic aperture radar (SAR)). Hyperspectral data contain detailed spectral signature of a material in a scene. On the other hand, LiDAR data contain the vertical structure of objects

in the scene. Integration of useful information from multiple data sources in an effective way for automatic interpretation still remains a challenge. These two types of data have been intensively investigated for tree species classification, forest biomass estimation and land-cover classification.

(Man *et al.* 2015) fused HS and LiDAR data at pixel-label for urban land use classification. They used LiDAR intensity and height information from LiDAR point cloud and spectral and texture information from HS data. (Luo *et al.* 2015) derived features from LiDAR point cloud and fused them with HS data. Several spatial resolutions were used to decide best resolutions for classification in (Luo *et al.* 2015). Spatial features provide important texture properties of a material which cannot be obtained from spectral and height information. So, such features can help to differentiate objects in which spectral and height features are similar. Recently, deriving spatial features on principal components (PCs) of HS and LiDAR data has become a very popular (Khodadadzadeh *et al.* 2015; Ghamisi *et al.* 2017; Dalla Mura *et al.* 2010; Rasti *et al.* 2017; Jahan *et al.* 2017) technique. (Ghamisi *et al.* 2017) used extinction profile for the extraction of spatial features from HS and LiDAR data. Also, spectral features from HS bands are used along with the spatial features. In (Luo *et al.* 2016) feature fusion were done by layer stacking and principal component analysis. Besides, layer stacking of the features derived from both data sources, other effective fusion approaches are also used. For instance, (Ghamisi *et al.* 2017) used a graph-based feature fusion and (Rasti *et al.* 2017) used a sparse and low rank feature fusion approach for fusing HS and LiDAR data for pixel-based classification. (Man *et al.* 2015) proposed an object-based classification by setting several object related parameters for better classification accuracies. In (Ghamisi *et al.* 2017), the authors used the deep convolutional neural network for classification.

From the above discussion on the existing literature, we can observe that effective derivation of features from both data sources should be the first step for high classification accuracy. The derived features must contain discriminative properties so that samples from different classes can be identified accurately. Also, an effective combination of features from different sources is important for generating more discriminative feature than individual sources. Fusion is a combination process that reduces dimensionality as well as increases the discriminative power of features without losing important information. In this work, we focus on effective information derivation from both HS and LiDAR data. Then we fuse them together for enhancing discrimination capability. The fusion of HS and LiDAR-derived features prove their effectiveness which is also observed in the classification accuracies.

The main contributions of this paper are as follows:

- We propose a novel feature extraction method by calculating the similarity between one band to the rest in terms of conditional probability. We calculate the inverse coefficient of variation (ICV) for converting the detailed similarity measures into a set of new spectral

responses. Using these new spectral responses we construct an HS cube from the original one. We name the new cube as an ICV cube.

- The probabilistic cube provides important and discriminative spatial information. This cube also contains complementary information regarding the spectral cube.
- We successfully derive spatial features from the ICV cube and LiDAR-derived DSM.

Fig. 1 The architecture of the proposed method.

2. METHODOLOGY

In this section, we describe the details of the proposed feature extraction and fusion approach. As shown in Fig. 1, the final feature vector for classification accumulates different features. The first one is the spectral responses from the HS cube, the second one is from the ICV cube, the third one is spatial features from the ICV cube and the forth part is spatial features from DSM and finally the DSM value itself, which is derived from LiDAR point cloud. The final feature vector is used for supervised classification of the sample pixels. In the following subsections, we discuss the details of these feature extraction techniques.

2.1 Generation of Inverse Coefficient of Variation (ICV) cube

For measuring the relationship between two bands, we extend the stochastic neighbourhood embedding (SNE) (Hinton *et al.* 2003) concept. SNE represents similarity between two points in terms of conditional probability. In our approach, we extend this concept for calculating neighbourhood attributes of the bands in the HS cube. Suppose, we have an HS cube H with dimensions $m \times n \times b$ where m , n and b , represent width, height and number of bands in the cube, respectively. So, if we consider a pixel at location (x, y) , we have a spectral feature vector $\langle q_{x,y}^1, q_{x,y}^2, \dots, q_{x,y}^b \rangle$, where $q_{x,y}^i$ represents the spectral intensity at location (x, y) in i -th band. The conditional probability $p_{i|j}$ for bands i and j at

pixel location (x, y) is proportional to the similarity between the bands and is calculated as

$$p_{i|j} = \frac{\exp\left(\frac{-\|q_{x,y}^i - q_{x,y}^j\|^2}{2\phi_i^2}\right)}{\sum_{k \neq i} \exp\left(\frac{-\|q_{x,y}^i - q_{x,y}^k\|^2}{2\phi_i^2}\right)}, \quad (1)$$

where ϕ^i is the bandwidth of the Gaussian kernel and is centered at $q_{x,y}^i$. We derive the relationship mentioned in Eq. (1) for each band in this way. Since, the cube has b bands, we get a similarity matrix, say S , of $b \times b$ dimensions, where $S^{ij} = p_{i|j}$ and i -th row of the matrix denotes the similarity between i -th band to the rest of the bands. For a pixel having b spectral values, we get b^2 values in the similarity matrix, where the curse of dimensionality increases exponentially. At this stage, we reduce the dimension of the similarity matrix by employing ICV, which is the ratio of mean to standard deviation (Sharma *et al.* 1994) as given in Eq. (4). For a band i , we calculate the mean of the i -th row and divide it with the standard deviation of that row. The operations can be described as

$$\mu^i = \frac{\sum_{j=1}^b S^{i,j}}{b}, \quad (2)$$

$$\sigma^i = \sqrt{\frac{\sum_{j=1}^b (S^{i,j} - \mu^i)^2}{b-1}}, \quad (3)$$

$$v^i = \frac{\mu^i}{\sigma^i}, \quad (4)$$

where μ , σ , v represent mean, standard deviation and inverse coefficient of variation respectively. For a pixel in the HS cube, Eq. (4) returns a vector of length b from similarity matrix of b^2 elements. We perform the same operations for every pixel in HS data; thus an ICV cube is produced of the same size as the

original HS cube. Fig. 2 shows the steps of this derivation process applied on HS data.

Fig. 2 ICV values derived from HS.

2.2 Spatial feature Extraction

In this section we aim to extract spatial features from the planes of the ICV cube. For better classification of hyperspectral and LiDAR data, it is important to analyse geometric properties of object(s) of the scene [10]. An effective technique for extracting spatial features from an image is mathematical morphological operation. Since, the ICV cube contains the same number of planes as the original HS cube which is still a large number, it is infeasible to apply DAPs on each plane of the cube because it produces $b \times (2h + 1)$ dimensional feature vector for every pixel which causes a curse of dimensionality as per Hughes phenomenon (Dalla Mura *et al.* 2011). In such a situation, it is more practical to reduce the number of planes from the bunch of planes without losing significant information. Additionally, to observe the linear dependency between two features, e.g., ICV values from two planes, we find their Pearson correlation coefficient (Fisher *et al.* 1992). This is performed for every individual feature with rest of the feature set. Figs. 3(a) and (b) pictorially represent such coefficient in matrix form for the training samples of HS and ICV respectively. It is clearly seen that the spectral responses of HS are more correlated than ICV. So, it is necessary to reduce such dependency between a pair of features. From the properties of PCA, we know that if features in samples are highly correlated they cause the PCA to overemphasise their contribution. Thus, applying PCA on less correlated features is more useful for generating discriminative feature vectors (Jolliffe *et al.* 2016). That is why, in order to increase the variance as well as reduce correlation among the features of ICV we apply PCA on ICV feature vectors. Aiming to reduce the number of planes we select largest principal components (PCs) with a cumulative variance of at least 99%. Then we apply DAP on the selected PCs as described in literature [3]. Like morphological profiles (MPs) [5], attribute profiles (APs) perform multiscale analysis of the image by using structuring element (SE) (Dalla Mura *et al.* 2010). APs incorporate different types of attributes to produce different characterisations of the scale of the structures of a scene captured in an image. Since, AP uses a sequence of opening and closing with SE of increasing size, it is able to extract properties invariant with scale. Given h attributes (e.g., area) for an image plane, AP produces a stack of $2h + 1$ planes, where h planes come from the closing profile, the original plane itself a plane and h planes from opening profile.

The differential attribute profile (DAP) (Dalla Mura *et al.* 2010; Dalla Mura *et al.* 2011) stores the residuals of the subsequent increasing transformations applied to the image. Since, important components of the profiles are more evident in DAP, it is more practical to use for obtaining important information.

For LiDAR DSM, which contains only one plane, we apply DAP directly on the plane and get spatial information from it.

2.3 Classification

Linear discriminant analysis (Fisher 1936) for multi-class classification incorporates the following two stages (Johnson et al. 1992; Li *et al.* 2006):

- In this technique, it is expected to decrease intra-class variation and increase inter-class variations for better classification. Say, we have total of C classes for training. The intra-class and inter-class scatter matrices then can be expressed by Eqs. 5 and 6 respectively:

$$\sum_w^{\wedge} = \sum_{c=1}^C \sum_{f \in f_c} (f - \bar{f}_c)(f - \bar{f}_c)^T \quad (5)$$

and

$$\sum_b^{\wedge} = \sum_{c=1}^C m_c (\bar{f}_c - \bar{f})(\bar{f}_c - \bar{f})^T, \quad (6)$$

where f denotes a sample or feature vector, m_c is the total number of training samples for class c , $\bar{f}_c$ is the mean of class c , f_c denotes the set of feature vectors in class c and $\bar{f}$ is the mean of all training samples. The transformation between intra-class and inter-class scatter matrices can be obtained by solving the generalized eigenvalue problem defined as

$$\sum_b^{\wedge} \varphi = \lambda \sum_w^{\wedge} \varphi, \quad (7)$$

where φ and λ are the corresponding eigen vector and eigen value respectively. The eigen vectors represent a new hyperspace.

- Given a new sample u , we transport the sample to the new space φ . Let the sample be $u\varphi$ in the new space. The classification of the sample is

performed by predicting the sample class using distance metric $d(\cdot)$ operator as given below:

$$\arg \min_i d(u\varphi, \bar{f}_i\varphi). \quad (8)$$

Table 1. Houston dataset: number of training and testing samples per class.

Class Name	Number of Samples	
	Training	Testing
Grass healthy	198	1053
Grass stressed	190	1064
Synthetic grass	192	505
Trees	188	1056
Soil	186	1056
Water	182	143
Residential	196	1072
Commercial	191	1053
Road	193	1059
Highway	191	1036
Railway	181	1054
Parking lot1	192	1041
Parking lot 2	184	285
Tennis court	181	247
Running track	187	473
Total	2832	12197

3. PERFORMANCE STUDY

3.1 Data description

For our experiment, we use Houston dataset that consists of a hyperspectral image and a LiDAR-derived DSM. This dataset was distributed at the 2013 GRSS data fusion contest. Both HS and LiDAR data were collected over the University of Houston campus and the neighbouring urban area. The HS data was acquired by a compact airborne spectrographic imager on June 23, 2012, and the average height of the sensor above ground was 5500 ft. The LiDAR data was acquired on June 22, 2012, and the average height of the sensor above ground was 2000 ft. The size of both HS data and LiDAR data is 349×1905 pixels with the spatial resolution of 2.5 m. The HS dataset consists of 144 spectral bands ranging from 0.38 to 1.05 μm . The 15 classes of interests are Grass Healthy, Grass Stressed, Grass Synthetic, Tree, Soil, Water, Residential, Commercial, Road, Highway, Railway, Parking Lot 1, Parking Lot 2, Tennis Court, and Running Track. The distribution of training and testing samples of 15 different classes are shown in Table 1. The “Parking Lot 1” class includes parking garages at the ground level and also in elevated areas, while the “Parking Lot 2” class corresponds to parked vehicles.

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

Fig.3 (a) Correlation coefficient matrix of samples using original hyperspectral responses (b) Correlation coefficient matrix of samples using ICV responses; (c) Band 20 from HS data; (d) Band 20 from ICV responses; (e) DSM, (f) Train Samples, (g) Test Samples, (h) Classification map obtained by proposed method (OA:90.40%)

3.2 Experimental setup

The input HS spectral cube, HS-derived ICV feature cube and DSM are normalised in the range of [0, 1]. Fig. 3(c), (d) and (e) shows band 20 of HS cube, band 20 of ICV cube and DSM respectively. The number of planes in the ICV cube is reduced by taking top PCs having at least 99% cumulative variance. In order to apply DAP on ICV cube, two types of attributes (e.g., area and diagonal of the bounding box) are considered in the experiment, which are areas of 10, 15, 20 and bounding box diagonals of 50, 100, 500. A total of 13 profiles are generated for the area and diagonal attributes of the bounding box.

For selecting suitable bandwidth of Gaussian kernel $\emptyset$, we follow the overall accuracy (OA) curve with respect to $\emptyset$. Fig. 4 shows OA for different values of $\emptyset$. Finally, we consider $\emptyset$ in the range of 111 to 120 which shows the best OA.

For classification, linear discriminant analysis (Fisher discriminant) is used as a classifier.

Fig.4 Overall Accuracy Vs. Values of Ø

Table 2. Classification accuracies of different methods. The dimension of the feature vector is provided in the parenthesis.

		HS (144)	ICV (144)	HS+ICV (288)	HS+ICV +DSM (289)	HS+DSM+ DAP(HS)+ DAP(DSM) (197)	HS+ICV +DSM +DAP(HS) +DAP(DSM) (341)	HS+ICV +DSM+ DAP(ICV) +DAP(DSM) (562)
	OA	74.81	75.35	80.85	81.79	86.23	87.63	90.40
	AA	77.10	78.30	83.25	84.09	87.39	88.74	91.05
	Kappa	0.73	0.73	0.79	0.80	0.85	0.87	0.90
1	Grass Healthy	82.24	92.97	83.10	83.57	83.29	83.38	83.29
2	Grass Stressed	98.50	95.58	96.80	97.18	99.06	98.12	98.40
3	Grass Synthetic	100.00	99.41	99.80	99.80	100.00	100.00	100.00
4	Tree	94.70	86.17	98.86	98.96	96.59	99.05	98.86
5	Soil	96.50	97.16	97.25	97.73	92.42	98.01	99.05
6	Water	86.71	97.90	99.30	99.30	99.30	99.30	99.30
7	Residential	82.18	69.78	76.40	77.24	87.87	85.07	77.05
8	Commercial	31.91	62.39	48.34	55.65	64.58	69.80	69.80
9	Road	59.96	71.95	76.20	78.66	85.84	87.16	91.88
10	Highway	64.96	45.66	75.19	71.72	68.24	74.42	88.61
11	Railway	65.28	62.43	78.08	79.51	98.96	93.07	93.74
12	Parking Lot 1	48.32	44.38	58.69	58.79	74.06	77.81	96.93
13	Parking Lot 2	49.82	52.63	64.56	67.02	63.16	68.77	68.77
14	Tennis Court	97.57	97.98	97.57	97.57	98.79	98.38	100.00
15	Running Track	97.89	98.10	98.52	98.73	98.73	98.73	100.00

3.3 Classification result

The classification results, shown in Table 2, are represented by overall accuracy (OA), average accuracy (AA) and kappa coefficient (Man *et al.* 2015). The metrics OA and AA are in percentage. Kappa coefficient does not have a unit. The best results are shown in boldface putting the number of features in parentheses. To represent the effectiveness of our proposed method in classification accuracy we successively compare it with different methods. We use the following titles to represent the associated descriptions in Table 2 (i) HS: Hyperspectral data, (ii) ICV: Inverse Coefficient of Variations Cube, (iii) HS+ICV: Stacked HS and ICV, (iv) HS+ICV+DSM: HS, ICV and DSM concatenated, (v) HS+DSM+DAP(HS)+DAP(DSM): Stacked HS, DSM, DAP on PCs of HS and DAP on DSM (vi) HS+ICV+DSM+DAP(HS)+DAP(DSM): Stacked HS, ICV, DSM, DAP on PCs of HS and DAP on DSM, (vii) HS+ICV+DSM+DAP(ICV)+DAP(DSM): Stacked HS, ICV, DSM, DAP on PCs of ICV and DAP on DSM.

Table 2 shows the classification accuracies of different methods. The ICV features achieve slightly higher OA than spectral features from HS cube. If we look at the OA of individual classes we can see that for some classes the accuracy is increased and for rest of them the accuracy is decreased. For example, OA using ICV feature for the classes “Water”, “Commercial”, “Road” and “Parking Lot2” are increased by 11.19%, 30.48%, 11.99%, 2.81% respectively in comparison to spectral features. On the other hand, spectral features perform better than ICV for the classes “Grass Stressed”, “Grass synthetic”, “Tree”, “Residential”, “Highway”, “Railway”, “Parking Lot 1” by 2.92%, 0.59%, 8.53%, 12.4%, 19.3%, 2.85%, 3.94% respectively compared to ICV features. But, considering the above result we hypothesis that the features obtained from the two spaces are complementary to each other. So, we combine these features using layer stacking and find that the combined features improve OA for the classes “Tree”, “Water”, “Road”, “Highway”, “Railway”, “Parking Lot 1” and “Parking Lot 2” by 4.16%, 1.4%, 4.25%, 10.23%, 12.80%, 10.37% and 11.93% respectively than using them individually which leads to improvement of overall accuracy for all classes by 6.04%. Adding DSM with the responses from HS and ICV, overall accuracy is improved by 0.94%. The spectral and spatial features from HS stacked with DSM and its spatial features achieves the OA of 86.23%. Stacking ICV features with them increases the OA by 1.4%. It is also evident from Table 2 that spatial features from HS data combined with spectral responses from HS, ICV, DSM and spatial feature from DSM gives OA (87.63%), which is 2.77% lower if we replace the spatial feature extracted from the HS with ICV. Concatenating spatial features from both ICV and DSM, HS, ICV and DSM together give 90.40% overall accuracy which is 8.61% higher than the accuracy obtained by HS, ICV and DSM together. Fig. 3(f) and Fig. 3(g) respectively depict the ground truth of train and test pixels using color map, and finally Fig. 3(f) shows the classification map of the test pixels obtained by our proposed approach respectively.

4. CONCLUSION

In this paper, we propose a novel framework for land cover classification. This framework proposes a novel signal strength based feature extraction technique from hyperspectral data. The proposed feature adds significant complementary information with the hyperspectral data and improves accuracy by as high as 6.04%. ICV responses preserve the local structure of the spectral responses by analysing band to band similarity. Also, the spatial feature derived from ICV has better discriminative capability than the original one. Our proposed framework achieves overall accuracy of 90.40% on Houston dataset which is competent with the existing literature. In our future work, we aim to devise an effective method for fusing data from multiple sources.

REFERENCES

- Man, Q., Dong, P. and Guo, H., 2015. Pixel-and feature-level fusion of hyperspectral and lidar data for urban land-use classification. *International Journal of Remote Sensing*, 36(6), pp.1618-1644.
- Luo, S., Wang, C., Xi, X., Zeng, H., Li, D., Xia, S. and Wang, P., 2016. Fusion of airborne discrete-return LiDAR and hyperspectral data for land cover classification. *Remote Sensing*, 8(1), p.3.
- Khodadadzadeh, M., Li, J., Prasad, S. and Plaza, A., 2015. Fusion of hyperspectral and LiDAR remote sensing data using multiple feature learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6), pp.2971-2983.
- Ghamisi, P., Höfle, B. and Zhu, X.X., 2017. Hyperspectral and LiDAR data fusion using extinction profiles and deep convolutional neural network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(6), pp.3011-3024.
- Dalla Mura, M., Benediktsson, J.A., Waske, B. and Bruzzone, L., 2010. Morphological attribute profiles for the analysis of very high resolution images. *IEEE Transactions on Geoscience and Remote Sensing*, 48(10), pp.3747-3762.
- Rasti, B., Ghamisi, P., Plaza, J. and Plaza, A., 2017. Fusion of Hyperspectral and LiDAR Data Using Sparse and Low-Rank Component Analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 55(11), pp.6354-6365.
- Jahan, F. and Awrangjeb, M., 2017. Pixel-Based Land Cover Classification by fusing Hyperspectral and LiDAR Data. *International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, pp.18-22.
- Hinton, G.E. and Roweis, S.T., 2003. Stochastic neighbor embedding. In *Advances in neural information processing systems* (pp. 857-864).
- Sharma, K.K. and Krishna, H., 1994. Asymptotic sampling distribution of inverse coefficient-of-variation and its applications. *IEEE Transactions on Reliability*, 43(4), pp.630-633.

- Dalla Mura, M., Villa, A., Benediktsson, J.A., Chanussot, J. and Bruzzone, L., 2011. Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis. *IEEE Geoscience and Remote Sensing Letters*, 8(3), pp.542-546.
- Fisher, R.A., 1992. Statistical methods for research workers. In *Breakthroughs in Statistics* (pp. 66-70). Springer, New York, NY.
- Jolliffe, I.T. and Cadima, J., 2016. Principal component analysis: a review and recent developments. *Phil. Trans. R. Soc. A*, 374(2065), p.20150202.
- Fisher, R.A., 1936. The use of multiple measurements in taxonomic problems. *Annals of human genetics*, 7(2), pp.179-188.
- Jonson, R.A. and Wichern, D.W., 1992. *Applied multivariate statistical analysis*.
- Li, T., Zhu, S. and Ogihara, M., 2006. Using discriminant analysis for multi-class classification: an experimental investigation. *Knowledge and information systems*, 10(4), pp.453-472.