Front. Earth Sci. 2017, 11(4): 774–783
DOI 10.1007/s11707-016-0611-2

RESEARCH ARTICLE

A mutual information-Dempster-Shafer based decision ensemble system for land cover classification of hyperspectral data

Parham PAHLAVANI, Behnaz BIGDELI (✉)

Center of Excellence in Geomatics Engineering in Disaster Management, School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran 1439957131, Iran

Abstract Hyperspectral images contain extremely rich spectral information that offer great potential to discriminate between various land cover classes. However, these images are usually composed of tens or hundreds of spectrally close bands, which result in high redundancy and great amount of computation time in hyperspectral classification. Furthermore, in the presence of mixed coverage pixels, crisp classifiers produced errors, omission and commission. This paper presents a mutual information-Dempster-Shafer system through an ensemble classification approach for classification of hyperspectral data. First, mutual information is applied to split data into a few independent partitions to overcome high dimensionality. Then, a fuzzy maximum likelihood classifies each band subset. Finally, Dempster-Shafer is applied to fuse the results of the fuzzy classifiers. In order to assess the proposed method, a crisp ensemble system based on a support vector machine as the crisp classifier and weighted majority voting as the crisp fusion method are applied on hyperspectral data. Furthermore, a dimension reduction system is utilized to assess the effectiveness of mutual information band splitting of the proposed method. The proposed methodology provides interesting conclusions on the effectiveness and potentiality of mutual information-Dempster-Shafer based classification of hyperspectral data.

Keywords mutual information, Dempster-Shafer, hyperspectral, classification, support vector machine

Received January 30, 2016; accepted August 8, 2016

E-mail: bigdeli@ut.ac.ir

1 Introduction and background

The rich spectral information of hyperspectral images (HSI) increases the capability to distinguish different physical materials, leading to the potential of more accurate image classification. In this context, hyperspectral images have been successfully used for classification problems that require very precise description in spectral feature space. Extensive literature is available on the classification of HSIs. Maximum likelihood or Bayesian estimation methods (Jia, 2002), decision trees (Goel et al., 2003), neural networks (Del Frate et al., 2007), genetic algorithms (Vaiphasa, 2003), and kernel-based techniques (Müller et al., 2001; Camps-Valls and Bruzzone, 2005) have been widely investigated in this direction. One of the most popular classification methods is Support Vector Machines (SVMs) defined by Vapnik, which is a large margin based classifier with a good generalization capacity in the small-size training set problem with high-dimensional input space (Vapnik, 1998). Recently, SVMs have been successfully applied in the classification of hyperspectral remote-sensing data. Camps-Valls and Bruzzone (2005) demonstrated that SVMs perform equal or better than other classifiers in terms of accuracy on HSI.

Akbari proposed a spatial-spectral classification method based on marker based minimum spanning forest (MMSF) algorithm and SVM classification. The proposed method achieved improvement on different hyperspectral data (Akbari et al., 2016). Li developed a new framework for the classification of hyperspectral scenes that pursues the combination of multiple features. The ultimate goal of the proposed framework is to be capable to cope with linear and nonlinear class boundaries present in the data, thus following the two main mixing models are considered for hyperspectral data interpretation. In addition of the above crisp classification strategies, some research focuses on fuzzy classification of HSI (Li et al., 2015).

Each pixel from different satellites might represent several kilometres of land. In this sense, fuzzy classification would represent more natural mixtures and transition zones. Also, in the presence of mixed coverage pixels, crisp classifiers produced errors of omission and commission. Pepe and colleagues compared the performance of crisp and fuzzy classification on remotely sensed data. The results showed that fuzzy classifiers outperformed the crisp classifiers (Pepe et al., 2010). In our remote sensing field of interest, a number of approaches have been considered for fuzzy classification of hyperspectral data. Yu and colleagues applied a fuzzy k-nearest neighbour classification method for classification of HS data (Yu et al., 2002). Experiments on AVIRIS hyperspectral data in this paper confirmed higher effectiveness of fuzzy classifiers in comparison to that of crisp classifiers on HSI. Borasca applied a fuzzy version of SVM classifier on hyperspectral data (Borasca et al., 2006). In this research, the proposed approach outperformed crisp classification. Kasiri Bidhendi et al. (2007) applied fuzzy unsupervised classification approaches on hyperspectral data. They used fuzzy C-means clustering (FCM) and fuzzy relational clustering (FRC). Zhang and Qiu successfully utilized an unsupervised neuro-fuzzy system for classification of hyperspectral data (Zhang and Qiu, 2012). They used Gaussian fuzzy self-organizing map (GFSOM). Di and colleagues applied fuzzy integral fusion on hyperspectral data in the aim of anomaly detection (Di et al., 2008).

A large number of features can become a curse in terms of accuracy if enough training samples are not available, i.e., due to the Hughes phenomenon in most of traditional classification techniques (Li et al., 2011). Conventional classification strategies often can-not overcome mentioned problem. Some researchers tried to use dimension reduction techniques to overcome this weakness. The existing dimension reduction methods mainly include two categories: the feature selection methods and the feature extraction methods. Imani and Ghasemian applied discriminant analysis and principal component analysis (PCA) as features extraction strategies on HSI (Imani and Ghasemian, 2015). The experimental results show good performance of their proposed method. Shen utilized a feature selection method based on particle swarm optimization (PSO) as feature selection method on HSI (Shen et al., 2015). Moreover, there exists significant research on dimension reduction techniques to overcome data redundancy of HSI. Nevertheless, the main drawback of feature selection methods is related to the loss of information through the elimination of some bands and the main drawback of feature extraction methods is related to change the spectral characteristics of HSI.

Recently, new research focuses on multiple classifier systems (MCS) to overcome the weaknesses of single classifiers (Benediktsson et al., 2007; Ceamanos et al., 2010). Ham successfully applied ensemble strategies to hyperspectral data (Ham et al., 2005). Also, Ceamanos et al. (2010) applied a multiple classifier system based on SVM for improvement of classification results on hyperspectral images. In their proposed method, the hyperspectral data set is decomposed into a few data sources according to the similarity of the spectral bands. Then, each source is processed separately by performing classification based on SVM. Finally, all outputs are used as input for final decision fusion performed by an additional SVM classifier. Results of the experiments show that the proposed multiple SVM system outperforms a standard classification of hyperspectral data.

This paper describes a new classifier ensemble approach based on mutual information (MI) and Dempster-Shafer (DS) to classify HSI. To address the capabilities of the proposed method, some conventional classification strategies are compared to the new results.

2 Proposed decision ensemble system based on mutual information and Dempster-Shafer

This paper proposes a decision ensemble system based on MI and DS for land cover classification of hyperspectral data. Figure 1 represents the general structure of the proposed methodology. First, MI is applied to group HSI cubes into partitions through a computation of the similarity measure of the spectral information. Second, the proposed methodology applies a fuzzy decision making system based on fuzzy maximum likelihood (FML) on each subset of hyperspectral bands. Then, a decision profile (DP) is established by using all fuzzy classifier decisions. Finally, a DS strategy is applied on DP to produce final classification result on HSI. In previous research, DS showed more accuracy and capability in comparison with those of other fusion procedures. The main motivations of the proposed methodology (especially in comparison with previous classification research) are considered as follows:

i) The proposed method tries to establish a decomposition of spectral information based on MI of spectral bands to overcome the high dimensionality problem of hyperspectral data. The main drawback of dimension reduction techniques in this field is related to the loss of information through the elimination of some bands in feature selection and to the changing of spectral information through transformation in feature extraction methods. By using MI, the proposed method tries to overcome these weaknesses by a system that enables the use of the entire high dimensional hyperspectral image space. Also, in experiments, we try to show this effectiveness by comparing the proposed method with conventional dimension reduction strategies.

ii) The proposed method applied a fuzzy decision

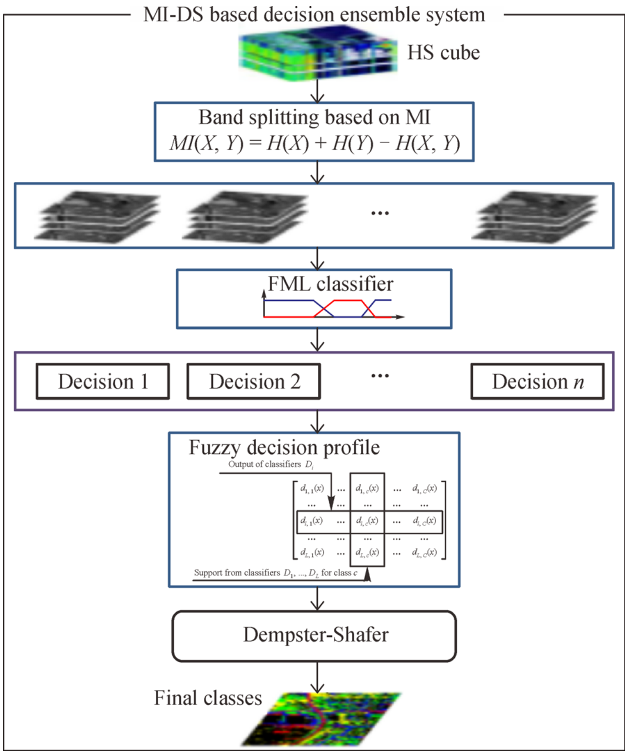


Fig. 1 Flowchart of the mutual Information-Dempster-Shafer based decision ensemble system for classification of HSI.

making algorithm based on fuzzy maximum likelihood to overcome the presence of mixed coverage pixels in HSI. Also, the proposed method tries to compare this fuzzy system with SVM as one of the most powerful crisp kernel based classification methods.

iii) Because of the higher performance of the ensemble learning systems rather than single methods, the proposed method fused the ensemble of fuzzy classifiers based on Dempster-Shafer. In addition, this fuzzy ensemble system is compared with crisp ensemble of SVMs. In this crisp version, we utilized weighted majority voting (WMV) as fusion method on SVMs.

2.1 Spectral information splitting by mutual information

The basic principle of spectral information decomposition is that the spectral bands which have high similarity can be grouped into one group and the ones with little redundancy can be separated into different groups (Su et al., 2011). This paper applied MI as a similarity measure between spectral bands. It is known that independence between bands is one of the key issues to obtain relevant subsets of bands for classification purposes. The use of information measures, such as MI, in order to quantify the degree of independence provides a methodology to find generalized correlations among image bands. Not only is MI widely used as a criterion for measuring the degree of independence between random variables but it also measures how much a certain variable can explain the information content about another variable, i.e., being a generalized correlation measure. MI can be interpreted as a generalized correlation measure, which includes the linear and nonlinear dependence between variables (Chen et al., 2003).

Thus, this technique exploits this concept for band selection in order to reduce data redundancy and non-useful information. Martínez-Usó and Li applied MI as a successful similarity measure between spectral bands in feature selection (Martínez-Usó et al., 2006; Li et al., 2011). In addition, Bigdeli utilized MI measure as a band splitting strategy through a crisp classification of hyperspectral data (Bigdeli et al., 2013).

Entropy is a measure of uncertainty of random variables, and is a frequently used evaluation criterion of feature selection. If a discrete random variable X has Φ alphabets and the probability density function is $p(x)$, $x \in \Phi$, the entropy of X is defined as:

$$H(X) = -\sum_{x \in \Phi} p(x) \log p(x). \tag{1}$$

In the task of band grouping, the entropy of each band is computed by using all spectral information of this band. For two discrete random variables X and Y, which have Φ and Ψ alphabets and their joint probability density function is $p(x, y)$, $x \in \Phi$, $y \in \Psi$, the joint entropy of X and Y is defined as:

$$H(X,Y) = -\sum_{x \in \Phi} \sum_{y \in \Psi} p(x,y) \log p(x,y). \tag{2}$$

The MI is usually used to measure the correlation between two random variables and it is defined as

$$MI(X,Y) = H(X) + H(Y) - H(X,Y), \tag{3}$$

where $H(X,Y)$ is the joint entropy.

The redundancy between two bands will be greater when the value of MI is larger. During the process of HSI cube splitting based on MI, the basic principle is that the bands are divided into groups according to local minima points of bands' MI. These local minima points can be obtained automatically by comparing the neighbourhoods of every point.

2.2 Fuzzy maximum likelihood

The proposed method tries to benefit from a fuzzy classification strategy based on fuzzy maximum likelihood. FML is a fuzzy version of the conventional ML parameters: fuzzy mean and fuzzy variance-covariance matrix. The fuzzy mean can be defined as:

$$\mu_c(x) = \left(\sum_{i=1}^{N} f_c(x_i)x_i\right) \Big/ \left(\sum_{i=1}^{N} f_c(x_i)\right), \tag{4}$$

where N is the total number of patterns, f_c is the membership function (MF) of class c, and x_i is the i^{th} pattern. The fuzzy variance-covariance matrix can be defined as

$$\sum\nolimits_c = \frac{\sum_{i=1}^{N} f_c(x_i)(x_i - \mu_c)(x_i - \mu_c)^{\mathrm{T}}}{\sum_{i=1}^{N} f_c(x_i)}. \tag{5}$$

The detailed description of FML was introduced in Chen (1999).

2.3 Dempster-Shafer as fuzzy decision fusion method

Decision ensemble systems are successfully applied on various types of data to improve single decision results. In the case of classification, these methods can improve classification accuracy in comparison to a single classifier by combining different classification algorithms or variants of the same classifier (Kuncheva, 2004). In such systems, a set of classifiers is produced at first and then combined by a specific fusion method. The resulting classifier is generally more accurate than any of the individual classifiers that make up the ensemble. The possible ways of combining the outputs of the L classifiers depend on what information (crisp or fuzzy) can be obtained from the individual members. In crisp category each classifier only outputs a unique class and finally a vector of classes is produced for each sample. Also the second type of classifier produces fuzzy output which means that in this case the classifier associates a confidence measurement for each class and finally produces a vector for every classifier and a matrix for the ensemble of classifiers. The output of the fuzzy classifiers that provides the class belongingness of an input pattern to different classes is arranged in a matrix form defined as a DP matrix. The DP matrix for L classifiers and C classes is shown in Fig. 2.

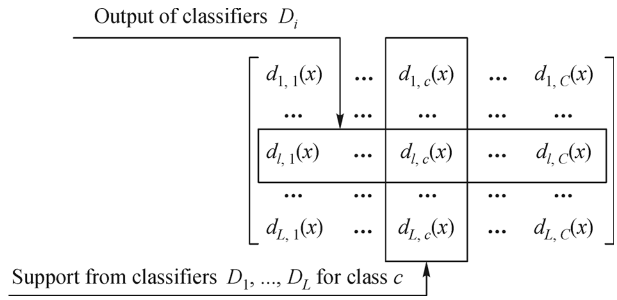


Fig. 2 Decision profile in fuzzy decision ensemble systems.

DS is one of the most powerful methods the can be applied on decision profile to fuse the results of fuzzy classifiers. Mathematical theory of evidence was first introduced by Dempster in the 1960's, and later extended by Shafer (Shafer, 1976). This theory, which represents both imprecision and uncertainty, appears as a more flexible and general approach than Bayesian Theory. The following steps will summarize our DS as a fuzzy decision fusion method:

Step 1: for $j = 1,...,C$ calculate the mean of the decision profile $DP(x)$ of all members of w_j from the data set Z. Call the mean a decision template DT_j:

$$DT_j = \frac{1}{N_j} \sum_{\substack{z_k \in w_j \\ z_k \in Z}} DP(Z_k), \tag{6}$$

where N_j is the number of element of Z from w_j.

Step 2: after defining $DT_1,...,DT_C$, let DT_j^i denote the i^{th} row of decision template DT_j. Denote by $D_i(x)$ the (soft label) output of D_i, that is, $D_i(x) = [d_{i,1}(x),...,d_{i,c}(x)]^{\mathrm{T}}$: the i^{th} row of decision profile $DP(x)$. We calculate the "proximity" Φ between DT_j^i and the output of classifier D_i for the input x as follows:

$$\Phi_{j,i}(x) = \frac{(1 + \|DT_j^i - D_i(x)\|^2)^{-1}}{\sum_{k=1}^{C} (1 + \|DT_k^i - D_i(x)\|^2)^{-1}}, \tag{7}$$

where $\|.\|$ is any matrix norm. For example, we can use the Euclidean distance between the two vectors. Thus, for each decision template we have L proximities.

Step 3: using previous equation we calculate for every class, $j = 1,...,c$; and for every classifier, $i = 1,...,L$, the following belief degrees:

$$b_j(D_i(x)) = \frac{\Phi_{ji}(x)\prod_{k\neq j}(1-\Phi_{k,i}(x))}{1-\Phi_{ji}(x)[1-\prod_{k\neq j}(1-\Phi_{k,i}(x)]}. \quad (8)$$

Step 4: the final degree of support are

$$\mu_j(x) = K\prod_{i=1}^{L} b_j(D_i(x)),\ j = 1,...,c, \quad (9)$$

where K is a normalizing constant.

The most important reason for using DS is that this method is more powerful than other fuzzy decision fusion methods. Moreover, some previous research showed the high performance of this method (Kuncheva, 2004; Breve et al., 2007).

Table 1 AVIRIS Indiana Pine land covers classes and available reference samples

Class	Land cover class	Samples
1	Corn-no till	1434
2	Corn-minimum till	834
3	Grass/pasture	497
4	Grass/trees	747
5	Hay-windrowed	489
6	Soybeans-no till	968
7	Soybeans-minimum till	2468
8	Soybeans-clean till	614
9	Woods	1294

3 Study areas and data sets

The performance of the proposed methodology is tested on two well-known hyperspectral data sets. The first data is made up of a 145×145 pixel portion of the AVIRIS image acquired over north-western US, Indian Pine, acquired in June 1992.

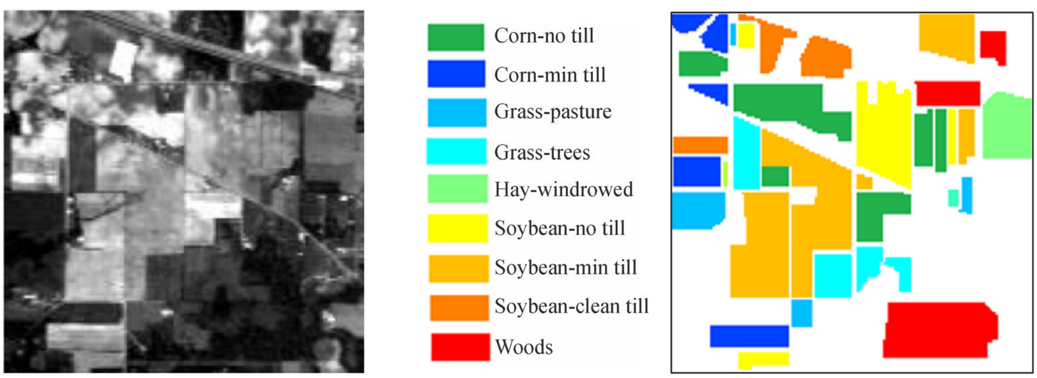
The Indian Pine data are available through Purdue University. The AVIRIS data set contains 220 spectral bands in the wavelength range 0.4–2.5 μm although not all of the 220 original bands are employed in the experiments since 18 bands are affected strongly by atmosphere absorption phenomena and are consequently discarded. Hence, the dimensionality for the AVIRIS Indian Pine data set is 202. Figure 3 shows the original and the reference data of AVIRIS Indian Pine data. From the 16 different land cover classes available in the original ground truth of AVIRIS data; seven were discarded since only few training samples were available for them.

The remaining nine land cover classes were used to generate a set of training data and a set of testing data (Table 1).

The second data set is from Pavia University as another common hyperspectral data. This data set has been acquired by the German ROSIS sensor during a flight campaign over Pavia, northern Italy. Figure 4 represents the original and the reference data of ROSIS data. In ROSIS data, the number of spectral bands is 103 covering the wavelength range from 0.43 μm to 0.86 μm. This data set contains 610×340 pixels with 1.3 m per pixel geometric sampling distance. Table 2 shows the reference data for Pavia University data. Training and testing samples were selected from different areas of the images. They are spatially disjointed.

4 Experimental results

To assess the performance of the proposed MI-DS system rather than dimension reduction, we use PCA and PSO, respectively, as common feature extraction and feature selection strategies. Also, a crisp decision ensemble system based on SVM as crisp classification and WMV as fusion method are utilized to compare with the proposed method. In all cases, we used the fixed training set to train classification methods.



(a) (b)

Fig. 3 AVIRIS Indian Pine data. (a) original data and (b) ground truth.

4.1 Experiments with the AVIRIS Indian Pine data set

Based on the proposed method, it is necessary to split the hyperspectral data into subsets based on MI as similarity measure between spectral bands. Figure 5 shows the MI results.

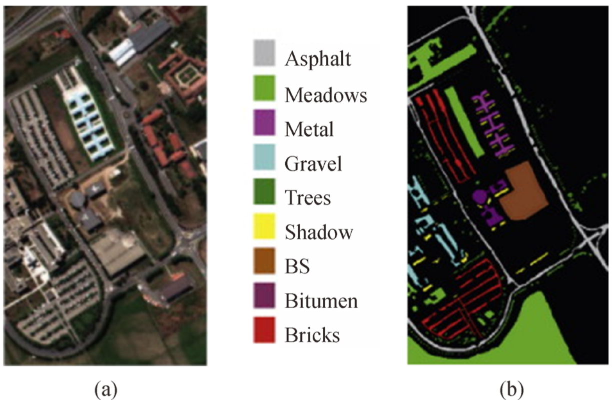


Fig. 4 ROSIS Pavia University data, (a) original data and (b) ground truth.

Table 2 ROSIS Pavia University lands cover classes and available reference samples

Class	Land cover class	Samples
1	Trees	524
2	Asphalt	548
3	Bitumen	375
4	Gravel	392
5	Painted metal sheets	265
6	Shadows	231
7	Self-blocking bricks	514
8	Meadows	540
9	Bare soil	532

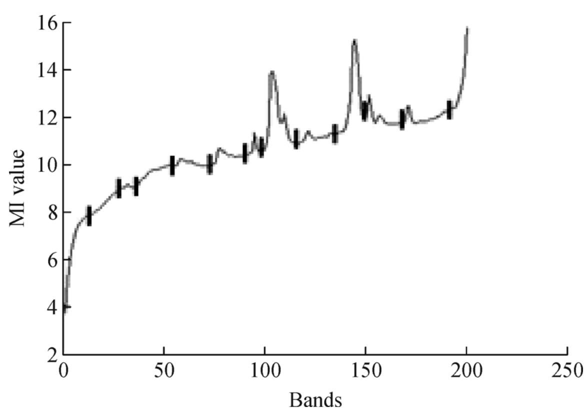


Fig. 5 Splitting of AVIRIS HSI based on local minima of MI (Bigdeli et al., 2013).

Local minima points of this figure (black short line) correspond to the bands with low redundancy. Table 3 illustrates 12 subsets of spectral bands and their overall accuracies were obtained by FML and SVM classification methods.

Table 3 Band numbers in splitting of spectral information of AVIRIS data and classification results by FML and SVM on bands subsets

No.	No. of bands	FML	SVM
1	1–18	59.28	57.33
2	19–33	62.12	64.67
3	34–44	68.24	65.33
4	45–57	57.36	53.56
5	58–77	64.45	62
6	78–105	65.68	66
7	106–125	57	54
8	126–131	48.44	48.89
9	132–147	50.18	50.22
10	148–157	55	51.11
11	158–170	53	49.33
12	171–202	58.46	53.11

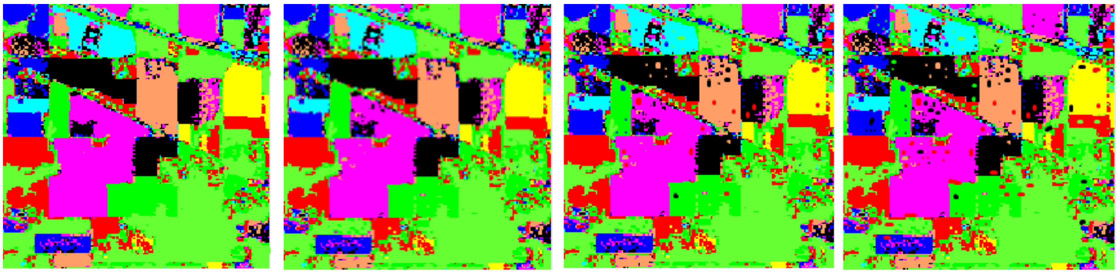
By analysing Table 3 and Fig. 5, AVIRIS data was decomposed into 12 spectral information groups. Afterwards, FML and SVM were applied to classify each subset of HSI data. In Table 3, we reported overall accuracies for these two classification strategies. In more detail, FML represented higher accuracy for most of the spectral groups. However, for some of the groups, SVMs were slightly better than FML.

After classifying each subset separately, DS and WMV were applied to fuse the final decision of all subsets. In order to compare the proposed method with dimension reduction results, PCA and PSO were utilized on HSI spectral bands. Table 4 illustrates overall accuracies and kappa coefficients for mentioned classification strategies.

Table 4 Overall accuracies and kappa coefficients for proposed DES and dimension reduction methods on AVIRIS data

	Decision ensemble system		Dimension reduction	
	DS	WMV	PCA	PSO
OA	97.42	94.22	90.02	88.96
Kappa	94.82	91.08	86.24	83.85

By comparing the different classification algorithms, it can be assessed that the proposed method based on MI for fuzzy version, i.e., DS, and crisp version, i.e., WMV, performed better than dimension reduction techniques. In particular, it is possible to observe that MI-DS system provided the best overall accuracy and kappa. The



(a) (b) (c) (d)

Fig. 6 Classification maps obtained by different classification methods of Table 4 on AVIRIS data. (a) MI-DS, (b) MI-WMV, (c) PCA, and (d) PSO.

improvements of MI-DS are 7.4% and 8.5% for PCA and PSO respectively.

Figure 6 shows the final classification maps of MI-based decision ensemble systems and dimension reduction techniques. These classification maps confirm the results of Table 4. Table 5 gives the results in more detail for class by class accuracy.

Table 5 Class by class accuracies for MI-based decision ensemble systems and dimension reduction techniques on AVIRIS data

Methods	MI-DS	MI-WMV	PCA	PSO
Corn-no till	**95.84**	94.54	86.90	86.21
Corn-minimum till	**92.42**	90.48	84.26	84.2
Grass/pasture	**96.84**	94.24	94.79	90.26
Grass/trees	**98.4**	95.8	97.70	93.1
Hay-windrowed	98.32	93.8	**99.51**	92.8
Soybeans-no till	93.89	**94.22**	84.78	85.3
Soybeans-minimum till	**99.6**	98.6	91.08	90.51
Soybeans-clean till	**98.84**	95.3	90.42	88.6
Woods	95.42	94.62	**96.79**	93.1

As it can be seen, the proposed MI-DS based ensemble system is characterized by a better performance with respect to the other methods. First, better values of accuracies (overall accuracy, kappa and class accuracies) are obtained using the proposed MI-DS system. Then, for most of the classes PCA and PSO as dimension reduction strategies provided smaller accuracies. However, for “Soybean-no till” MI ensemble system based on weighted majority voting and for “Hay-windrowed” and “Woods” classes PCA provided better accuracy than the proposed MI-DS method.

4.2 Experiments with the ROSIS Pavia data set

The present paper tries evaluating classification accuracies of the proposed framework using ROSIS HSI data. Same as the previous experiment on AVIRIS data, Pavia hyperspectral data is converted into spectral subsets based on MI as similarity measure between spectral bands. Figure 7 illustrates the MI results.

As shown by Fig. 7, eight spectral subsets are obtained by the MI proposed method. In addition, Table 6 shows spectral subsets and overall accuracies obtained by fuzzy maximum likelihood and SVM on these spectral subsets.

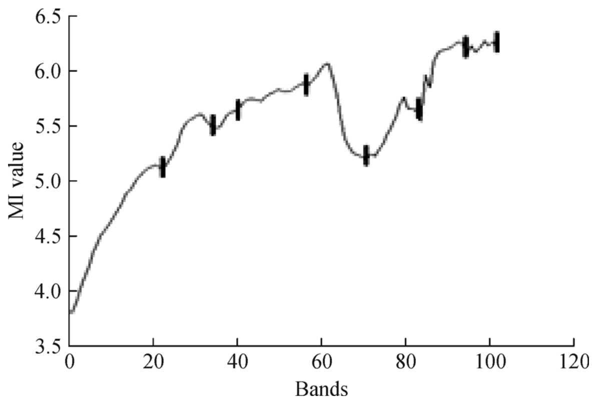


Fig. 7 Splitting of Pavia ROSIS HSI based on local minima of MI (Bigdeli et al., 2013).

Table 6 Band numbers in splitting of spectral information of ROSIS data and classification results by FML and SVM on bands subsets

No	No of bands	FML	SVM
1	1–22	65.96	62.4
2	23–33	62.6	58.56
3	34–46	68.28	64.11
4	47–57	58.28	54.86
5	58–73	70.45	68
6	74–84	68.6	66.7
7	85–95	64.3	60.4
8	96–103	58.9	58.1

Table 7 provides a comparison between the proposed MI-DS framework with dimension reduction systems and decision ensemble system based on WMV-MI.

The overall accuracies and kappa coefficients are similar to the previous experiment on AVIRIS data. The proposed fuzzy decision ensemble system based on the MI-DS

Table 7 Overall accuracies and kappa coefficients for proposed DES and dimension reduction methods on Pavia ROSIS data set

Accuracy	Decision ensemble system		Dimension reduction	
	DS	WMV	PCA	PSO
OA	98.52	96.1	87.24	84.48
Kappa	94.2	92.07	84.2	81.8

system provided the best accuracy rather than the crisp decision ensemble system based on WMV and dimension reduction methods. This improvements are 11.2% and 14.04% for PCA and PSO, respectively. These results confirm that the proposed method based on the using of all spectral bands through an ensemble system based on MI-DS provides significant improvement rather than dimension reduction techniques. Figure 8 illustrates classification maps of the decision ensemble and dimension reduction methods.

Furthermore, Table 8 compared all classification strategies for class by class accuracy. These results exhibit the higher performance of the proposed MI-DS method for

Table 8 Class by class accuracies for MI-based decision ensemble systems and dimension reduction techniques on ROSIS data

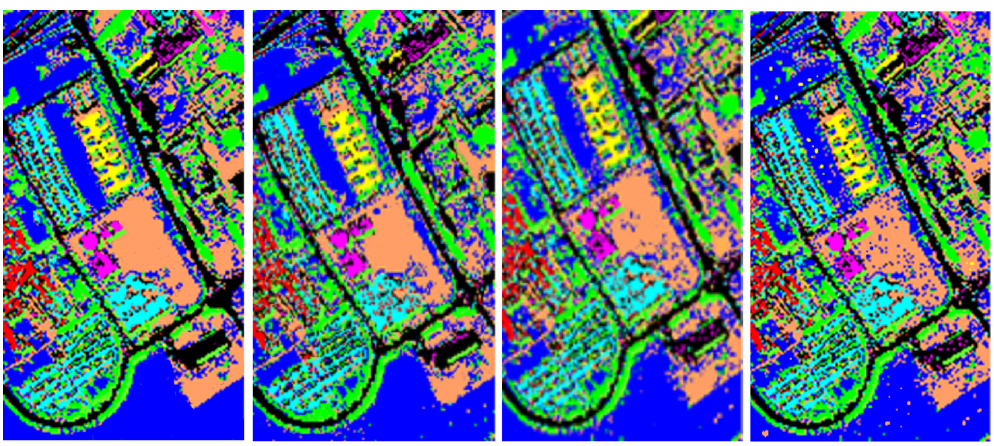
Methods	MI-DS	MI-WMV	PCA	PSO
Trees	**98.20**	97.8	92.4	91.83
Asphalt	**98.82**	96.92	85.76	82.55
Bitumen	89.47	89.2	**91.42**	77.37
Gravel	**99.93**	98.76	89.78	87.18
Painted metal sheets	**99.85**	96.98	93.58	90.53
Shadows	97.69	97.02	**98.68**	89.52
Self-blocking bricks	**98.40**	92.12	**90.9**	90.45
Meadows	**94.69**	93.84	92.84	92.32
Bare soil	98.43	97.68	79.6	72.44

most of the classes. Further analysis of Table 8 shows that the mutual information based DS system produced higher accuracies for most of the classes; however, for "Bitumen" and "Shadows" classes dimension reduction based on PCA provided higher accuracies than the proposed method. This can be explained by smaller number of training samples for these two classes.

It should also be noted that the running time of the experiments has a major role in assessment of the results. First, because of the band splitting on hyperspectral data and then classification of each subset; i.e., in each single classification (12 times for AVIRIS and 8 times for ROSIS data), we did not use all spectral bands. Based on Tables 3 and 6, the number of spectral bands in each subset is small; consequently, classification of them takes just a few seconds. Regarding the above comment, the classification of subsets required only a few seconds, whereas the fusion of these classifiers needed more time. Second, comparison of the time of dimension reduction methods with decision fusion methods demonstrates that DS (as decision fusion method) method improved PSO 14% for ROSIS data. The time of the DS is 5 minutes longer than PSO, because PSO is a searching optimization method that requires more time. Also, PCA is a faster method than decision fusion methods and PSO. However, decision fusion methods provided 8%–10% improvement over PCA in overall accuracy. Finally, it can be concluded that the proposed method provided more improvement in classification of hyperspectral data although it is slightly slower than PCA and PSO.

5 Discussion and conclusions

In this paper, we have developed a new framework for a fuzzy decision ensemble system on hyperspectral data which is based on MI and DS. A main contribution of the



(a) (b) (c) (d)

Fig. 8 Classification maps obtained by different classification methods of Table 4 on ROSIS data. (a) MI-DS, (b) MI-WMV, (c) PCA, and (d) PSO.

presented approach is the classification of HSI through a multiple decisions system without losing spectral information. In this work, we give a first step in computation of MI of spectral bands to split HSI into subsets. Then, we designed a fuzzy decision ensemble system based on maximum likelihood as the fuzzy classification technique and DS as the fuzzy decision fusion method. Also, the proposed method executed a crisp decision ensemble system based on SVM as a the crisp classification method and WMV as a crisp fusion method. In addition, two dimension reduction methods (PCA and PSO) were applied to compare with the proposed methodology. Our experimental results show some important points:

• Our band grouping system based on MI tries to solve the high dimensionality problem of HSI. Some previous research tried to select just useful bands in dimension reduction techniques to overcome data redundancies. Nevertheless, the main drawback of dimension reduction techniques is related to the loss of information through the elimination of some bands. The proposed method tries to overcome this weakness by a system that enables the use of the entire high dimensional hyperspectral image space. Comparison of the results in Tables 4 and 7 confirm the effectiveness on the proposed method over dimension reduction.

• Comparing the results of the crisp decision ensemble system based on SVM-WMV with fuzzy system based on DS show further improvement of our fuzzy proposed method. This is because of the existence of mixed coverage pixels in HSI that belong to more than one class.

The proposed method obtained approximately 97% and 98% overall accuracies for Indiana and Pavia hyperspectral data, respectively. Also, the proposed method improved common dimension reduction methods from 7% to 10%. These results suggest that the decomposition of the HSI classification problem into multiple decisions represents an effective way of improving the overall discrimination capability especially for fuzzy classification system.

References

Akbari D, Homayuoni S, Saffari A, Mehrshad N (2016). Mapping urban land cover based on spatial-spectral classification of hyperspectral remote- sensing data. Int J Remote Sens, 37(2): 440–454

Benediktsson J A, Chanussot J, Fauvel M (2007). Multiple classifier systems in remote sensing: from basics to recent developments. In: Haindl M, Kittler J, Roli F, eds. Multiple Classifier Systems. Heidelberg: Springer, 501–512

Bigdeli B, Samadzadegan F, Reinartz P (2013). A multiple SVM system for classification of hyperspectral remote sensing data. Journal of the Indian Society of Remote Sensing, 41(4): 763–776

Borasca B, Bruzzone L, Carlin L, Zusi M (2006). A fuzzy-input fuzzy-output SVM technique for classification of hyperspectral remote sensing images. Signal Processing Symposium, 2006. NORSIG 2006. Proceedings of the 7th Nordic, 2–5

Breve F, Ponti M, Mascarenhas N (2007). Multilayer Perceptron Classifier Combination for Identification of Materials on Noisy Soil Science Multispectral Images. Computer Graphics and Image Processing, 2007. SIBGRAPI 2007. XX Brazilian Symposium on. 239–244

Camps-Valls G, Bruzzone L (2005). Kernel-based methods for hyperspectral image classification. IEEE Trans Geosci Rem Sens, 43(6): 1351–1362

Ceamanos X, Waske B, Benediktsson J, Chanussot J, Fauvel M, Sveinsson J (2010). A classifier ensemble based on fusion of support vector machines for classifying hyperspectral data. International Journal of Image and Data Fusion, 1(4): 293–307

Chen C F (1999). Fuzzy training data for fuzzy supervised classification of remotely sensed images. Asian Conference Remote Sensing (ACRS 1999)

Chen H, Varshney P K, Arora M K (2003). Performance of mutual information similarity measure for registration of multitemporal remote sensing images. IEEE Trans Geosci Rem Sens, 41(11): 2445–2454

Del Frate F, Pacifici F, Schiavon G, Solimini C (2007). Use of neural networks for automatic classification from high-resolution images. IEEE Trans Geosci Rem Sens, 45(4): 800–809

Di W, Pan Q, He L, Cheng Y (2008). Anomaly detection in hyperspectral imagery by fuzzy integral fusion of band-subsets. Photogramm Eng Remote Sensing, 74(2): 201–213

Goel P K, Prasher S O, Patel R M, Landry J A, Bonnell R B, Viau A A (2003). Classification of hyperspectral data by decision trees and artificial neural networks to identify weed stress and nitrogen status of corn. Comput Electron Agric, 39(2): 67–93

Ham J, Yangchi Chen, Crawford M M, Ghosh J(2005). Investigation of the random forest framework for classification of hyperspectral data. IEEE Trans Geosci Rem Sens, 43(3): 492–501

Imani M, Ghasemian H (2015). Feature reduction of hyperspectral images: discriminant analysis and the first principal component. Journal of AI and Data Mining, 3(1): 1–9

Jia X (2002). Simplified maximum likelihood classification for HS data in cluster space. In: IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS ‘02), Toronto, Ontario, Canada, 5: 2578–2580

Kasiri Bidhendi S, Shirazi A, Fotoohi N, Ebadzadeh M (2007). Material Classification of hyperspectral images using unsupervised fuzzy clustering methods. Third International IEEE Conference on Signal-Image Technologies and Internet-Based System, SITIS '07. 619 –623

Kuncheva L (2004). Combining Pattern Classifiers methods and algorithms. Hoboken, New jersey. Canada: John Wiley&sons, INC.

Li J, Huang X, Gamba P, Bioucas-Dias J M B, Zhang L, Benediktsson J A, Plaza A (2015). Multiple feature learning for hyperspectral image classification. IEEE Trans Geosci Rem Sens, 53(3): 1592–1606

Li S, Wu H, Wan D, Zhu J (2011). An effective feature selection method for hyperspectral image classification based on genetic algorithm and support vector machine. Knowl Base Syst, 24(1): 40–48

Martínez-Usó A, Pla F, Sotoca J M, Garcia-Sevilla P (2006). Clustering based multispectral band selection using mutual information, Proceedings of the 18th International Conference on Pattern Recognition (ICPR), 2: 760–763

Müller K L, Mika S, Rätsch G, Tsuda K, Schölkopf B (2001). An introduction to kernel-based learning algorithms. IEEE Trans Neural Netw, 12(2): 181–201

Pepe M, Boschetti L, Brivio P A, Rampini A (2010). Comparing the performance of fuzzy and crisp classifiers on remotely sensed images: a case of snow classification. Int J Remote Sens, 31(23): 6189–6203

Shafer G (1976). A Mathematical Theory of Evidence. Princeton University Press

Shen J, Wang C, Wang R, Huang F, Fan C, Xu L (2015). A band selection method for hyperspectral image classification based on improved particle swarm optimization. International Journal of Signal Processing. Image Processing and Pattern Recognition, 8(4): 325–338

Su H, Yang H, Du Q, Sheng Y (2011). Semi-supervised band clustering for dimensionality reduction of hyperspectral imagery. IEEE Geosci Remote Sens Lett, 8(6): 1135–1139

Vaiphasa C (2003). Innovative genetic algorithm for hyperspectral image classification. In: Proc. Int. Conf. Map Asia.

Vapnik V N (1998). Statistical Learning Theory. New York: Wiley & Sons, Inc.

Yu S, De Backer S, Scheunders P (2002). Genetic feature selection combined with composite fuzzy nearest neighbor classifiers for hyperspectral satellite imagery. Pattern Recognit Lett, 23(1−3): 183–190

Zhang C, Qiu F (2012). Hyperspectral image classification using an unsupervised neuro-fuzzy system. J Appl Remote Sens, 6(1): 1–15

Author Biographies

Parham Pahlavani received the B.Sc. degree in Civil Eng.-Surveying & Geomatics Eng. from industrial K.N.Toosi University, Tehran, Iran in 2001, and the M.Sc. degree in geo-spatial information system (GIS) engineering from the Tehran University, Tehran, Iran in 2005, and the PhD in Geomatics Eng.-GIS from the Tehran University, Tehran, Iran in 2012. Since 2013 he has been working in School of Surveying & Geospatial Eng., College of Eng., University of Tehran as an Assistant Professor. Also, he is a member of the Iran's National Elites Foundation. His current research interests are using soft computing algorithms, machine learning algorithms, agent-based spatial modeling, learning systems, fuzzy control and approximate reasoning in GIS and Remote Sensing Integration.

E-mail: pahlavani@ut.ac.ir

Behnaz Bigdeli has a B.Sc. in Civil Eng.-Surveying & Geomatics Eng. from industrial K.N.Toosi University, Tehran, Iran in 2007 and M.Sc. degrees in Remote Sensing from the University of Tehran, Iran in 2010, and the PhD in Remote Sensing from the University of Tehran, Iran in 2014. Her research are data fusion of remote sensing sensors, machine learning and object extraction. She has published more than 20 peer-reviewed articles in journals and selective conferences.

E-mail: bigdeli@ut.ac.ir