Comparative Study of Intrinsic Dimensionality Estimation and Dimension Reduction Techniques on Hyperspectral Images Using K-NN Classifier

Mahdi Hasanlou, *Student Member, IEEE*, and Farhad Samadzadegan

***Abstract*—Nowadays, hyperspectral remote sensors are readily available for monitoring the Earth's surface with high spectral resolution. The high-dimensional nature of the data collected by such sensors not only increases computational complexity but also can degrade classification accuracy. To address this issue, dimensionality reduction (DR) has become an important aid to improving classifier efficiency on these images. The common approach to decreasing dimensionality is feature extraction by considering the intrinsic dimensionality (ID) of the data. A wide range of techniques for ID estimation (IDE) and DR for hyperspectral images have been presented in the literature. However, the most effective and optimum methods for IDE and DR have not been determined for hyperspectral sensors, and this causes ambiguity in selecting the appropriate techniques for processing hyperspectral images. In this letter, we discuss and compare ten IDE and six DR methods in order to investigate and compare their performance for the purpose of supervised hyperspectral image classification by using K-nearest neighbor (K-NN). Due to the nature of K-NN classifier that uses different distance metrics, a variety of distance metrics were used and compared in this procedure. This letter presents a review and comparative study of techniques used for IDE and DR and identifies the best methods for IDE and DR in the context of hyperspectral image analysis. The results clearly show the superiority of the hyperspectral signal subspace identification by minimum, second moment linear, and noise-whitened Harsanyi–Farrand–Chang estimators, also the principal component analysis and independent component analysis as DR techniques, and the norm L1 and Euclidean distance metrics to process hyperspectral imagery by using the K-NN classifier.**

***Index Terms*—Dimension reduction, distance metric, feature extraction, hyperspectral images, intrinsic dimension estimation (IDE), K-nearest neighbor (K-NN) classifier.**

I. Introduction

As a result of current developments in remote sensing sensors, hyperspectral remote sensing imagery is now widely available and is becoming viable tools for monitoring the Earth's surface [1], [2]. Hyperspectral sensors offer a dense sampling of the spectral range of the sensor, thus facilitating better discrimination among similar ground cover classes than traditional multispectral scanners with low spectral resolution. These images provide valuable observations at hundreds of frequency bands (very narrow spectral bands) that are highly correlated with each other and may contain redundant information [3]. This high dimensionality not only increases computational complexity but also may degrade classification accuracy [4]. As the dimensionality of remotely sensed imagery increases, the classification accuracy improves at first but then declines when the number of the training samples is low and finite and remains constant. This problem is referred to as Hughes phenomenon or Bellman's curse of dimensionality [3]. Subsequently, dimensionality reduction (DR) in hyperspectral image data without losing relevant information about objects for the purpose of classification has become a topic of great interest in recent years [1], [2]. Hyperspectral DR is the transformation of a given hyperspectral image into a meaningful representation with reduced dimensionality. Ideally, the reduced image has a dimensionality that corresponds to the intrinsic dimensionality (ID) of the data. The ID of a data set is the minimum number of free variables needed to model the data without loss [5]. A DR technique transforms given hyperspectral images into new hyperspectral images with dimensionality $\hat{d}$, while preserving the geometry of the hyperspectral image as much as possible.

The true dimensionality of hyperspectral imagery is difficult to determine in practice, since its ID simply cannot be determined by the dimensionality of image data. There are different techniques that estimate ID for hyperspectral images, most of which cannot exactly determine the dimensionality of hyperspectral images [6]. The results of this study can be used to decrease effects of inaccurate ID estimation (IDE) in classifying hyperspectral imagery. Furthermore, the most effective and optimum methods for IDE have not been determined for hyperspectral sensors, and this causes some ambiguity when selecting appropriate techniques for processing hyperspectral images. This letter investigates the performance of ten IDE techniques for hyperspectral images with a variety of six DR methods and the K-nearest neighbor (K-NN) as supervised classifier with variable parameters [7]. The K-NN classifier is one of the most convenient and widely used nonparametric classifier techniques and has been successfully used to classify hyperspectral images by researchers. In addition, the K-NN classifier uses different distance metrics for measuring the distance between their objects, and this may be helpful to process hyperspectral imagery.

Manuscript received May 9, 2011; revised February 20, 2012; accepted February 21, 2012.

The authors are with the Division of Remote Sensing, Department of Surveying and Geomatics Engineering, College of Engineering, University of Tehran, Tehran 1439957131, Iran (e-mail: hasanlou@ut.ac.ir; samadz@ut.ac.ir).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2012.2189547

II. IDE

Several techniques have been proposed in order to estimate the ID in hyperspectral images [5], [6], [8]–[11]. Due to high spectral resolution of a hyperspectral image, classifiers cannot extract many unknown spectral signal sources. Therefore, it is very challenging and difficult to determine how many spectral signal sources are present in a hyperspectral image [8]. In our case (classification of hyperspectral imagery), the accuracy of a classified image depends on the correct extraction of features from the high-dimensional image. This issue can be handled by better estimating the ID in the hyperspectral image. In addition, the value of the ID actually varies with different image frames depending upon the content of the image. Techniques for IDE can be used to circumvent the problem of selecting proper target dimensionality of hyperspectral images. For convenience, we denote the estimation of the ID in hyperspectral images by $\hat{d}$. Methods for IDE can be divided into two groups [12]: 1) estimators based on the analysis of local properties of the data and 2) estimators based on the analysis of global properties of the data. In this letter, we discuss and compare ten IDE techniques for hyperspectral DR. The first three are local IDE techniques which are the correlation dimension (CD) estimator [13], the maximum likelihood (ML) estimator [14], and the Takens (TAK) estimator [15]. The next seven are global IDE techniques: the eigenvalue (EV) estimator [16], the geodesic minimum spanning tree (GMST) estimator [12], the second moment linear (SML) [6], the hyperspectral signal subspace identification by minimum error (Hysime) [8], the Hein (HEI) estimator [17], the Harsanyi–Farrand–Chang (HFC) estimator [5], and the noise-whitened HFC (NWHFC) estimator [5].

III. Dimension Reduction Techniques

DR is the transforming of data to a lower dimensional space in such a way that irrelevant variance in the data is eliminated and detected. DR has an applicable method for data visualization and for extracting critical low-dimensional features. In hyperspectral studies, the favored and optimal low-dimensional features depend on the hyperspectral application. In general, using hyperspectral images increases our ability to classify land use/cover types. However, the image classification approach that has been successfully applied to multispectral data in the past is not as effective for hyperspectral data [18]. One major application of DR is hyperspectral image classification. Using insufficient training samples without DR will cause inaccurate parameter estimation and, consequently, will lead to decreased classification accuracy [19]. One of the approaches to improving the classification performance is to reduce dimensionality via a preprocessing method, which takes the high-dimensional space properties into account. Image DR is the transformation that converts image data from a high-order dimension to a low-order dimension (bands/features). In the case of hyperspectral image classification, to obtain better result, there is no way to incorporate a DR method to achieve a feature space with low dimensionality. By using an IDE method, the dimensionality of the feature space can be determined. In this letter, we discuss and compare six DR techniques for hyperspectral images: principal component analysis (PCA) [20], vertex component analysis (VCA) [21], independent component analysis (ICA) [22], linear discriminate analysis (LDA) [12], factor analysis (FA) [23], and Gram–Schmidt (GS) [24].

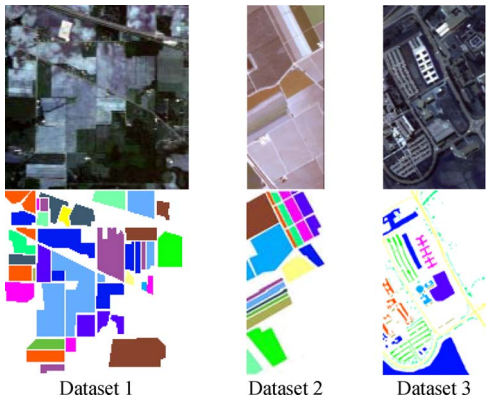

Fig. 1. Hyperspectral images and related ground truth data.

TABLE I
Number of Training Pixels in Each Data Set

Class Name	Dataset 1	Dataset 2	Dataset 3
1	19	1106	2021
2	241	590	5582
3	295	415	645
4	925	803	954
5	937	1190	411
6	22	1049	1531
7	102	3366	393
8	157	1907	1089
9	10	993	279
10	384	303	-
11	91	611	-
12	572	280	-
13	401	350	-
14	25	2193	-
15	243	557	-
16	58	400	-

IV. Experiments and Results

In order to evaluate the potential of IDE methods, three different hyperspectral images [25] with different image pixel sizes were used (Fig. 1). Data set 1 was gathered by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over the Indian Pines test site in Northwestern Indiana and consists of 145 × 145 pixels and 224 spectral reflectance bands in the wavelength range of 0.4–2.5 μm. This scene is a subset of a larger one. The ground truth for data set 1 is composed of 16 classes. We have also decreased the bands to 200 by eliminating bands covering the region of water absorption: {104–108}, {150–163}, and 220 [8]. Data set 2 was collected by the 224-band AVIRIS sensor over Salinas Valley, CA, and is characterized by high spatial resolution (3.7-m pixels). This image has 512 × 217 pixels. As with data set 1, we removed the 20 water absorption bands, which, in this case, are the following bands: {108–112}, {154–167}, and 224. The data set 2 ground truth contains 16 classes. Data set 3 is acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor during a flight campaign over Pavia, Northern Italy. The number of spectral bands is 103 for the Pavia University data set. Data set 3 is 610 × 340 pixels. The geometric resolution of data set 3 is 1.3 m, and the ground truth includes nine classes.

Thirty percent of the ground truth pixels are used for training the classifier. These data sets are selected by considering

TABLE II
DETAILS OF IDE

IDE Estimator	$\hat{d}$ (dataset 1)	$\hat{d}$ (dataset 2)	$\hat{d}$ (dataset 3)
EV	5	6	5
ML	7	7	11
CD	4	2	3
GMST	8	6	16
Hysime	15	23	49
HFC (P_f=10^{-5})	22	14	9
NWHFC (P_f=10^{-5})	10	14	9
HEI	9	3	6
TAK	7	6	7
SML (α=25×10^{-3} to 53×10^{-3})	30	14	59

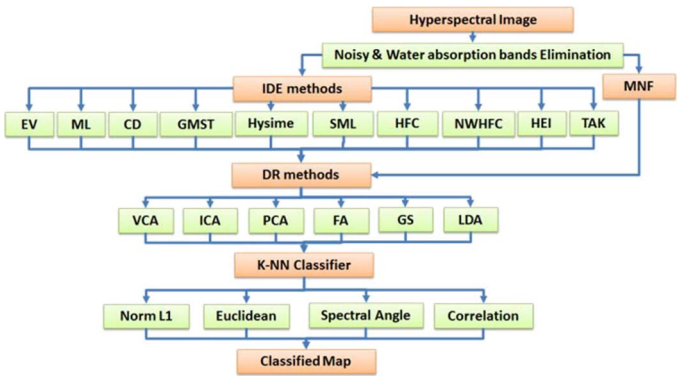

Fig. 2. Flowchart of IDE and DR methods for hyperspectral images.

TABLE III
DIFFERENT K-NN CLASSIFIER DISTANCE METRICS

Distance metric	Formulation
Euclidean	$d_{rs}^2 = (x_r - x_s)(x_r - x_s)^T$
Norm L1	$d_{rs} = \sum_{j=1}^{n} \lvert x_{rj} - x_{sj} \rvert$
Spectral Angle	$d_{rs} = \cos^{-1}(x_r x_s^T / (x_r x_r^T)^{1/2} (x_s x_s^T)^{1/2})$
Correlation	$d_{rs} = 1 - \frac{(x_r - \bar{x}_r)(x_s - \bar{x}_s)^T}{((x_r - \bar{x}_r)(x_r - \bar{x}_r)^T)^{1/2}((x_s - \bar{x}_s)(x_s - \bar{x}_s)^T)^{1/2}}$

different types and the variety of pattern classes. Data set 1 has an agricultural pattern with complex classes, and data set 2 has wide spread classes with different textures. Data set 3 has more complicated classes compared to other data sets with urban facilities. The actual pixels used to train the classifiers of these three data sets are represented in Table I.

Most of the IDE techniques that we used in this study do not have any specific parameter, although Hysime, HFC, NWHFC, and SML have some initial parameters. The Hysime method begins by estimating the signal and the noise correlation matrices; it then selects the subset of eigenvectors. In this manner, the Poisson noise type is selected for noise estimation for every pixel and noise correlation matrix estimates. The Hysime algorithm not only is used for IDE but also prepares a matrix to project the hyperspectral image and, therefore, can be used as a DR technique [8]. In the HFC and NWHFC methods, the false-alarm probability is set to $P_f = 10^{-5}$ [5]. In the SML algorithm, the significance α level must be considered, and dimensionality must be selected for α between 25×10^{-3} and 53×10^{-3} [6]. Table II presents the value of different IDEs $(\hat{d})$ for the three image data sets. From Table II, it is clear that all IDE methods have the same trends in extracting the ID dimension from the hyperspectral image. Also, these methods appear to exhibit a good degree of consistency and behave in a predictable manner for the range of data that they were tested on.

TABLE IV
PERFORMANCE OF K-NN CLASSIFIER IN DATA SET 1 (IN PERCENT)

		EV	ML	CD	GMST	Hysime	HFC	NWHFC	HEI	TAK	SML
ICA	Ed	81.41	83.08	79.12	83.76	85.31	76.57	83.97	84.17	83.09	71.91
	L1	81.23	82.76	79.01	83.92	85.60	79.19	83.88	83.79	82.68	74.70
	Sa	77.53	80.51	74.52	82.44	84.90	76.17	82.83	82.98	80.35	71.93
	Co	72.97	79.31	66.86	81.15	84.93	76.20	81.76	81.73	78.97	71.96
VCA	Ed	78.87	81.49	75.67	82.02	82.45	74.47	83.61	82.85	81.38	68.55
	L1	78.66	81.58	75.52	81.95	82.34	74.24	83.39	82.76	81.41	68.36
	Sa	78.69	81.31	75.30	81.84	82.01	74.11	83.51	82.92	81.26	67.90
	Co	74.36	79.75	67.43	80.39	81.56	73.90	82.60	82.04	79.67	67.82
PCA	Ed	79.25	80.19	77.03	80.41	82.16	81.49	80.48	80.78	80.17	81.56
	L1	79.58	80.86	77.32	81.59	84.38	83.62	82.01	82.33	80.94	83.63
	Sa	77.28	79.03	73.16	79.46	81.85	81.09	79.88	79.84	79.15	81.30
	Co	73.62	76.91	66.43	78.39	81.42	80.94	79.28	79.10	77.32	81.07
GS	Ed	74.10	76.24	72.88	77.03	78.97	79.79	77.78	78.08	76.24	80.40
	L1	74.14	76.14	72.76	76.81	78.68	79.70	77.52	77.76	76.05	80.24
	Sa	72.63	76.39	68.30	77.03	80.02	80.66	78.05	78.22	76.47	81.23
	Co	69.10	73.87	63.50	75.32	79.64	80.47	77.22	76.88	74.23	81.28
FA	Ed	70.66	76.45	70.59	76.92	76.74	77.13	73.59	75.72	74.63	77.50
	L1	70.14	76.07	70.16	77.14	76.41	76.80	73.39	75.45	74.69	77.45
	Sa	73.73	73.85	70.64	74.50	74.38	75.77	73.97	74.34	74.77	76.26
	Co	70.83	72.61	67.02	75.91	75.41	77.66	74.98	76.15	74.64	76.91
LDA	Ed	66.21	68.66	66.00	69.30	71.44	71.30	70.18	70.05	68.58	70.98
	L1	66.25	68.33	66.02	69.39	71.22	70.96	69.77	69.49	68.83	70.54
	Sa	64.23	67.00	63.62	67.67	70.65	70.57	68.97	68.83	66.90	70.54
	Co	62.95	65.69	61.11	66.23	69.78	70.47	67.81	67.49	65.35	70.03

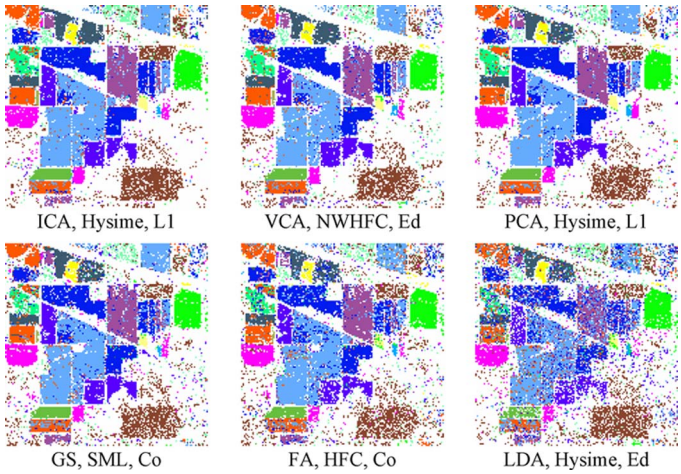

Fig. 3. Best classified map for each DR method for data set 1.

The next step after estimating the ID number is to employ DR methods (Fig. 2). For example, if $\hat{d}$ for the HFC estimator from Table II is 22, it means that 22 features are retained at the DR step and feature extraction method reduced the feature space to 22 bands. Before the extraction of features by using ID numbers, a minimum noise fraction (MNF) was used to remove and reduce noise in the hyperspectral data caused by the image sensor. The MNF transform is a linear transformation which is essentially two cascaded PCA transformations [26]. This algorithm does not have much influence on the classification result but reduces image noise in hyperspectral images. To study the properties of the IDE and DR methods, it is necessary to apply a supervised classification procedure to the extracted features. Here, the K-NN classifier has been adopted for supervised classification. A number of distance metrics were used, which were as follows (Table III): the Euclidean distance (Ed), the norm L1 distance (L1), the spectral angle distance (Sa), and the correlation distance (Co).

In Table III, x_r and x_s are two vectors with the dimension of n, and $\bar{x}_r$ and $\bar{x}_s$ are the mean values of those vectors. There are a lot of indices that are used to evaluate the performance of supervised classification; from these indices, the overall accuracy shows better results in classifying remote sensing images [27]. Then, in this letter, the overall accuracy is used for evaluating the hyperspectral classification procedure.

TABLE V
PERFORMANCE OF K-NN CLASSIFIER IN DATA SET 2 (IN PERCENT)

		EV	ML	CD	GMST	Hysime	HFC	NWHFC	HEI	TAK	SML
ICA	Ed	90.92	91.23	80.06	91.16	89.98	92.53	92.45	86.86	91.09	92.59
	L1	90.84	91.28	80.03	91.12	90.24	92.70	92.69	86.82	90.97	92.82
	Sa	89.19	90.17	68.26	89.43	89.61	92.16	91.96	81.17	89.32	92.29
	Co	87.99	89.47	52.99	88.19	89.46	92.08	91.93	72.47	88.08	92.16
VCA	Ed	90.48	89.99	82.55	90.35	89.45	90.94	91.06	87.57	89.78	91.22
	L1	90.64	90.25	82.50	90.29	90.01	91.29	91.51	87.50	90.08	91.55
	Sa	89.40	89.23	71.44	89.17	89.84	91.26	91.54	82.06	88.58	91.45
	Co	88.13	88.63	52.94	87.91	89.81	91.22	91.40	72.91	87.68	91.36
PCA	Ed	90.99	91.23	82.40	90.94	92.29	91.99	92.18	87.02	90.96	92.10
	L1	91.16	91.46	82.39	91.08	93.11	92.74	92.87	86.94	91.18	92.79
	Sa	90.51	90.89	73.20	90.35	92.56	92.27	92.44	82.45	90.42	92.32
	Co	89.62	90.12	38.32	89.48	92.48	92.12	92.27	72.29	89.46	92.18
GS	Ed	90.05	90.13	82.84	90.01	91.55	91.26	91.25	85.87	89.79	91.26
	L1	89.99	90.06	82.82	89.90	91.65	91.10	91.16	85.72	89.86	91.16
	Sa	87.90	88.03	71.51	87.91	91.39	90.21	90.31	79.83	87.87	90.35
	Co	87.34	87.83	53.04	87.29	91.31	90.06	90.23	63.68	87.24	90.28
FA	Ed	89.84	90.73	82.64	90.27	90.32	90.23	89.93	87.05	90.04	89.87
	L1	89.68	90.70	82.60	90.18	90.24	90.16	89.74	86.92	90.00	89.65
	Sa	88.97	89.80	72.77	89.37	90.51	90.16	90.47	82.38	88.88	89.94
	Co	88.53	89.32	53.00	89.41	91.08	90.44	90.66	73.03	89.15	90.44
LDA	Ed	79.71	82.46	70.57	79.54	90.71	88.92	88.89	71.56	79.75	88.82
	L1	79.60	82.27	70.55	79.59	90.34	88.67	88.66	71.50	79.57	88.67
	Sa	76.29	79.90	63.13	76.10	89.75	88.31	88.11	64.74	76.35	88.08
	Co	73.79	77.86	53.00	73.45	89.45	87.30	87.19	63.02	73.90	87.21

TABLE VI
PERFORMANCE OF K-NN CLASSIFIER IN DATA SET 3 (IN PERCENT)

		EV	ML	CD	GMST	Hysime	HFC	NWHFC	HEI	TAK	SML
ICA	Ed	83.00	82.20	80.47	79.38	76.07	81.97	82.18	83.88	83.64	76.15
	L1	83.00	82.10	80.48	79.33	75.89	82.04	82.20	83.86	83.51	76.02
	Sa	80.51	80.28	77.71	78.59	76.00	80.44	80.37	81.51	81.56	76.20
	Co	79.83	80.19	75.92	78.46	75.74	80.32	80.43	80.78	81.29	75.97
VCA	Ed	80.98	81.09	80.11	81.54	79.65	81.23	81.09	83.19	82.10	80.44
	L1	81.04	81.36	80.05	81.80	79.70	81.44	81.40	83.25	82.23	80.45
	Sa	81.25	79.89	77.91	79.97	79.14	79.87	79.64	81.43	80.95	80.21
	Co	79.12	79.96	75.84	80.00	79.15	79.78	79.54	81.05	80.62	80.21
PCA	Ed	82.61	83.49	80.48	83.58	83.65	83.50	83.60	83.13	83.43	83.84
	L1	82.62	83.54	80.53	83.61	83.75	83.51	83.67	83.17	83.43	83.94
	Sa	82.39	83.95	77.87	84.03	84.13	83.78	83.79	82.85	83.45	84.18
	Co	79.69	83.87	76.26	84.03	84.08	83.56	83.60	81.91	82.98	84.20
GS	Ed	80.33	82.34	79.98	82.57	83.50	81.82	81.89	80.72	81.41	83.44
	L1	80.33	82.47	80.04	82.57	83.38	81.90	81.83	80.76	81.42	83.42
	Sa	78.85	81.31	77.85	81.83	82.65	80.78	80.82	79.56	80.58	82.68
	Co	78.23	81.22	75.89	81.57	82.67	80.62	80.67	79.47	80.38	82.66
FA	Ed	81.13	82.74	80.36	82.97	83.34	82.78	82.36	81.89	81.78	83.30
	L1	81.04	82.68	80.35	82.77	83.30	82.67	82.07	81.86	81.64	83.24
	Sa	81.83	83.59	77.79	83.45	83.53	83.26	83.12	82.47	82.69	83.69
	Co	80.23	83.27	76.28	83.01	83.84	83.03	83.00	81.59	82.80	83.70
LDA	Ed	75.48	76.00	75.56	76.14	75.92	75.91	75.77	75.45	75.46	75.58
	L1	75.55	76.05	75.53	75.86	75.61	75.83	75.78	75.45	75.38	75.35
	Sa	75.35	76.42	75.16	77.03	77.62	76.12	76.03	75.22	75.51	77.54
	Co	75.26	76.39	75.18	76.96	77.65	76.15	76.20	75.36	75.48	77.48

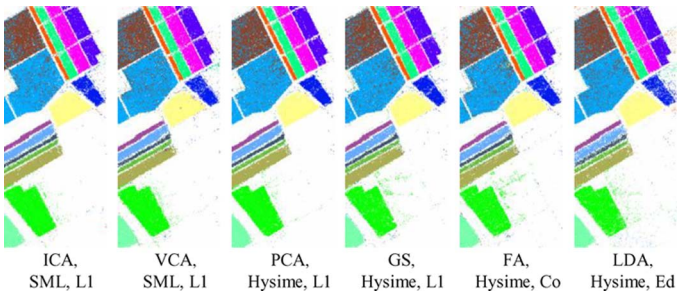

Fig. 4. Best classified map for each DR method for data set 2.

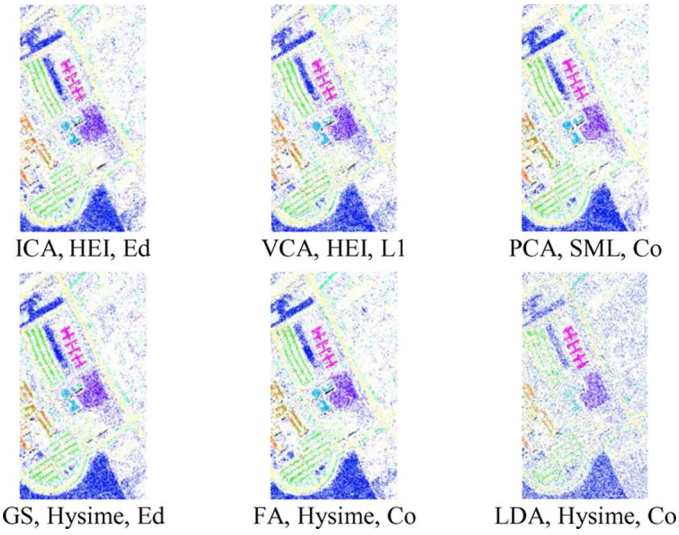

Fig. 5. Best classified map for each DR method for data set 3.

The performance of supervised classification by using features extracted from hyperspectral images (data set 1) with a variety of metric distances is presented in Table IV. It is obvious that the best classification performance is obtained by the ICA method on data set 1. In addition, the Euclidean and norm L1 distance metrics perform more favorably in classifying data set 1 using most IDE and DR methods. Fig. 3 shows the best extracted classified map for each DR method by considering the best IDE estimator and the best distance metric. In data set 1, the Hysime method has the best performance in extracting the ID number. On the other hand, the worst IDE technique is CD, the worst DR technique is LDA, and the worst distance metric is the correlation distance in data set 1.

The same computation is done for data set 2; Table V presents the performance of supervised classification on this data set. From Table V, it is clear that the best feature classification performance is obtained by the PCA method for data set 2. Again, the Euclidean and norm L1 distance metrics achieve the best rank in classifying data set 2 with most IDE and DR techniques. Fig. 4 shows the top extracted classified feature map for each of the DR methods by considering the best IDE estimator and the best distance metric. As before, for data set 2, the Hysime method is the best in estimating the ID number. On the other hand, the worst IDE technique is the CD method, the worst DR technique is the LDA method, and the worst distance metric is the correlation distance for data set 2.

Finally, we have the results obtained from data set 3 for the study and comparison of IDE and DR techniques on hyperspectral images. A comparison of the overall accuracy for data set 3 by using features extracted with various distance metrics is presented in Table VI. From this table, the best outcome of classification is obtained by the PCA technique for data set 3. Similar to previous data sets, the Euclidean and norm L1 distance metrics achieve higher performance in classifying this data set using most IDE and DR methods. In Fig. 5, the best extracted classified feature map for each of the DR methods utilizing the best IDE estimator and distance metric is shown. In this data set, the SML method performs best in extracting the ID number, and the worst IDE technique is the CD method, the worst DR technique is the LDA method, and the worst distance metric is the correlation distance.

To summarize the results obtained from Tables IV–VI and investigate the potential of each DR and IDE technique with comparative respect in three data sets, Fig. 6 shows the average overall accuracy value for each DR method. Similar trends of classifying hyperspectral data sets by using IDE techniques are shown in Fig. 6(a). From this figure, it is clear that the best three IDE methods are Hysime, SML, and NWHFC. From Fig. 6(b), it is obvious that the average overall accuracy for component analysis methods (PCA, ICA, and VCA) achieved higher performance in reducing the dimensionality of hyperspectral images than the remaining DR methods in the three utilized data sets. Also, extracted outputs show the similar trend of procedure in these three data sets.

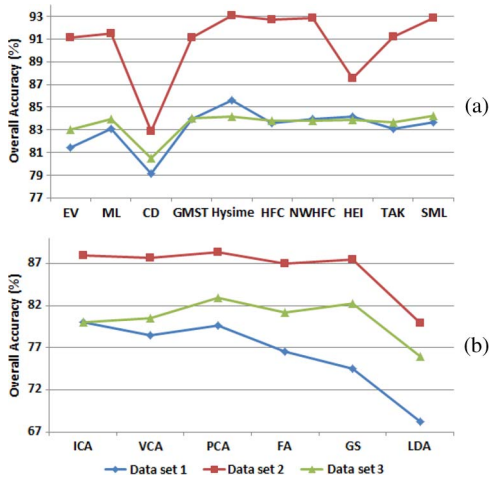

Fig. 6. Comparative output for three data sets. Average overall accuracy in (a) IDE techniques and (b) DR techniques.

V. Conclusion

In this letter, an extensive comparative study of the performance of different IDE methods has been presented. The IDE methods used were the EV estimator, the ML estimator, the CD estimator, the GMST estimator, Hysime, the HFC estimator, the NWHFC estimator, the SML, the HEI ID estimator, and the TAK estimator. The experimental results in this letter clearly show that the Hysime estimator has superior performance in comparison to other IDE techniques. Moreover, different feature extraction techniques (PCA, VCA, ICA, GS, FA, and LDA) have been utilized in a hyperspectral dimension reduction process. As shown in this letter, using the PCA and ICA techniques for feature extraction has better performance in supervised image classification (higher than 80% in the overall accuracy) in the three data sets. Also, as part of this supervised classification process (K-NN classifier), different distance metrics (Euclidean, norm L1, spectral angle, and correlation) have been compared. As can be seen in the results, the norm L1 and Euclidean metrics have higher performance in hyperspectral image classification. Looking more closely to the trend of estimating ID numbers and extracting features by using the ID number in these three data sets (Tables IV–VI), it is observed that our implementation has good robustness in estimating the ID number for the purpose of classifying hyperspectral images. However, this topic still needs more investigation in the area of the automatic determination of optimum IDEs, in selecting suitable bands for classifying hyperspectral images regarding the ID number, and in extracting appropriate features for cube bands of hyperspectral images; also, this exercise should be repeated using a classifier that uses parametric techniques, for example, the support vector machine which is more compatible and feasible to hyperspectral date processing, particularly for feature reduction in hyperspectral images.

References

[1] C. I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*, 1st ed. New York: Springer-Verlag, 2003.
[2] P. F. Hsieh and D. Landgrebe, "Classification of high dimensional data," School ECE, Atlanta, GA, Tech. Rep. 5, 1998.
[3] D. W. Scott, *Multivariate Density Estimation*. Hoboken, NJ: Wiley, 1992.
[4] F. Keinosuke, "Intrinsic dimensionality extraction," in *Classification Pattern Recognition and Reduction of Dimensionality*. Amsterdam, The Netherlands: Elsevier, 1982, pp. 347–360.
[5] C. I. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 608–619, Mar. 2004.
[6] P. Bajorski, "Second moment linear dimensionality as an alternative to virtual dimensionality," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 672–678, Feb. 2011.
[7] T. M. Mitchell, *Machine Learning*, 1st ed. New York: McGraw-Hill, 1997.
[8] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.
[9] C. I. Chang, W. Xiong, W. Liu, M. L. Chang, C. C. Wu, and C. C. C. Chen, "Linear spectral mixture analysis based approaches to estimation of virtual dimensionality in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 3960–3979, Nov. 2010.
[10] O. Eches, N. Dobigeon, and J. Y. Tourneret, "Estimating the number of endmembers in hyperspectral images using the normal compositional model and a hierarchical Bayesian algorithm," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 3, pp. 582–591, Jun. 2010.
[11] N. Acito, M. Diani, and G. Corsini, "Hyperspectral signal subspace identification in the presence of rare signal components," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 4, pp. 1940–1954, Apr. 2010.
[12] L. Van der Maaten, "An introduction to dimensionality reduction using Matlab," Maastricht Univ., Maastricht, The Netherlands, Tech. Rep. MICC 07-07, 2007, vol. 1201.
[13] P. Grassberger and I. Procaccia, "Measuring the strangeness of strange attractors," *Nonlinear Phenom.*, vol. 9, no. 1/2, pp. 189–208, Oct. 1983.
[14] E. Levina and P. J. Bickel, "Maximum likelihood estimation of intrinsic dimension," in *Proc. NIPS*, 2005, vol. 48109, p. 1092.
[15] F. Takens, "On the numerical determination of the dimension of an attractor," in *Proc. Dyn. Syst. Bifurcations*, 1985, pp. 99–106.
[16] K. Fukunaga and D. R. Olsen, "An algorithm for finding intrinsic dimensionality of data," *IEEE Trans. Comput.*, vol. C-100, no. 2, pp. 176–183, Feb. 1971.
[17] M. Hein, J. Y. Audibert, and U. Von Luxburg, "Intrinsic dimensionality estimation of submanifolds in Euclidean space," in *Proc. 22nd ICML*, 2005, pp. 289–296.
[18] P. H. Hsu, Y. H. Tseng, and P. Gong, "Dimension reduction of hyperspectral images for classification applications," *Ann. GIS*, vol. 8, no. 1, pp. 1–8, Jun. 2002.
[19] G. F. Hughes, "On the mean accuracy of statistical pattern recognition," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 1, pp. 55–63, Jan. 1968.
[20] M. D. Farrell and R. M. Mersereau, "On the impact of PCA dimension reduction for hyperspectral detection of difficult targets," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 2, pp. 192–195, Apr. 2005.
[21] J. M. P. Nascimento and J. M. B. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.
[22] J. Wang and C. I. Chang, "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1586–1600, Jun. 2006.
[23] I. K. Fodor, "A survey of dimension reduction techniques," US DOE Office Scientific Technical Information, Livermore, CA, 2002, vol. 18.
[24] O. Kuybeda, A. Kagan, and Y. Lumer, "Determining hyperspectral data-intrinsic dimensionality via a modified Gram–Schmidt process," in *Proc. Elect. Electron. Eng. Israel*, 2004, pp. 380–383.
[25] 2012. [Online]. Available: http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes
[26] A. Dadon, E. Ben-Dor, and A. Karnieli, "Use of derivative calculations and minimum noise fraction transform for detecting and correcting the spectral curvature effect (smile) in Hyperion images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 6, pp. 2603–2612, Jun. 2010.
[27] C. Liu, P. Frazier, and L. Kumar, "Comparative assessment of the measures of thematic classification accuracy," *Remote Sens. Environ.*, vol. 107, no. 4, pp. 606–616, Apr. 2007.