

Systematic Comparison of Linear Feature Extraction Methods for Classification of Hyperspectral Images with Noises

Farid Muhammad Imran, Mingyi He and Yifan Zhang

Shaanxi Key Laboratory of Information Acquisition and Processing, Center for Earth Observation, School of Electronics and Information, Northwestern Polytechnical University, Xi'an, 710129, China
imran@mail.au.edu.pk, myhe@nwpu.edu.cn, yifanzhang@nwpu.edu.cn

Abstract

Hyperspectral Image processing is usually time consuming, due to its huge data size. Nowadays Hyperspectral Imaging is used in many fields where real-time solutions are required. A systemic comparison study of linear feature extraction methods for classification of hyperspectral images with various types of noises is carried out in this paper, in which the performance of different linear feature extraction methods for classification and their computation cost reduction are compared. In practice, hyperspectral images are often contaminated by different types of noises, as the atmosphere around hyperspectral cameras may change all the time. In this paper, to make it more realistic, different types of noises, including Salt-and-Pepper noise, Gaussian noise, Speckle noise and their mixtures, are artificially imposed on the hyperspectral image. Support Vector Machine based classification is employed for classification performance comparison. The experimental results are very helpful for selecting linear feature extraction methods for classification of hyperspectral images that are usually affected with noises.

Keywords: *Classification, feature extraction, hyperspectral images, support vector machine, noise*

1. Introduction

The applications of hyperspectral images (HSI) are no more limited to remote sensing and space research. Nowadays it has been widely used in material identification, crime scene investigation, high throughput in-line inspection, optical biopsy, drug spectroscopy, border protection, Unmanned Aerial Vehicles (UAV) surveillance, mineral mapping, LCD quality control, document inspection, semiconductor operations and many other fields. For most of the applications, real-time solutions are needed, which should be fast and reliable. HSI has a huge data set and sometimes huge data set will reduce the effectiveness of the data mining. Some attributes in data may not contribute to a meaningful model. The irrelevant attributes often adds noises and processing time, which affects the model accuracy and real-time processing performance. Some attributes in data may represent the same feature, which adds skewness in the logic of algorithm and results in affecting the model accuracy as well. Apart from this, higher the dimensionality of processing space, higher the computational cost. Sometimes to minimize the effects of noise, correlation and high dimensionality, dimension reduction methods including linear and nonlinear ones need to be conducted as a preprocessing step for hyperspectral data mining [1-3].

Feature Extraction (FE) and dimensionality reduction are important tasks in many fields where signal processing and analysis is involved [4]. FE is performed by combining bands into a smaller number of new features. It provides the compressed version of the original data set. Each band in the original data set has contribution determined by the

transformation matrix associated with a given extraction method. As a result, dimensionality is reduced, accuracy is increased and solutions are more robust [5].

HSI are often disturbed by radiometric noise such as sensor noise, photon (or shot) noise, calibration error, atmospheric scattering and absorption [6], which not only degrades the visual quality of the HSI data but also limits the precision of the subsequent image interpretation and analysis [7]. The noise in HSI can be divided into two categories: random noise and fixed-pattern noise [8]. This paper will focus on random noise. In real-time situations different kinds of noises can be observed in HSI as the atmosphere around the HSI cameras may change all the time. Therefore, to make it more realistic, different types of noises, including Salt-and-Pepper (S&P) noise, Gaussian noise, Speckle noise and their mixtures, are artificially imposed on the hyperspectral image. In this paper, Support Vector Machine (SVM) [9] based classification, a widely used classifier for HSI, is employed for performance comparison.

Although various linear and nonlinear feature extraction methods have been proposed and used for classification of hyperspectral images, systemic study on the performance of various feature extraction methods for practical hyperspectral images are rare.

In this paper, for the lower computational cost, reliability and robustness, linear feature extraction (LFE) methods for classification of hyperspectral image with noises are systematically compared. The effect of different types of noises on the classification accuracy and computational cost is studied as well.

2. Selected Linear Feature Extraction Methods

In this paper we have studied the well-known linear feature extraction techniques, such as Principal Component Analysis (PCA), Partial Least Square- SB (PLS-SB), Partial Least Square (PLS), Orthogonal Partial Least Square (OPLS), Canonical Correlation Analysis (CCA) and Maximum Noise Fraction (MNF). PCA [10] is an unsupervised feature extraction method as it depends only on the input. It transforms data into new coordinates such that the greatest variance in data lies on the first coordinate and the second greatest variance on the second and so on. PLS [11] comes in the families of bilinear factor models as it uses both X 's (the predictor variables) and Y (the dependent variable) data to project it into new space. PLS finds the linear regression model by projecting the predicted variables and the observed variables into the new space. There are many other variants of PLS [12], such as OPLS [13]. OPLS removes the unwanted variation of X 's which is not correlated with Y , while large data

Table 1. Summary of Linear Feature Extraction Methods

PCA	PLS	OPLS	CCA
$u^T C_x u$	$u^T C_{xy} v$	$u^T C_{xy} C_{xy}^T u$	$u^T C_{xy} v$
$U^T U = I$	$U^T U = I$ $V^T V = I$	$U^T C_x U = I$	$U^T C_x U = I$ $V^T C_y V = I$
$r(X)$	$r(X)$	$r(C_{xy})$	$r(C_{xy})$

blocks are integrated. CCA [14] maximize the correlation between the projected input and the output data. MNF [15] is basically the same as PCA. The main difference is how it normalizes the raw input spectra for noise before the PCA.

If X and Y are taken as the column wise-centered input and target data matrices then the input data sample covariance matrix is given as $C_x = (1/L)X^T X$ and the output sample covariance matrix is given as $C_y = (1/L)Y^T Y$. The covariance between input and output

data is $C_{xy} = (1/L)X^T Y$; where L is the number of training samples. The transformation is obtained via a linear transformation of the original data, $X' = XU$, where $U = [u_1, u_2, \dots, u_{n_F}]$ is the projection matrix, u_i being the i th projection vector and n_F is the number of extracted features [4]. Some methods also considered the feature extracted in the output space, so $Y' = YV$, with $V = [v_1, v_2, \dots, v_{n_F}]$. Table 1 describes the summary of the linear feature extraction methods used in the experiments. The objective is to maximize the first row with constraint to the second row; the last row represents the maximum number of features. Vector u is the column vectors in matrix U and $r(\cdot)$ is the rank of a matrix [4].

3. Image Noises

In the experiments, three different types of noises named: Salt-and-Pepper (S&P), Gaussian (Gau) and Speckle (Spe) are studied.

3.1. Salt-and-Pepper Noise

It is called impulsive, spot or peak noise. It appears randomly as white or black spot in image.

$$I_{sp}(i, j) = \begin{cases} I(i, j) & x < l \\ I_{\min} + y(I_{\max} - I_{\min}) & x \geq l \end{cases} \quad (1)$$

$x, y \in [0, 1]$ are two uniformly distributed random variables.

3.2. Gaussian Noise

Gaussian noise is a statistical noise having probability density function equals to the normal distribution. Gaussian $n(i, j)$ is a random variable with distribution

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2)$$

where μ is the mean and σ is the variance.

3.3. Speckle Noise

Speckle is the granular noise that inherently exists to degrade the image. It is also called the multiplicative noise.

$$g(i, j) = f(i, j) * u(i, j) + \xi(i, j) \quad (3)$$

where u is the randomly distributed noise with zero mean and varying variance.

In the experiments, influences of noises with varying variances on classification accuracy are illustrated. The types of noises and their mixtures studied in the experiments are shown in Table 2, with SNR of 15.7dB, 5.7dB and -4.3dB. SNR is calculated as

$$SNR = 10\log_{10}(\text{var}(\text{image}) / \text{var}(\text{noise})) \quad (4)$$

Table 2. Noises and Mix-noises Used for Analysis

Noise
S&P, Gau, Spe, S&P+Gau, S&P+Spe, Gau+Spe, S&P+Gau+Spe

4. Experimental Results

The classical 220 bands hyperspectral Image of Indian Pine Scene generated by the AVIRIS instrument in 1992 is used for the test. It consists of 145×145 pixels and 16 ground-truth classes, ranging from 20 to 2468 pixels in size [16]. We remove 20 noisy bands covering the region of water absorption, and work with 200 spectral bands. We are using all 16 classes and 50% pixels from each class are randomly chosen to make training set. Average results are taken after repeating each experiment 5 times. For our experimentation we use PC with specifications: Processor Intel(R) Core(TM) i5-2400 CPU @ 3.10GHz, 4 Core(s) with 4.0 GB RAM. The criteria used to compare classification results involve Overall Accuracy (OA), Time (s) and the kappa coefficient (k). The time mentioned in our experiment is the time for the SVM to build model and to predict the final Overall Accuracy (OA).

Table 3. SVM OA, Time and Kappa Coefficient with and without Preprocessed Data by Different LFE Methods when “No Noise” is used on Indian Pines HSI

	No FE	PCA	PLS-SB	PLS	OPLS	CCA	MNF
OA (%)	90.61	83.71	87.63	88.12	72.11	71.08	78.82
Time (s)	28.13	2.91	2.50	2.58	1.98	1.99	1.74
Kappa	0.893	0.814	0.859	0.865	0.673	0.659	0.756

Table 3 represents the case when no noise is imposed on HSI. We get best OA when no FE technique is used compare to other FE techniques. Table 4 summarizes the case when we use the SNR of 5.7dB for different noises. Compared with the result without Feature Extraction (No FE), there is obvious improvement in overall accuracy when PCA, PLS-SB and PLS LFE methods are used. The difference is more prominent when the noises get severe. The case where all three noises are used together i.e. S&P+Gau+Spe with PLS method, there is a 20% improvement in accuracy and the processing time is 33 times faster. Overall PLS and PLS-SB methods perform better

Table 4. SVM OA, Time and Kappa Coefficient with and without Preprocessed Data by Different LFE Methods when “SNR of 5.7dB” is used on HIS

		No FE	PCA	PLS-SB	PLS	OPLS	CCA	MNF
S&P	OA (%)	76.66	78.18	80.30	81.10	52.74	53.17	65.44
	Time (s)	47.76	3.06	2.70	2.71	2.35	2.38	1.86
	Kappa	0.731	0.750	0.775	0.784	0.426	0.431	0.593
Gau	OA (%)	62.84	70.27	71.89	71.79	48.80	47.93	57.91
	Time (s)	83.41	3.52	3.00	3.01	2.45	2.51	1.96
	Kappa	0.559	0.658	0.677	0.676	0.375	0.362	0.497
Spe	OA (%)	75.55	77.54	79.49	79.92	51.98	52.58	66.33
	Time (s)	42.58	3.06	2.69	2.71	2.35	2.38	1.86
	Kappa	0.718	0.743	0.765	0.770	0.417	0.424	0.604

S&P+ Gau	OA (%)	53.15	68.52	68.46	69.00	47.50	47.81	55.01
	Time (s)	100.37	3.67	3.25	3.20	2.50	2.52	1.99
	Kappa	0.429	0.638	0.637	0.644	0.357	0.360	0.457
S&P+ Spe	OA (%)	67.96	73.17	75.43	74.89	49.37	50.09	60.15
	Time (s)	59.71	3.30	2.87	2.88	2.44	2.46	1.89
	Kappa	0.625	0.692	0.718	0.712	0.382	0.391	0.526
Gau+ Spe	OA (%)	54.61	67.80	68.84	68.76	47.13	47.66	55.34
	Time (s)	95.13	3.67	3.18	3.18	2.50	2.52	2.01
	Kappa	0.449	0.629	0.641	0.641	0.351	0.358	0.462
S&P+ Gau+ Spe	OA (%)	46.47	65.78	65.00	66.69	46.43	46.62	54.53
	Time (s)	111.29	3.85	3.32	3.36	2.51	2.57	2.07
	Kappa	0.335	0.605	0.608	0.617	0.341	0.342	0.451

Table 5. SVM OA, Time and Kappa Coefficient with and without Preprocessed Data by Different LFE Methods when “SNR of -4.3dB” is used on HIS

		No FE	PCA	PLS-SB	PLS	OPLS	CCA	MNF
S&P	OA (%)	23.80	58.26	57.24	56.96	41.81	42.26	50.72
	Time (s)	259.11	5.34	4.50	4.45	2.61	2.66	2.38
	Kappa	0.000	0.515	0.504	0.502	0.275	0.280	0.400
Gau	OA (%)	23.80	50.47	49.37	49.98	38.95	39.32	46.02
	Time (s)	275.86	6.90	5.86	5.70	2.69	2.76	2.72
	Kappa	0.000	0.422	0.412	0.420	0.233	0.238	0.338
Spe	OA (%)	41.80	60.08	59.48	59.10	43.71	43.79	49.35
	Time (s)	105.60	5.21	4.34	4.318	2.65	2.70	2.50
	Kappa	0.264	0.535	0.529	0.526	0.299	0.301	0.380
S&P+ Gau	OA (%)	23.80	46.43	44.71	45.07	37.22	36.63	44.68
	Time (s)	308.19	7.75	6.43	6.39	2.79	2.84	2.93
	Kappa	0.000	0.372	0.356	0.360	0.207	0.198	0.320
S&P+ Spe	OA (%)	23.80	52.61	50.91	50.58	38.44	38.55	46.89
	Time (s)	259.01	6.71	5.65	5.59	2.81	2.80	2.77
	Kappa	0.000	0.445	0.428	0.425	0.224	0.225	0.349
Gau+ Spe	OA (%)	23.80	47.51	45.78	46.16	36.90	37.12	44.96
	Time (s)	276.40	7.50	6.30	6.12	2.77	2.82	2.92
	Kappa	0.000	0.383	0.367	0.372	0.202	0.205	0.323
S&P+ Gau+ Spe	OA (%)	23.80	43.90	42.35	42.14	36.41	36.27	43.60
	Time (s)	307.94	8.33	6.71	6.79	2.83	2.92	3.10
	Kappa	0.000	0.339	0.326	0.323	0.195	0.193	0.305

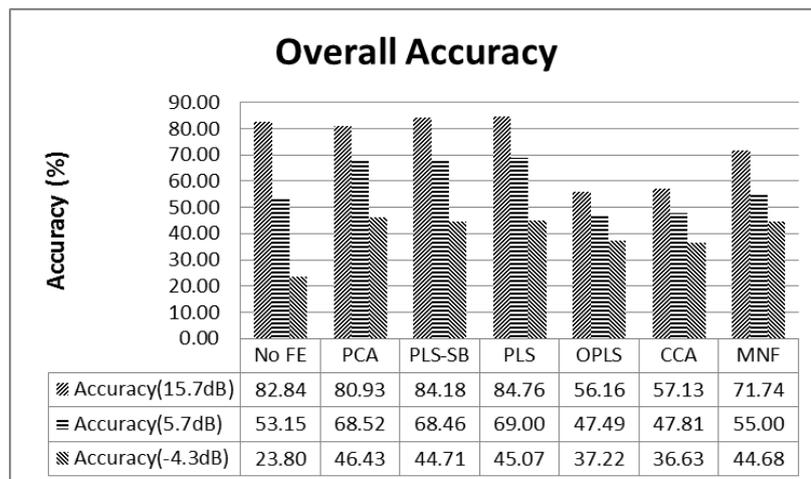


Figure 1. SVM Overall Accuracy when Mix-noise Salt-and-Pepper + Gaussian is Artificially Imposed in HIS

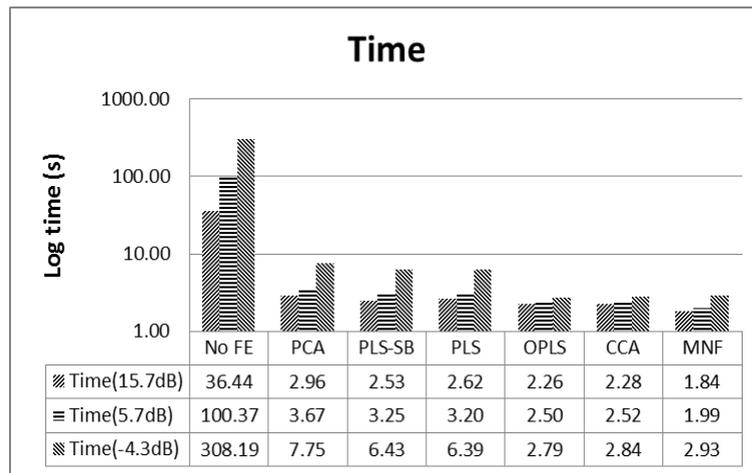


Figure 2. SVM Processing Time when Mix-noise Salt-and-Pepper + Gaussian is Artificially Imposed in HSI

than the other LFE methods in terms of overall accuracy because they are bilinear factor models. MNF method performs the best in terms of time, as it pre-processes the noisy data. Table 5 summarizes the results when we use the SNR of -4.3dB for different noises. The case when no FE is used, SVM saturates at the accuracy of 23.80% and the kappa coefficient becomes zero mostly. We get good improvement in accuracy and tremendous reduction in time when LFE are used. PCA performs best in terms of accuracy, as it is variance dependent and OPLS and MNF perform best in terms of time.

In Figure 1, mix-noise Salt & Pepper + Gaussian is used, with different SNR of 15.7dB, 5.7dB and -4.3dB. Using SNR of 15.7dB we get an OA of 82.84% when no FE is used. Compare to this when PCA, PLS-SB and PLS LFE techniques are used we get an OA of 80.93%, 84.18% and 84.76% respectively. PLS performs the best compare to other LFE

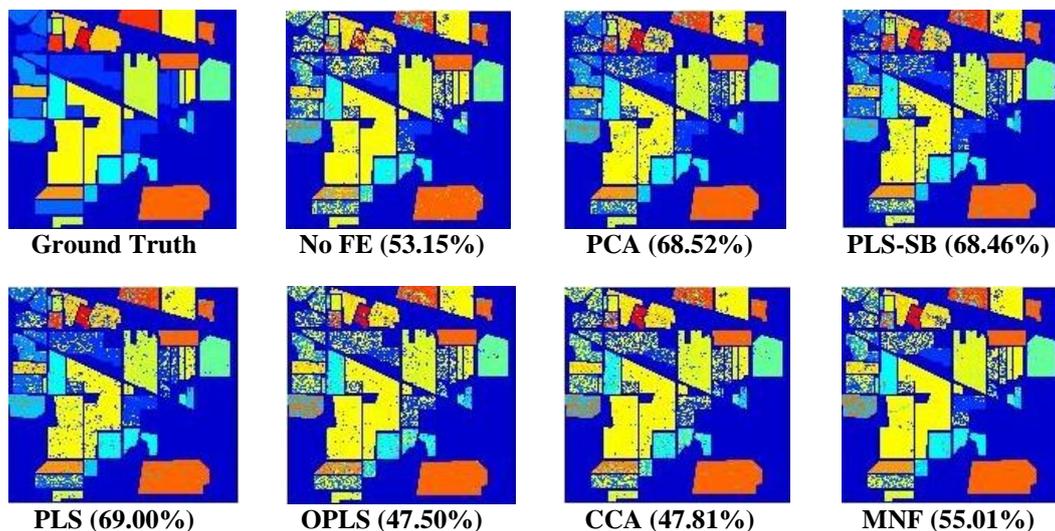


Figure 3. Classification maps using different linear feature extraction techniques when mix-noise Salt-and-Pepper + Gaussian is used with the "SNR of 5.7dB".

techniques when SNR of 15.7dB is used. When SNR of 5.7dB is used, the OA when No FE is used drops to 53.15%. PLS performs the best and PCA performs the second best. When SNR is drops to -4.3dB we get the OA of 23.80% when No FE is used and get the best OA of 46.43% when PCA is used for LFE. It can be observed that the lower the SNR, the better results acquired by using LFE methods compare to the No FE case. From Figure 2 we can see that there is a tremendous reduction in time when the selected LFE techniques are used as compare to the case when No FE is used. MNF performs the best. Time is approximately 40 times faster when LFE methods are used.

Figure 3, shows the Classification maps using different linear feature extraction techniques when mix-noise Salt-and-Pepper + Gaussian is used with the SNR of 5.7dB. Comparing the top left side of the classification maps of PCA, PLS-SB and PLS with No FE classification map we can see clearly that classification maps improves when LFE methods are used.

5. Conclusion

In this paper, the performance of different linear feature extraction techniques are compared considering classification accuracy and computation cost. HSI is contaminated by different types of noises and then classification is performed by using SVM. The experimental results have shown that FEs before classification with PLS and PLS-SB perform the best no matter what the SNR is chosen. Additionally, when SNR is very low i.e. in minus, FEs before classification with PCA performs the best. The processing time for all FEs studied is much shorter than that of without FEs, and the speed with MNF is the fastest. It is also observed that classification maps are improved when LFE methods are used on noisy HSI. FE is no doubt the critical step if HSI is used in real-time applications as all sorts of noises may emerge. FE generally compresses the original data set into a few new features, which results in improvement in classification accuracy and the processing speed.

Acknowledgements

This research is supported by National Natural Science Foundation of China (key project number 61420106007 and project number 61101188)

References

- [1] X. Jia, K. Bor-Chen, and M. M. Crawford, "Feature Mining for Hyperspectral Image Classification," *Proceedings of the IEEE*, (2013), vol. 101, no. 3, pp. 676-697.
- [2] X. Jia and M. He, "Feature Mining for Hyperspectral Data," Invited Tutorial at IEEE GRSS 4th Workshop on Hyperspectral Image and Signal Processing, (2012), Shanghai, China.
- [3] M. He, "Signal and Information Processing in Hyperspectral Imaging Remote Sensing," Invited Distinguished Lecture IEEE ICIEA (2014), Hangzhou, China.
- [4] J. Arenas-Garcia, K. Petersen, G. Camps-Valls, and L. K. Hansen, "Kernel Multivariate Analysis Framework for Supervised Subspace Learning: A Tutorial on Linear and Kernel Multivariate Methods," *IEEE Signal Processing Magazine*, (2013), vol. 30, no. 4, pp.16-29.
- [5] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*, Wiley, Hoboken, NJ, (2003).
- [6] J. P. Kerekes, and J. E. Baum, "Full-spectrum spectral imaging system analytical model," *IEEE Trans. Geosci. Remote Sens.*, (2015), vol. 43, no. 3, pp. 571-580.
- [7] S. Matteoli, N. Acito, M. Diani, and G. Corsini, "An automatic approach to adaptive local background estimation and suppression in hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, (2011), vol. 49, no. 2, pp. 790-800.
- [8] X. Guo, X. Huang, L. Zhang and Lefei Zhang, "Hyperspectral image noise reduction based on rank-1 tensor decomposition", *ISPRS Journal of Photogrammetry and Remote Sensing*, (2013), vol. 83, pp. 50-63.
- [9] V. Vapnik, "The Nature of Statistical Learning Theory", Springer, (1999).
- [10] I. T. Jolliffe, "Principal Component Analysis", Springer Verlag, New York, (1986).
- [11] H. Wold, "Non-linear estimation by iterative least squares procedures," in *Research Papers in Statistics*, F. David, Ed. Wiley, New York, (1966), pp. 411-444.

- [12] P. Geladi, "Notes on the history and nature of partial least squares (PLS) modeling," *J. Chemometr.*, **(1988)**, vol. 2, no. 4, pp. 231-246.
- [13] J. Trygg, and S. Wold, "Orthogonal projections to latent structures (O-PLS)," *J. Chemometr.*, **(2002)**, vol.16, pp. 119-128.
- [14] H. Hotelling, "Relations between two sets of variates," *Biometrika*, **(1936)**, vol. 28, no. 3-4, pp. 321-377.
- [15] A. A. Green, M. Berman, etc., "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sens.*, **(1988)**, vol. 26, no. 1, pp. 65-74.
- [16] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, **(2009)**, vol. 113, pp. 110-122.

Authors



Muhammad Imran Farid, he received his BSc degree in both mathematics and physics from University of the Punjab, Pakistan, in 1998 and the M.Sc. degree in electronics from Quaid-i-Azam University, Islamabad, Pakistan, in 2000. He also received his MSc degree in radio frequency communication systems from University of Southampton, U.K, in 2005. He is currently pursuing the Ph.D. degree in image processing at Northwestern Polytechnical University, Xian, China. He is currently working as a Lecturer in Air University, Islamabad, Pakistan and is on study leave for PhD. His research interests include remote sensing, pattern recognition and astronomy.



Mingyi He received B.Eng. and M.Eng. from Northwestern Polytechnical University (NPU) in 1982 and 1985, respectively, and Ph.D. from Xidian University, China, in 1994. Currently Prof. He is working with the School of Electronics and Information, NPU. He is the Founder and Director of Shaanxi Key Laboratory of Information Acquisition and Processing and the Director and Chief Scientist of the Center for Earth Observation Research. His research interests include hyperspectral image processing, computer vision and image processing, neural networks and intelligent information processing with notable applications to X-ray image processing for luggage inspection, laser-finder test system for airborne systems etc.