Hyperspectral Image Classification Using Factor Analysis and Convolutional Neural Networks

A. F. M. Minhazur Rahman and Boshir Ahmed

Abstract Hyperspectral image sensors can provide valuable data for land covers, oceans, and the earth atmosphere at various spatial and spectral scales. Rich spectral and spatial information of a location makes hyperspectral image (HSI) an excellent way to work with materials, identify them, or define their properties. However, computer-automated analysis and classification of hyperspectral image is a challenging problem. Most of the spectral information in hyperspectral image is correlated, containing redundant information. High number of bands in input image contributes to the curse of dimensionality problem that reduces classifier performance. In many applications, the amount of labelled hyperspectral data that can be acquired is minimal. The complexities associated with HSI motivate us to propose a method named FA-CNN. We have used factor analysis (FA) dimension reduction technique to remove band correlation while maintaining useful spectral information in a lower number of bands. Then, we have applied convolutional neural network (CNN) for combining spectral and spatial features of HSI. Finally, multilayer perceptron classifier is used for classifying each of the input pixels in HSI. Our proposed method achieved 99.59% overall accuracy and 99.75% average accuracy on Indian Pines dataset; 99.95% overall accuracy and 99.90% average accuracy on Pavia University dataset while requiring a lower number of trainable parameters and training data compared to other methods.

Keywords Hyperspectral image classification · Dimension reduction · Factor analysis · Spectral and spatial feature extraction · Convolutional neural networks

1 Introduction

Hyperspectral remote sensing aims to collect spectral and spatial data from objects on the earth's surface based on their reflectance property acquired by airborne or spaceborne sensors. It has finer wavelength resolution, contiguous wavelength bands, and

A. F. M. Minhazur Rahman (✉) · B. Ahmed
Department of Computer Science and Engineering, Rajshahi University of Engineering & Technology, Rajshahi, Bangladesh

M. S. Arefin et al. (eds.), *Proceedings of the International Conference on Big Data, IoT, and Machine Learning*, Lecture Notes on Data Engineering and Communications Technologies 95, https://doi.org/10.1007/978-981-16-6636-0_11

a higher range than traditional image acquisition techniques. It makes precise analysis of soils, waterbody, and materials possible [1]. Because of the large amount of detailed information available, hyperspectral image classification of ground objects has become a common research topic among researchers. However, high dimensionality [2] of hyperspectral images combined with limited training samples leads to Hughes phenomenon [3]. Hughes phenomenon has the potential to significantly reduce classifier performance. To overcome the problem of high dimensionality, various dimension reduction techniques like PCA [4], LDA, etc. are commonly used. Spectral features extracted from these dimension reduction techniques are insufficient for providing satisfactory classification performance using SVM [5], Multilayer Perceptron (MLP) etc.

Deep learning-based frameworks have become a popular method for hyperspectral image classification task. Support vector machine, random forest, decision tree, and other conventional machine learning approaches require laborious feature engineering. In comparison, deep learning-based methods automatically extract features that are effective in minimizing task errors [6]. On the other hand, combining spatial context with spectral features has shown great promise [7]. Rather than treating hyperspectral image cubes as a simple pixel array, designing a classifier that integrates spectral and spatial features is an efficient way to enhance classification efficiency. Spatial features provide additional information correlated with the shape and structure of a material [8]. The rise of deep learning techniques and the effectiveness of spectral–spatial classification framework have led to broader adoption of convolutional neural networks (CNN) [9–12]. Major drawbacks of CNN include increased training time, large training data, and high computational power requirements compared to traditional machine learning techniques. However, hyperspectral images typically do not have a large amount available training samples [10]. This constraint forces the development of a robust CNN framework that can capture the underlying discriminant features of the HSI using a minimal amount of training samples.

One of the recent CNN-based method—Makantasis et al. [12] used Randomized-PCA to reduce dimension and represent spectral information in a compressed form while keeping spatial information intact. This dimension reduced image was fed into convolutional neural network (CNN) using 2D convolution, which conducts high-level feature construction and a multilayer perceptron, which classifies each pixel of the input image. This method performed extremely well using 30 PCA bands. Haque et al. [11] used a similar method to [12]. However, they considered multiple spatial contexts like 3×3, 5×5 and 7×7 at the same time to extract different level of spatial features and passed it to CNN using 2D convolution, 2D max-pooling, and MLP classifier. This method performed well but increased the number of trainable parameters compared to other methods. Roy et al. [13] proposed a joint 2D and 3D-CNN-based method called HybridSN. This method utilizes 3D-CNN for spectral–spectral feature learning from dimensional reduced spectral bands. 2D-CNN follows the 3D-CNN framework. 2D-CNN learns deep spatial features, which further ensures proper utilization of available spatial information. Usage of both 2D and 3D-CNN and a MLP classifier resulted in high accuracy in HSI classification task. One downside of this method is the added complexity of joint 3D and 2D convolution and high number of trainable parameters.

In this research work, we have proposed a method called FA-CNN, which uses factor analysis dimension reduction technique to overcome the Hughes phenomenon by finding the original image bands' underlying factors and representing spectral information of the original image using those factors. After that, we have used a convolutional neural network to combine spectral and spatial information of the image in a single step and a multilayer perceptron classifier to classify each pixel in the input hyperspectral image. Our proposed method overcomes some of the drawbacks in the hyperspectral image classification mentioned above.

2 Methodology

Our research methodology is comprised of three major steps. In the first step, the input hyperspectral image cube is projected into a new subspace using factor analysis dimension reduction technique. After that, multiple overlapping three-dimensional patches are created from the dimensionally reduced image for providing spatial–spectral information to the CNN framework. Finally, fully connected layers in the CNN are used to classify each input pixel fed into our proposed framework. The following subsections elaborate on these main components of our research methodology.

2.1 Dataset Description

Indian Pines (IP) hyperspectral dataset (Fig. 1) was captured using AVIRIS sensor in Northwest Indiana, USA. There are 200 spectral bands in this dataset. The spatial dimension of the hyperspectral image cube is 145×145. After removing background samples, 10249 samples from 16 classes are left for experimentation.

Pavia University (PU) hyperspectral dataset (Fig. 2) was captured using ROSIS sensor in Northern Italy. This dataset contains 103 spectral bands. The spatial dimension of the hyperspectral image cube is 610×340. After removing background samples, 42776 samples from 9 classes are left for experimentation.

2.2 Dimension Reduction Using Factor Analysis

Factor analysis is an unsupervised feature extraction method that finds the underlying factors that explain the original spectral information of hyperspectral data in fewer dimensions while keeping the spatial information intact. This dimension reduction

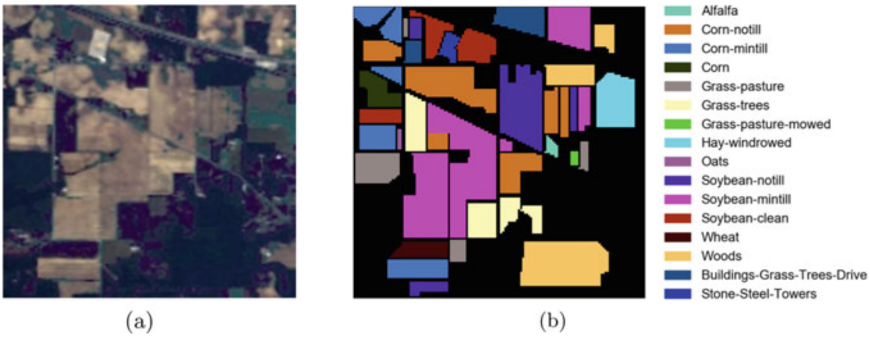


Fig. 1 Indian Pines dataset. **a** False colour composite image; **b** ground truth

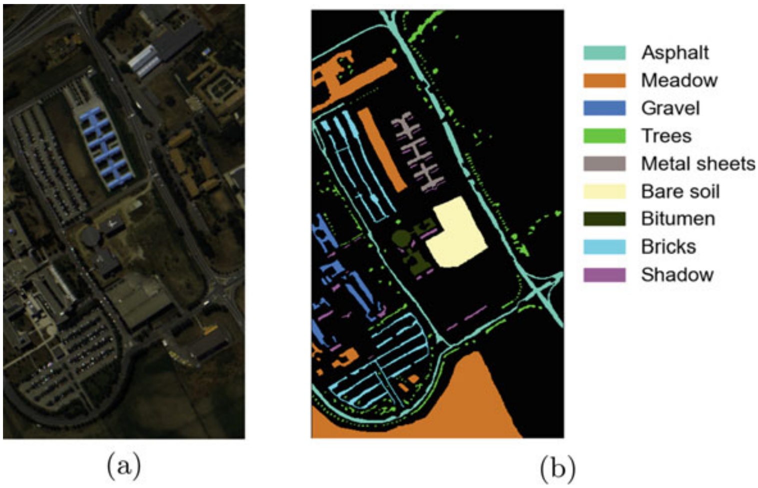


Fig. 2 Pavia University dataset. **a** False colour composite image; **b** ground truth

process alleviates the curse of dimensionality problem, reduces training parameters, and compresses the original data for memory efficiency.

Factor analysis assumes that all p observed variables X_1, X_2, ..., X_p in a dataset is linearly dependent on m unobservable, common factors $F_1, F_2, \ldots, F_m$. Observed variables also have unique variances $\epsilon_1, \epsilon_2, \ldots, \epsilon_p$ [14]. For a single observation,

$$X_i - \mu_i = l_{i1}F_1 + l_{i2}F_2 + \cdots + l_{im}F_m + \epsilon_i \tag{1}$$

l_{ij} is loading or weight of the ith variable on the jth factor. μ_i is the mean of X_i. For n observation, the matrix form of factor model,

$$X - \mu = LF + E \tag{2}$$

A maximum likelihood estimation technique will find factors that increase the likelihood of generating the covariances matrix of the original data. Assumption is made that the data are independently sampled from a multivariate normal distribution with mean vector μ, and covariance matrix of the form $LL^T + \psi$, where ψ is the covariance matrix of E [14]. The log likelihood function to find MLE estimators μ, L, ψ is,

$$\mathcal{L}(\mu, L, \psi) = -\frac{np}{2}log2\pi - \frac{n}{2}log|LL^T + \psi| - \frac{1}{2}(X_i - \mu)^T(LL^T + \psi)(X_i - \mu) \tag{3}$$

According to Bartlett method [15], we can obtain the factor score matrix F,

$$F = (L^T\psi^{-1}L)^{-1}L^T\psi^{-1} \tag{4}$$

Finally, projection of X into new subspace is obtained by,

$$Y = F^T(X - \mu) \tag{5}$$

By following the above method, a hyperspectral image cube X with $W \times H \times \lambda$ dimension can be reduced to $W \times H \times B$ where $B <= \lambda$.

2.3 Neighbourhood Patch Creation

The input of the hyperspectral image classification pipeline is normally a single image cube. This differs from conventional classification problems, where many images are used to train the model.

To create multiple image cubes for CNN model, HSI cube Y of dimension $W \times H \times B$ needs to be split into many neighbourhood patches(also called spatial context) of dimension $S \times S \times B$. This can be achieved by considering a neighbourhood window of $S \times S$ size for each pixel in the input HSI (Fig. 3). This procedure creates $(W - S + 1) \times (H - S + 1)$ number of 3D-patches from Y [13]. We added padding of size $\frac{S-1}{2}$ to the original image before creating the patches to ensure that the number of samples remains the same after creating the 3D-patches.

2.4 CNN Architecture

We will use CNN for constructing deep high-level spectral–spatial features. Spectral and spatial feature extraction is performed by using a 2D convolution layer followed by one 2D max-pooling layer and finally another 2D convolution layer. Since input

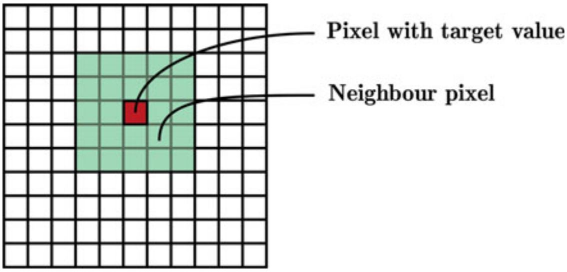


Fig. 3 Creating neighbourhood patch from input image (2D view)

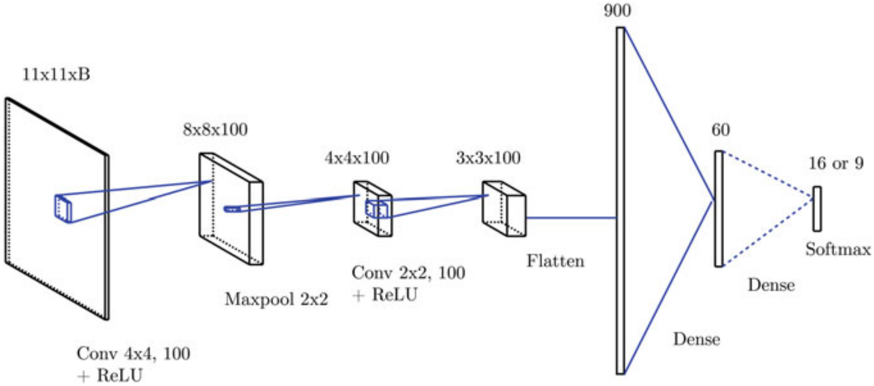


Fig. 4 Proposed CNN architecture

image cubes have extremely low spatial dimension, filter size is restricted to 4×4 (100 filters) and 2×2 (100 filters). The activation function of choice for all convolution layers is ReLU. ReLU has proven to converge faster in HSI classification [10]. Max-pooling layer accounts for spatial invariance. Performing convolution and pooling operation generates invariant, nonlinear, and discriminant features that will facilitate effective classification [9].

Multilayer perceptron with one hidden layer and softmax output layer is used for classification of each pixel in the input image cube. The output layer has 16 units (Indian Pines) or 9 units (Pavia University), each representing the likelihood of an input image cube belonging to a certain class. The choice of loss function for the network is categorical cross-entropy (Fig. 4).

The proposed CNN architecture is shallow because using deep network with small training data available could lead to overfitting problem [10].

3 Experimental Analysis

3.1 Experimental Setup

We have trained our deep learning model using Keras. Keras is a framework for creating deep neural networks that provides an abstraction on top of TensorFlow library.

We used Google Colab Jupyter Notebook powered by Intel(R) Xeon(R) CPU@2.20 GHz, 12 GB RAM, NVIDIA Tesla T4 GPU running on Linux-based operating system for training and testing our architecture.

Indian Pines and the Pavia University datasets are divided into training and test sets. Only 40% of the initial dataset is used as the training sample, while 60% is used as the test set. For hyperparameter tuning, 60% of the training set is used as the validation set. Training on such a small number of samples is critical in developing reliable classification method because hyperspectral images seldom have a large number of training data available for the classification task [10, 16].

The model was trained for 130 epochs using mini batch gradient descent. Selected batch size was 16 and the learning rate was empirically determined to be 0.001.

3.2 Result and Performance Analysis

We considered different numbers of spatial context or neighbourhood window sizes $S = \{7, 9, 11, 13\}$ and extracted bands $B = \{1, 3, 5, 7, 9, 11\}$ for our experiment to find the optimal configuration that would give maximum overall accuracy and average accuracy.

Figure 5 shows that increasing the number of extracted band and spatial context positively affects accuracy. However, increasing the spatial context increases trainable parameters and consequently increases training time. Also, a higher value of S can lead to overlapping problem [17]. So we settled on $S = 11$. A lower number of extracted band does not contain enough discriminating information to classify HSI correctly. Increasing the number of extracted bands up to a certain extent adds useful spectral information for the classifier to make correct decisions without introducing the dimensionality problem. Based on the above results, the optimal value is $S = 11$ and $B = 11$. So we have used, $11 \times 11 \times 11$ neighbourhood patches, constructed from original hyperspectral image cube, as input to our CNN framework for best performance.

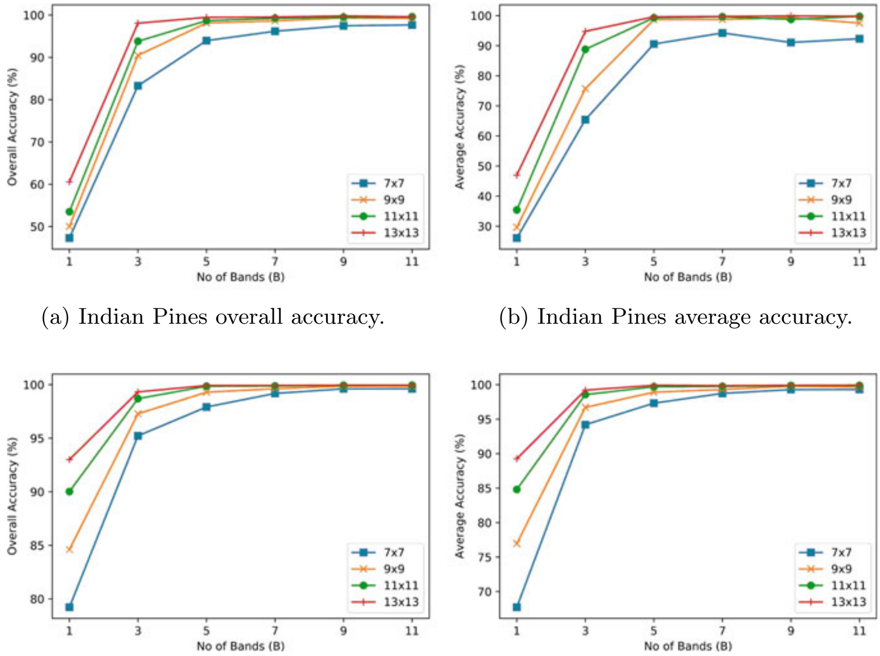
Tables 1 and 2 demonstrate that FA-CNN performs well on overall accuracy, average accuracy, per-class accuracy, Cohen's Kappa(κ), and $F1$ score metrics on both Indian Pines and Pavia University dataset. The classifier only makes 17 misclassification out of 4100 test samples on Indian Pines dataset and 9 misclassification out of 17111 samples on Pavia University dataset.

Table 1 Performance of FA-CNN using $S = 11$ and $B = 11$ on Indian Pines

#	Class	Accuracy (%)	Training samples	Testing samples
1	Alfalfa	100	16	19
2	Corn-notill	99.47	514	571
3	Corn-mintill	99.40	299	332
4	Corn	100	85	95
5	Grass-pasture	98.96	174	193
6	Grass-trees	100	263	292
7	Grass-pasture-mowed	100	10	11
8	Hay-windrowed	100	172	191
9	Oats	100	7	8
10	Soybean-notill	98.97	350	389
11	Soybean-mintill	99.59	884	982
12	Soybean-clean	100	213	237
13	Wheat	100	74	82
14	Woods	99.60	455	506
15	Buildings-Grass-Trees-Drives	100	139	155
16	Stone-Steel-Towers	100	34	37
Average accuracy		99.75		
Overall accuracy		99.59		
κ score		99.53		
$F1$ score		99.66		

Table 2 Performance of FA-CNN using $S = 11$ and $B = 11$ on Pavia University

#	Class	Accuracy (%)	Training samples	Testing samples
1	Asphalt	100	2387	2652
2	Meadows	100	6173	7460
3	Gravel	99.40	756	839
4	Trees	100	1103	1226
5	Metal sheets	100	484	538
6	Bare Soil	100	1810	2012
7	Bitumen	100	479	532
8	Bricks	99.73	1326	1473
9	Shadows	100	341	379
Average accuracy		99.90		
Overall accuracy		99.95		
κ score		99.93		
$F1$ score		99.91		



(a) Indian Pines overall accuracy. (b) Indian Pines average accuracy.

(c) Pavia University overall accuracy. (d) Pavia University average accuracy.

Fig. 5 Accuracy results for different values of extracted bands (B) and spatial context (S)

3.3 Comparison with Other Methods

We compared our proposed FA-CNN with two recent spectral–spatial classification methods—HybridSN [13] and PCA-MS-CNN [11]. For Indian Pines dataset, 30 PCA bands accounting for 99.24% variance of the original dataset were extracted, as suggested by those papers. For Pavia University dataset, 15 PCA bands accounting for 99.90% variance of the original dataset were extracted for both methods. Spatial context for HybridSN and PCA-MS-CNN is, respectively, $S = 25$ and $S = \{3, 5, 7\}$, as recommended by those papers.

We have implemented all of these methods on our machine and used the same training split of 36% from the original dataset to train the models and the same test split for evaluating performance.

From Table 3, we can see that our method FA-CNN compares favourably to HybridSN and PCA-MS-CNN on accuracy metrics while requiring a vastly lower number of trainable parameters. A lower number of trainable parameters indicates that our network can be trained faster compared other two approaches.

Table 3 Comparison of accuracy and trainable parameters between various methods

Dataset	Measurements	Methods		
		Our method	HybridSN	PCA-MS-CNN
IP	Overall accuracy	**99.59**	98.37	97.61
	Average accuracy	**99.75**	96.15	97.79
	Total trainable parameters	**112,836**	5,122,176	10,035,746
PU	Overall accuracy	**99.95**	99.78	99.68
	Average accuracy	**99.90**	99.41	99.40
	Total trainable parameters	**112,409**	4,844,793	5,680,283

Bold indicates the best result among all the compared methods for a specific dataset and measurement

4 Conclusion

This research work proposed a hyperspectral image classification method called FA-CNN. FA-CNN extracted spectral features from the hyperspectral data using an unsupervised dimension reduction technique called factor analysis. Factor analysis extracted the underlying factors that explain the original spectral information in fewer dimensions. Combining effective spectral feature extraction with spatial information using CNN, FA-CNN achieved high accuracy using a relatively small training set and lower number of trainable parameters. Our method compared favourably to several state-of-the-art methods. In future, we want to apply this method to other hyperspectral datasets to leverage the power of hyperspectral imaging in solving various problems.

References

1. Bioucas-Dias JM, Plaza A, Camps-Valls G, Scheunders P, Nasrabadi N, Chanussot J (2013) Hyperspectral remote sensing data analysis and future challenges. IEEE Geosci Remote Sens Mag 1(2):6–36
2. Donoho DL et al (2000) High-dimensional data analysis: the curses and blessings of dimensionality. AMS Math Challenges Lecture 1(2000):32
3. Hughes GF (1968) On the mean accuracy of statistical pattern recognizers. IEEE Trans Inf Theory IT 14(1):55-63
4. Rodarmel C, Shan J (2002) Principal component analysis for hyperspectral image classification. Surveying Land Inf Sci 62(2):115–122
5. Melgani F, Bruzzone L (2004) Classification of hyperspectral remote sensing images with support vector machines. IEEE Trans Geosci Remote Sens 42(8):1778–1790
6. Audebert N, Le Saux B, Lefèvre S (2019) Deep learning for classification of hyperspectral data: a comparative review. IEEE Geosci Remote Sens Mag 7(2):159–173
7. Li S, Song W, Fang L, Chen Y, Ghamisi P, Benediktsson JA (2019) Deep learning for hyperspectral image classification: an overview. IEEE Trans Geosci Remote Sens 57(9):6690–6709
8. Ahmad M (2020) A fast 3D CNN for hyperspectral image classification. arXiv:2004.14152

9. Chen Y, Jiang H, Li C, Jia X, Ghamisi P (2016) Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. IEEE Trans Geosci Remote Sens 54(10):6232–6251
10. Gao Q, Lim S, Jia X (2018) Hyperspectral image classification using convolutional neural networks and multiple feature learning. Remote Sens 10(2):299
11. Haque M, Zaman S (2019) Spectral-spatial feature extraction using PCA and multi-scale deep convolutional neural network for hyperspectral image classification, pp 1–6
12. Makantasis K, Karantzalos K, Doulamis A, Doulamis N (2015) Deep supervised learning for hyperspectral data classification through convolutional neural networks. In: 2015 IEEE International geoscience and remote sensing symposium (IGARSS). IEEE, pp 4959–4962
13. Roy SK, Krishna G, Dubey SR, Chaudhuri BB (2019) HybridSN: exploring 3-D-2-D CNN feature hierarchy for hyperspectral image classification. IEEE Geosci Remote Sens Lett 17(2):277–281
14. Kassim S, Hasan H, Mohd Ismon A, Muhammad Asri F (2013) Parameter estimation in factor analysis: maximum likelihood versus principal component. AIP Conf Proc 1522(1):1293–1299
15. Bartlett MS (1937) The statistical conception of mental factors. British J Psychol 28(1):97
16. Huang L, Chen Y (2020) Dual-path siamese CNN for hyperspectral image classification with limited training samples. IEEE Geosci Remote Sens Lett
17. Lange J, Cavallaro G, Götz M, Erlingsson E, Riedel M (2018) The influence of sampling methods on pixel-wise hyperspectral image classification with 3d convolutional neural networks. In: IGARSS 2018-2018 IEEE international geoscience and remote sensing symposium. IEEE, pp 2087–2090