A Mutual Guide Framework for Hyperspectral Image Classification with Small Training Dataset

Yaomin Xue
Department of Electrical and Computer Engineering
Memorial University of Newfoundland
St. John's, NL, Canada
yxue@mun.ca

Weimin Huang
Department of Electrical and Computer Engineering
Memorial University of Newfoundland
St. John's, NL, Canada
weimin@mun.ca

***Abstract*—A mutual guide framework (MGF) with two multi-kernels classifier models is proposed for hyperspectral images (HSIs) classification using a small number of training data. HSIs spectral and spatial features obtained by superpixelwise principal component analysis (SuperPCA) are input of the classifiers. The two base SVM classifiers are first trained with the same but small training dataset. Then, new classification results with high confidence of each SVM classifier are input to the other SVM classifier to augment their training data size iteratively till the size is increased from 5% to 10% of total data. Unlike existing MGF in which SuperPCA is implemented as a feature extractor and the extreme learning machines (ELMs) are applied as feature classifiers, the ELMs are replaced by support vector machine (SVM) in this study. Experiment results from three public datasets, i.e., Indian Pines, University of Pavia, and Salinas, demonstrate the classification accuracy of the proposed framework reaches 97%, 96%, and 99%, respectively.**

***Index Terms*—hyperspectral images (HSIs) classification, SVM, small training data**

I. Introduction

Recent advances in airborne and spaceborne hyperspectral imaging technology have endowed hyperspectral sensors with the capability to provide measurements about the surface of the Earth with a very high spectral resolution [2]. The sensors achieve this by sampling the reflective portion of the electromagnetic spectrum, ranging from the visible spectrum to the short-wave infrared spectrum [3]. As one of the most common issues in hyperspectral images (HSIs), HSIs classification focuses on extracting effective features and combining complex spectral and spatial information [4].

In the last two decades, many spectral feature-based methods have been proposed and applied to solve the HSIs classification problem effectively. Initially, various conventional methods were applied to HSIs feature extraction by exploiting certain spectral-spatial structures or patterns in the neighbouring pixels of a hyperspectral image. For instance, in [5]–[7], the strategies of predetermined masks, filters, and the Laplacian matrix were employed to extract spectral-spatial co-occurrence, wavelets, and Gabor features [8] in the neighbouring pixels of an HSIs image. Additionally, convolutional neural networks (CNN) [9], [10] under the feedback of classification output were utilized to adjust the neighbourhood contributions to spectral-spatial features [8]. These methods adopt a certain image-sensitive factor to tune the importance of pixels in a neighbourhood. However, such a capacity tends to make spatial dependency adapt to the specific characteristics of an image and it is explicitly data-dependent [8].

Generally, pre-processing-based methods rather than integrated methods have shown superior performance in HSIs classification. The classification process can be typically divided into two phases: spectral-spatial feature extraction and feature-wise classification. For example, Tang et al. [11] proposed a 3-D scattering wavelet transform to extract scattering features and employed a support vector machine (SVM) based on the Gaussian kernel as a classifier. Similarly, Jiang et al. [12] proposed a SuperPCA approach that involved principal component analysis (PCA) to extract contextual information from HSIs. Subsequently, various classifiers were implemented, including sparse representation [13], nearest regularized subspace [14], and extreme learning machines (ELMs) [15]. Although these pre-processing-based methods have achieved remarkable performance, HSIs classification is a challenging problem as the number of suitable training samples is scarce [16]. Many of the powerful classification capabilities require training several samples and may result in overfitting. Additionally, noisy labels in a significant proportion of samples may lead to misleading training samples.

Considering the significance of HIC with small training data and the vulnerability of the preprocessing-based methods subject to noisy labels, Tai et al. [1] developed a novel mutual guide framework (MGF) and implemented SuperPCA and ELMs to form a baseline for HSIs classification. However, ELMs are not effective when dealing with a high number of spectral bands, since they only focus on the shallow features and ignore the deeper information. In this paper, we implement MGF as the framework and apply SuperPCA as a feature extractor and SVMs as the pixel classifier. We trained two multi-kernel SVM classifiers simultaneously in an iterative manner using a small number of training data. The two base SVM classifiers are first trained with the same but small training dataset. Then, new classification results with high confidence of each SVM classifier are input to the other SVM classifier to augment their training data size iteratively till the size is increased from 5% to 10% of total data. Such a framework is formulated to also solve the noise reduction and model selection issue.

II. Proposed Method for HSIs Classification

A. Framework of the Proposed Method

MGF develops a base framework and is applied for training two multi-kernel classifiers in this project. Fig. 1 shows the framework of the MGF.

The two classifiers are trained iteratively. Let C_A^k and C_B^k denote two base classifiers with different classification kernels. Let X_L^1 denotes the randomly 5% of the total HSIs data as the initial training data for the C_A^1 and C_B^1. Then, Y_A^1 is the new classification results of the C_A^1 and Y_B^1 is the new classification results of the C_B^1. Training the two classifiers at the kth iteration is described as follows

$$C_A^k\left(X_{L+B}^k\right) = C_A^k\left(X_A^{k-1} + \Psi\left(Y_B^{k-1}\right)\right) \tag{1}$$

where X_{L+B}^k and X_{L+A}^k represent the training input of the classifier C_A^k and C_B^k, respectively. X_{L+B}^k and X_{L+A}^k are increasing iteratively with $\Psi\left(Y_B^{k-1}\right)$ and $\Psi\left(Y_A^{k-1}\right)$ as the added parts at the kth iteration. Y_A^{k-1} and Y_B^{k-1} represent the classification results of the classifier C_A^{k-1} and C_B^{k-1}. $\Psi(\cdot)$ is the selection function that randomly chooses 0.5% of the classification result with accurate labels (high confidence). Thus, the two base classifiers initially trained with the original small data independently keep providing high confidence classification results as guide labels to each other and in turn enhancing their training effectiveness separately. The classification result comes from one of the classifiers with better performance after the 10 iterations.

B. Classification Architecture Selection

Although the SuperPCA and ELMs in MGF are implemented as feature extractor and classifiers for the HSIs classification, ELMs are based on the linear transformation matrix and only focus on shallow features. The SVM replaces the ELMs in this project due to its success and popularity in the HSIs classification field and its lower sensitivity to the curse of dimensionality [2].

Given a hyperspectral image composed of N pixels characterized by a M dimensional spectral vector of features, $X = \{X_1, X_2, \ldots, X_N\}$, with $X_N = [X_{n1}, X_{n2}, \ldots, X_{nM}]$. We further suppose that the set of pixels can be classified into one and only one of the K unordered classes denoted by $S = [U_1, U_2, \ldots, U_K]$. Let $Y = \{Y_1, Y_2, \ldots, Y_N\}$ denotes the label indicator for X, and $Y_i \in S$.

Firstly, the SuperPCA acts as a feature extractor and employed to extract low-dimension features, which can be described by

$$\hat{X}_N = \mathrm{Cov}\left(X_N\right) \tag{2}$$

where $\mathrm{Cov}(\cdot)$ represents the covariance matrix, and $\hat{X}_N = \left[X_{n1}, X_{n2}, \ldots, X_{n\hat{M}}\right]$ is the feature vector where $\hat{M}$ is much smaller than M. Therefore, the dimensionality of the data X is reduced from M to $\hat{M}$.

Then, SVM is a kernel-based algorithm creates hyperplanes that $w * X + b$, to divide the X data into K classes. The optimization function of SVM can be described by

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j Y_i Y_j * K\left(X_i X_j\right) - \sum_{j=i}^{N} \alpha_i \tag{3}$$

where $\alpha_i \alpha_j$ is the solution to find the hyperplanes for linear SVM model with nonlinear kernel function $K\left(X_i X_j\right)$ mapping from the input spectral space, $\hat{X}_i = \left[X_{i1}, X_{i2}, \ldots, X_{i\hat{M}}\right]$ and $\hat{X}_j = \left[X_{j1}, X_{j2}, \ldots, X_{j\hat{M}}\right]$, to the feature space. We implement radial basis kernel function to SVM classifier C_A^k and polynomial kernel function to SVM classifier C_B^k. The two SVM models based on different kernels are implemented and show different performances in making the hyperplane decision boundary between the classes. The result $\alpha_i \alpha_j$ that minimizes (3) defines the hyperplane for spectral space $\hat{X}_i$ and $\hat{X}_j$ to classify them. There is a hyperplane between every two pairs of spectral vectors. After classifying spectral features, probability estimate values are obtained for comparison with ground truth label Y. The accurate prediction results of one classifier are used as guided labels of another classifier to enhance the training effectiveness. The final prediction result is chosen from one of the classifiers with better performance after completing all of the iterations.

III. Experiments and Discussion

We empirically evaluate the classification methods by making qualitative and quantitative comparisons and applying them on three public HSIs datasets, i.e., Indian Pines (IP), University of Pavia (UP), and Salinas (SA).

A. Training Data

The data sets employed in this work are captured by different sensors and cover different areas with different spectral bands range. The first dataset is IP, which was captured in northwestern Indiana, USA, and possesses 145 * 145 spatial pixels and 200 available reflectance spectral bands ranging from 0.4 to 2.5 m with 16 different land coverings (classes). The second dataset is UP, which was captured in UP, Italy, and possesses 610 * 340 spatial pixels and 103 available reflectance spectral bands with 9 classes in the ground truth. The third dataset is SA, which was captured over the SA Valley, California, USA. It possesses 512 * 217 spatial pixels and 184 available reflectance spectral bands with 16 classes in the ground truth.

B. Experimental Configurations

The experiments have been executed using two different classification architectures: the proposed SVM based on MGF with SuperPCA (SMS) taken for comparison with the mutual guide method [1]. For both two classification architectures, 5% of the total data are randomly selected as the training data and the remaining 95% of total data are for testing. 10 iterations are conducted. 10% of the total data are applied as training data finally.

To measure the classification performance, we apply three critical metrics of HSIs classification, ie., overall accuracy

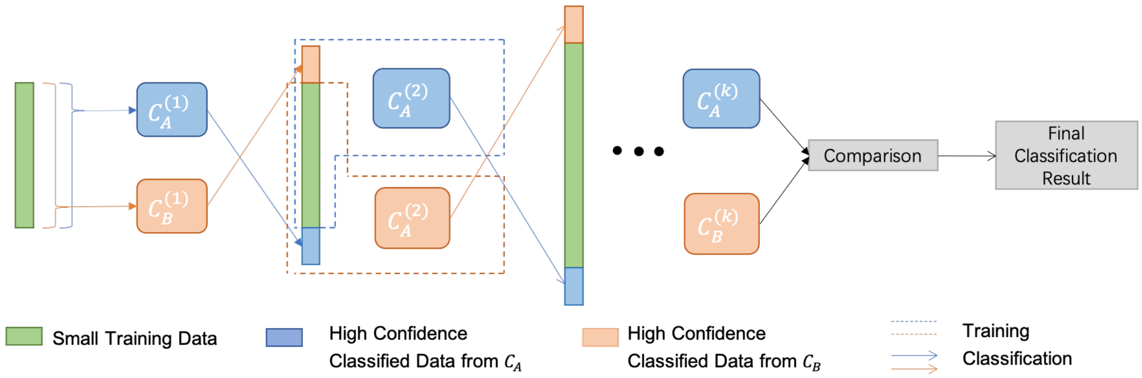


Fig. 1: The framework of MGF.

(OA), average accuracy (AA), and increased accuracy (IA). OA is the percentage of all accurate classified pixels over the total pixels. AA is the average accurate prediction percentage for each class. IA indicates the increase in accuracy after iterations, which also be used to measure the effectiveness of the MGF framework.

C. Qualitative Evaluations

The classification maps obtained for three data sets by the proposed SMS method and the mutual guide are visually demonstrated in Fig. 2. The ground truth and the classification result for different classes are illustrated with different colors.

At the same time, the performance trend of the two methods is compared with respect to iterations in Fig. 3. The curves A and B in Fig. 3 represent the OA and AA in SMS, and the curves C and D represent the OA and AA in mutual guide. With the training iterations increasing, the classification accuracy of both two methods is improving. Comparing the OA and AA (for the two methods) in Fig. 3(a) and Fig. 3(b), the classification performance for SMS is better than the mutual guide. Although the OA and AA for the mutual guide in Fig. 3(c) are slightly higher than the SMS, the classification accuracy for SMS is more than 99.6% which means there is only a slight misclassification. This demonstrates the SMS method improves the training and classification performance gradually and can be a general approach to improve the representational power of classifiers. Additionally, the prediction results on test data achieve acceptable performance and suggest the good robustness of the SMS. The specific accuracy for each class on the IP datasets, UP datasets and SA datasets is shown in Table. I, Table. II and Table. III, respectively.

CONCLUSION

A method for HSIs classification based on MGF with small training data is proposed. Specifically, SuperPCA carries out dimensionality reduction for HSIs spectral and spatial features, and then input features to the SVM classifier for classification. The two SVM classifiers with different kernels are first trained with 5% of the total data and then applied to predict the rest 95% data. The classification results with high confidence of each SVM classifier are input to the other SVM classifier to

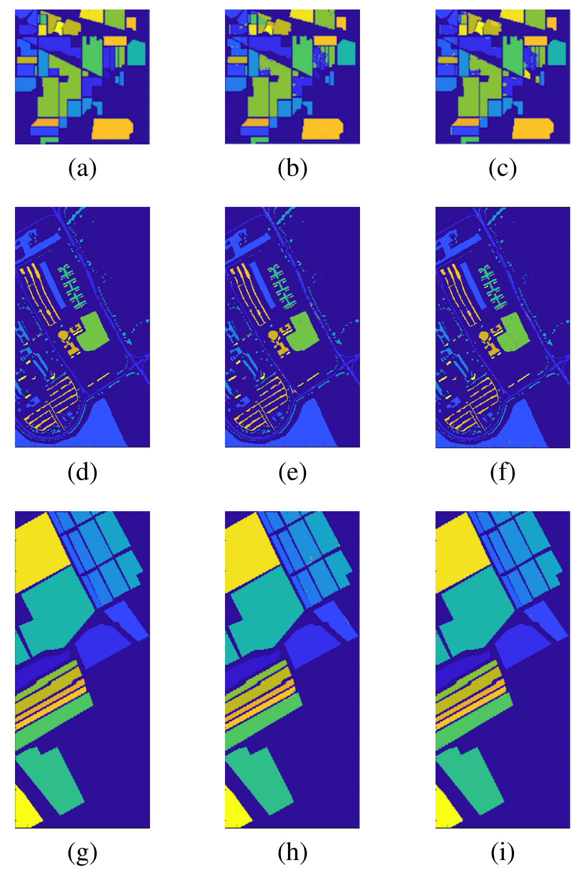


Fig. 2: Visual illustration of classification. (a) Ground truth on the IP datasets. (b) Proposed SMS on the IP datasets. (c) Original MGF on the IP datasets. (d) Ground truth on the UP datasets. (e) Proposed SMS on the UP datasets. (f) Original MGF on the UP datasets. (g) Ground truth on the SA datasets. (h) Proposed SMS on the SA datasets. (i) Original MGF on the SA datasets.

TABLE I
Classification Accuracy on the IP Dataset (%)

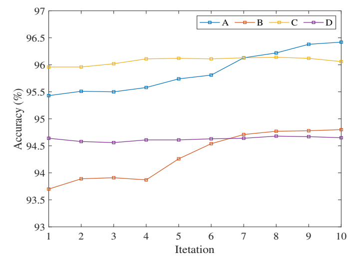
class	The mutual guide	Proposed SMS
1	**100**	**100**
2	91.80	94.47
3	94.95	90.60
4	96.40	98.73
5	96.27	96.27
6	99.85	**100**
7	96.00	96.43
8	99.57	99.58
9	**100**	**100**
10	97.70	95.88
11	97.82	94.62
12	83.16	**90.89**
13	99.49	99.51
14	99.75	99.68
15	98.64	99.22
16	60.44	**97.85**
OA	96.01	**96.22±0.15**
AA	94.65	**94.8±0.23**
IA	0.1	0.09

TABLE II
Classification Accuracy on the UP Dataset (%)

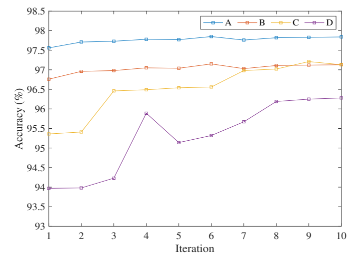
class	The mutual guide	Proposed SMS
1	92.76	94.98
2	99.22	99.07
3	98.95	97.86
4	89.10	90.01
5	96.80	97.03
6	99.29	98.39
7	97.14	93.08
8	97.55	98.21
9	95.97	**99.05**
OA	97.13	**97.34±0.09**
AA	96.30	**96.45±0.25**
IA	1.23	0.21

TABLE III
Classification Accuracy on the SA Dataset (%)

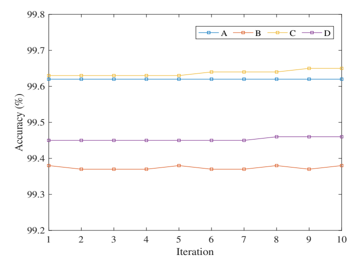
class	The mutual guide	Proposed SMS
1	**100**	**100**
2	99.97	99.33
3	99.57	**100**
4	99.62	98.78
5	99.37	99.07
6	99.42	99.42
7	99.47	99.44
8	99.87	99.96
9	**100**	**100**
10	98.40	98.69
11	99.50	97.75
12	99.95	**100**
13	98.16	98.25
14	98.24	97.48
15	99.97	99.94
16	**100**	**100**
OA	99.65	**99.62±0.02**
AA	99.45	**99.29±0.05**
IA	0.02	0.03



(a)



(b)



(c)

Fig. 3: Performance trend of the two methods with respect to iterations. (a) IP datasets. (b) UP datasets. (c) SA datasets..

augment their training data size to 10% iteratively and enhance the training effectiveness. The performance evaluated across three public datasets indicates that the proposed SMS is an effective approach to improve the accuracy and robustness of HSIs classification. Future research could be implementing CNN as the feature extractor to classify the spectral and spatial information.

References

[1] X. Tai, M. Li, M. Xiang, and P. Ren, "A mutual guide framework for training hyperspectral image classifiers with small data," *IEEE Trans. Geosci. Remote Sens.*, early access, Jul. 2021, doi: 10.1109/TGRS.2021.3092351.

[2] Y. Bazi and F. Melgani, "Toward an optimal SVM classification system for hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, pp. 3374-3385, Nov. 2006.

[3] P. Ghamisi et al., "Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art," *IEEE Geosci. Remote Sens.* Mag., vol. 5, no. 4, pp. 37–78, Dec. 2017.

[4] F. Wang, R. Zhang, and Q. Wu. "Hyperspectral image classification based on PCA network," *Workshop Hyperspectral Image Signal Processing: Evolution Remote Sens*, Aug. 2016, pp. 1-4.

[5] F. Tsai and J.-S. Lai, "Feature extraction of hyperspectral image cubes using three-dimensional gray-level cooccurrence," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 6, pp. 3504–3513, Jun. 2013.

[6] L. He, Y. Li, X. Li, and W. Wu, "Spectral–spatial classification of hyperspectral images via spatial translation-invariant wavelet-based sparse representation," *IEEE Trans. Geosci. Remote Sens*., vol. 53, no. 5, pp. 2696–2712, May 2015.

[7] Z. Zhong et al., "Discriminant tensor spectral–spatial feature extraction for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett*., vol. 12, no. 5, pp. 1028–1032, May. 2015.

[8] L. He, J. Li, C. Liu, and S. Li. "Recent advances on spectral–spatial hyperspectral image classification: An overview and new guidelines", *IEEE Trans. Geosci. Remote Sens*., vol. 56, no. 3, pp. 1579-1597, Jun. 2013.

[9] W. Zhao and S. Du, "Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544–4554, Aug. 2016.

[10] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens*., vol. 54, no. 10, pp. 6232–6251, Oct. 2016.

[11] Y. Y. Tang, Y. Lu, and H. Yuan, "Hyperspectral image classification based on three-dimensional scattering wavelet transform," *IEEE Trans. Geosci. Remote Sens*., vol. 53, no. 5, pp. 2467–2480, May 2015.

[12] J. Jiang, J. Ma, C. Chen, Z. Wang, Z. Cai, and L. Wang, "SuperPCA: A superpixelwise PCA approach for unsupervised feature extraction of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens*., vol. 56, no. 8, pp. 4581–4593, Aug. 2018.

[13] L. Huo, L. Zhao, and P, Tang, "Hyperspectral image classification with sparse representation classifier and active learning," *Workshop Hyperspectral Image Signal Processing: Evolution Remote Sens*, Aug. 2016, pp. 1-5.

[14] B. Peng, X. Xie, W. Li, and Q. Du, "Hyperspectral image classification using multiple features and nearest regularized subspace," *Workshop Hyperspectral Image Signal Processing: Evolution Remote Sens*, Jun. 2015, pp. 1-4.

[15] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, Dec. 2006.

[16] B. Pan, Z. Shi, and X. Xu, "R-VCANet: A new deep-learning-based hyperspectral image classification method," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens*., vol. 10, no. 5, pp. 1975–1986, May 2017.