A NOVEL CLASSIFICATION PROCESSING BASED ON THE SPATIAL INFORMATION AND THE CONCEPT OF ADABOOST FOR HYPERSPECTRAL IMAGE CLASSIFICATION

Bor-Chen Kuo[1] *Shih-Syun Lin*[1] *Huey-Min Wu*[1] *Chun-Hsiang Chuang*[2]
kbc@mail.ntcu.edu.tw catchylss@hotmail.com lhswu@seed.net.tw cch.chuang@gmail.com

[1]Graduate School of Educational Measurement and Statistics,
National Taichung University, Taiwan, R.O.C.
[2]Department of Electrical and Control Engineering, National Chiao Tung University, Taiwan, R.O.C.

1. INTRODUCTION

In this paper, a novel classification processing based on the spatial information and the concept of Adaboost [1] for hyperspectral image classification is proposed. This classification process is named adaptive feature extraction with spatial information (AdaFESI). The main idea is adaptive in the sense that subsequent feature spaces are tweaked in favor of those instances misclassified by spectral or spatial classifiers in the previous feature space. This processing includes two concepts for classifying hyperspectral image: (1) For avoiding the Hughes phenomenon, the feature extraction is the important for hyperspectral image classification [2]-[3]. Hence, the feature space at the next round is varied at every round such that it suits for the misclassified samples at this round. The weights of the terms of the scatter matrices corresponding to the samples which are classified correctly at this round will be decreased in the next round. Otherwise, the weights will be increased for the samples which are misclassified. (2) Many studies [4]-[5] show that the performance of the classifier with spatial information outperforms than of the original one. Hence, which one of the spatial classifier and spectral classifier used at every round is determined by their classification performances at this round. The traditional hyperspectral image classification procedure is a special case of our proposed processing because it is the same to perform our proposed method one round without using spatial classifier. Fig. 1 shows the flowchart of AdaFESI. Note that any type of classifier and feature extraction method can be used in our proposed procedure.

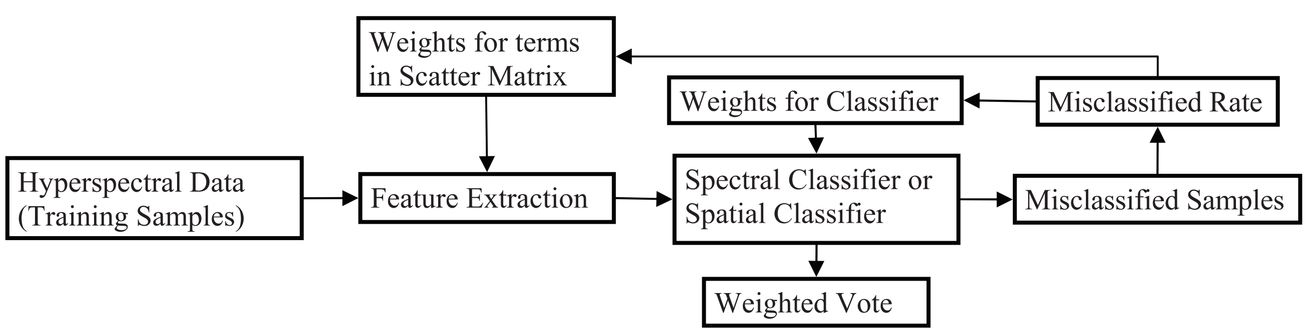


Fig. 1 The flowchart of AdaFESI

2. NONPARAMETRIC WEIGHTED FEATURE EXTRACTION BASED ADAPTIVE FEATURE EXTRACTION

Many researches about hyperspectral image classification also show that nonparametric weighted feature extraction (NWFE) [6] is a powerful tool for extracting hyperspectral image features [7]-[8]. Hence, NWFE will be used in AdaFESI. Let $x_\ell^{(i)}$ be the ℓ-th samples in class i, P_i denote the prior probability of class i, and $dist(x,z)$ be the Euclidean distance from x to z. At each round, the effects of terms in the scatter matrices are modified by the weights $\eta_t(x_\ell^{(i)})$. The modified between-class scatter matrix $S_b^{NW}(t)$ and the within-class scatter matrix $S_w^{NW}(t)$ of NWFE at the round t are defined as

$$S_b^{NW}(t)=\sum_{i=1}^{L}P_i\sum_{\substack{j=1\\ j\neq i}}^{L}\sum_{\ell=1}^{N_i}\frac{\eta_t(x_\ell^{(i)})\lambda_\ell^{(i,j)}}{N_i}(x_\ell^{(i)}-M_j(x_\ell^{(i)}))(x_\ell^{(i)}-M_j(x_\ell^{(i)}))^T \text{ and } S_w^{NW}=\sum_{i=1}^{L}P_i\sum_{\ell=1}^{N_i}\frac{\eta_t(x_\ell^{(i)})\lambda_\ell^{(i,i)}}{N_i}(x_\ell^{(i)}-M_i(x_\ell^{(i)}))(x_\ell^{(i)}-M_i(x_\ell^{(i)}))^T$$

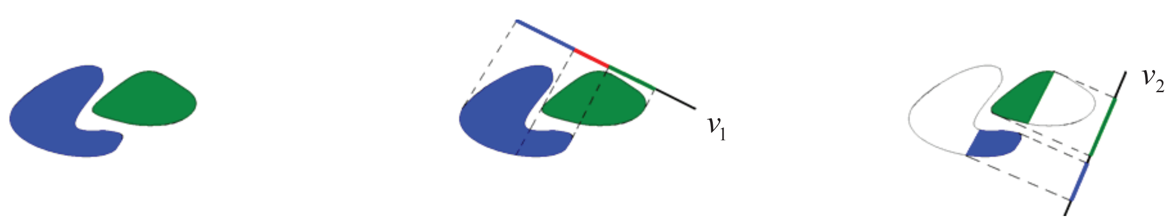
where the scatter matrix weight $\lambda_\ell^{(i,j)}$ is defined by

$$\lambda_\ell^{(i,j)}=\frac{dist(x_\ell^{(i)},M_j(x_\ell^{(i)}))^{-1}}{\sum_{t=1}^{N_i}dist(x_t^{(i)},M_j(x_t^{(i)}))^{-1}},\quad M_j(x_\ell^{(i)})=\sum_{k=1}^{N_j}w_{\ell k}^{(i,j)}x_k^{(j)},\text{ and } w_{\ell k}^{(i,j)}=\frac{\mathrm{dist}(x_\ell^{(i)},x_k^{(j)})^{-1}\eta_t(x_k^{(j)})}{\sum_{t=1}^{N_j}\mathrm{dist}(x_\ell^{(i)},x_t^{(j)})^{-1}\eta_t(x_t^{(j)})}$$

denotes the weighted mean with respect to $x_\ell^{(i)}$ in class j at the round t.

3. ADAPTIVE FEATURE EXTRACTION WITH SPATIAL INFORMATION

The main idea of AdaFESI is to find a suitable feature space for those data which are difficult to identify in the previous feature space. As shown in Fig. 2, there are two distributions. The red region in Fig. 2(b) is composed of the projections with respect to misclassified samples at the first round. The weights with respect to these samples are increased in the second round. Hence, in Fig. 2(c), the extracted feature at the second round will be more suitable for those hard identified data at the first round.



(a) Two distributions

(b) The extracted feature v_1 at the first round

(c) The extracted feature v_2 at the second round

Fig. 2 Adaptive feature extraction with spatial information

The algorithm of AdaFESI with spatial information is described in the following.

Algorithm 1. Adaptive feature extraction with spatial information

Input:

(1) The training data $x_\ell^{(i)}$, $\ell = 1,\ldots,N_i, i = 1,\ldots,L$.

(2) The test sample z .

(3) The classifier methods, $\psi^{spectral}(\cdot)$, $\psi^{spatial}(\cdot)$ with output at the round *t*.

(4) The reduced dimension, p .

Output:

The label y of the test sample z is given by the ensemble.

A. Training Procedure:

(1) Initialize weight: $\eta(x_\ell^{(i)}) = \frac{1}{N}$, $\ell = 1,\ldots,N_i$, $i = 1,\ldots,L$

(2) Let stop parameter q

- Do for $t = 1$ to ... until $|err_t\text{-}err_{t\text{-}1}| < q$
- To build the classification including the $A_t \in \Re^{d\times p}$ linear transformation.
- Applying NWFE with $S_b^{Ada\_NW}(t)$ and $S_w^{Ada\_NW}(t)$.
  - $S_b^{Ada\_NW}(t) = \sum_{i=1}^{L} P_i \sum_{\substack{j=1 \\ j\neq i}}^{L} \sum_{\ell=1}^{N_i} \frac{\eta_t(x_\ell^{(i)})\lambda_\ell^{(i,j)}}{N_i} (x_\ell^{(i)} - M_j(x_\ell^{(i)}))(x_\ell^{(i)} - M_j(x_\ell^{(i)}))^T$
  - $S_w^{Ada\_NW}(t) = \sum_{i=1}^{L} P_i \sum_{\ell=1}^{N_i} \frac{\eta_t(x_\ell^{(i)})\lambda_\ell^{(i,i)}}{N_i} (x_\ell^{(i)} - M_i(x_\ell^{(i)}))(x_\ell^{(i)} - M_i(x_\ell^{(i)}))^T$
- Creating the $h_t(x_\ell^{(i)}) = \psi^{spectral}\left(A_t^T x_\ell^{(i)}\right)$ and $h'_t(x_\ell^{(i)}) = \psi^{spatial}\left(A_t^T x_\ell^{(i)}\right)$.
- To estimate ε_t and ε'_t .
  - $\varepsilon_t = \frac{\sum \tau(\cdot)}{i}$, where $\tau(\cdot) = \begin{cases} 1 & h_t(x_\ell^{(i)}) = \text{true label} \\ 0 & h_t(x_\ell^{(i)}) \neq \text{true label} \end{cases}$.
  - $\varepsilon'_t = \frac{\sum \vartheta(\cdot)}{i}$, where $\vartheta(\cdot) = \begin{cases} 1 & h'_t(x_\ell^{(i)}) = \text{true label} \\ 0 & h'_t(x_\ell^{(i)}) \neq \text{true label} \end{cases}$.
  - $\lambda_t = \begin{cases} 1 & if \quad \varepsilon_t < \varepsilon'_t \\ 0 & otherwise \end{cases}$
  - $err_t = \lambda_t \varepsilon_t + (1-\lambda_t)\varepsilon'_t$
- To estimate $\eta_{t+1} = f_1\left(x_\ell^{(i)}, err, \lambda_t\right)$.
  - $f_1(x_\ell^{(i)}, \lambda_t) = \begin{pmatrix} \eta_t(x_\ell^{(i)}) \times \lambda_t\left((1-\delta(h_t(x_\ell^{(i)}), i)) + \delta(h_t(x_\ell^{(i)}), i) \times (10)^{-r}\right) \\ +(1-\lambda_t)\left((1-\delta(h'_t(x_\ell^{(i)}), i)) + \delta(h'_t(x_\ell^{(i)}), i) \times (10)^{-r}\right) \end{pmatrix} / Z_t$.

  Where Z_t is the normalization constants, and $\delta(a,b) = \begin{cases} 1 \; if & a = b \\ 0 \; if & a \neq b \end{cases}$.
- To estimate $\alpha_t = f_2(err)$, where $f_2 = -\log(err/(1-err))$ is used to estimate classifier's weights at the round *t*.

B. Classification Procedure:

$y = \arg\max_{c\in\{1,\ldots,L\}} \sum_{t=1}^{10} \alpha_t (\lambda_t \delta(h_t(z), c) + (1-\lambda_t)\delta(h'_t(z), c))$

4. SOME EXTERIMENTAL RESULTS

There are two data sets in our experiments. They are the Washington, DC Mall hyperspectral image [9] as an urban site and the Indian Pine Site [10]. In this paper, for investigating the influences of training sample sizes to the dimension, three distinct cases, $N_i = 20 < N < d$ (case 1), $N_i = 40 < d < N$ (case 2), and $d < N_i = 300 < N$ (case 3), will be discussed. Due to these sample size constraints, some of the classes in selected hyperspectral images for the experiment are used. At each experiment, ten training and testing datasets are randomly selected for estimating system parameters and computing the average accuracy of testing data (100 testing samples per class) of different algorithms respectively. The experimental result of methods is as following:

Table 1. The experimental result of methods

Feature Extraction	Classifier	case 1		case 2		case 3	
		DC	SITE	DC	SITE	DC	SITE
NWFE	Gaussian	87.7 % (4)	77.3 % (8)	92.0 % (4)	82.6 % (8)	94.1 % (15)	89.3 % (13)
NWFE	*k*nn	88.9 % (4)	80.9 % (11)	91.2 % (7)	85.0 % (9)	95.1 % (8)	93.5 % (13)
NWFE	CART	79.8 % (2)	61.1 % (5)	86.0 % (4)	69.6 % (7)	91.5 % (8)	82.3 % (8)
AdaFESI	Gaussian	**93.1 % (8)**	80.2 % (8)	94.6 % (14)	87.0 % (11)	95.0 % (14)	91.1 % (15)
AdaFESI	*k*nn	92.6 % (3)	**83.7 % (14)**	94.3 % (7)	**89.1 % (6)**	**98.1 % (8)**	**96.4 % (8)**
AdaFESI	CART	91.6 % (3)	75.3 % (6)	**94.7 % (10)**	81.8 % (5)	97.9 % (8)	94.9 % (7)

5. REFERENCES

[1] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences,* Vol. 55, No. 1, pp. 119-139, 1997.

[2] D. A. Landgrebe, "Information Extraction Principles and Methods for Multispectral and Hyperspectral Image Data," Chapter 1 of *Information Processing for Remote Sensing*, edited by C. H. Chen, published by the World Scientific Publishing Co., Inc., 1060 Main Street, River Edge, NJ 07661, USA 1999.

[3] David Landgrebe, *Signal Theory Methods In Multispectral Remote Sensing*, 508 pages plus a CD containing exercises and data. John Wiley & Sons, January 2003, ISBN 0-471-42028-X.

[4] Bruzzone, L. and Persello, C., " A Novel Context-Sensitive Semisupervised SVM Classifier Robust to Mislabeled Training Samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, issue 7, pp. 2142-2154, 2009.

[5] Bor-Chen Kuo, Chun-Hsiang Chuang, Chih-Sheng Huang and Chih-Cheng Hung (2009), "A Nonparametric Contextual Classification Based on Markov Random Fields," First Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, Aug. 2009, Grenoble, France.

[6] B.C. Kuo and D.A. Landgrebe, "Nonparametric weighted feature extraction for classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 5,pp. 1096–1105, May 2004.

[7] J. A. Richards, "Analysis of remotely sensed data: The formative decades and the future," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 422– 432, Mar. 2005.

[8] D. Landgrebe, "Multispectral land sensing: Where from, where to?" *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 414–421, Mar. 2005.

[9] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.

[10] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*, John Wiley and Sons, Hoboken, NJ: Chichester, 2003.