Discriminatively Guided Filtering (DGF) for Hyperspectral Image Classification

Ziyu Wang[a,b], Huafeng Hu[b], Lefei Zhang[c], Jing-Hao Xue[b,*]

[a] *Department of Security and Crime Science, University College London, London, UK*
[b] *Department of Statistical Science, University College London, London, UK*
[c] *Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China*

Abstract

In this paper, we propose a new filtering framework called discriminatively guided image filtering (DGF), for hyperspectral image (HSI) classification. DGF integrates a discriminative classifier and a generative classifier by the guided filtering (GF), considering the complementary strength of these two types of classification paradigms. To demonstrate the effectiveness of the proposed framework, the combination of support vector machine (SVM) and linear discriminative analysis (LDA), which serve as a discriminative classifier and a generative classifier respectively, is investigated in this paper. Specifically, the original HSI is projected into the low-dimensional space induced by LDA to serve as guidance images for filtering the intermediate classification results induced by SVM. Experiment results show the superior performance of the proposed DGF compared with that of the principal component analysis (PCA)-based GF.

Keywords: Hyperspectral image (HSI) classification, guided image filtering (GF), discriminative classifiers, generative classifiers, linear discriminant analysis (LDA), principal component analysis (PCA), support vector machine (SVM).

*Corresponding author.
Email addresses: `ziyu.wang.12@ucl.ac.uk` (Ziyu Wang), `huafeng.hu.15@ucl.ac.uk` (Huafeng Hu), `cslfzhang@comp.polyu.edu.hk` (Lefei Zhang), `jinghao.xue@ucl.ac.uk` (Jing-Hao Xue)

1. Introduction

Classification of pixels in hyperspectral images (HSIs) is an important and challenging task in remote sensing [1]. Recently, many classifiers have been developed on the basis of state-of-the-art machine learning techniques such as support vector machine (SVM) [2, 3], manifold and subspace learning [4, 5, 6, 7], sparse representation [8, 9, 10], collaborative representation [11], active learning [12, 13, 14], multitask learning [15], domain adaptation [16], object-based classifiers [17] and deep learning [18].

Because HSIs contain rich information in both spectral and spatial dimensions, the strategies simultaneously exploiting them prevail in the processing and analysis of HSIs [3, 19, 20, 21, 22, 23, 24]. In the pixel-wise classification, it often happens with undesired salt and pepper appearance if the spatial smoothness has not been adequately addressed [25, 26], therefore it is appealing to introduce an image filtering process to relieve this issue. Recently in 2D image processing, a new method called guided filtering (GF) [27] has proved an effective approach to edge-preserving smoothing (now a function *imguidedfilter* in the Image Processing Toolbox of MATLAB, The MathWorks, Inc.). The GF uses the content of a guidance image to guide the smoothing of the input image. It has been pioneered by [28] to facilitate the HSI classification. They first use SVM to classify the HSI resulting in binary classification maps for each class, then adopt GF to filter the classification maps. To build the guidance image, they conduct principal component analysis (PCA) on the HSI, and project the data onto the first principal component to obtain a virtual greyscale guidance image (or onto the first three principal components for a virtual colour guidance image). They show that the resultant PCA-based GF (short as PGF hereafter) can improve the classification performance of SVM on the HSI.

However, we shall show that PGF can be further revised, methodologically and empirically, if we can design a better-founded scheme to produce a superior guidance image for classification. In terms of classification, an ideal guidance image should be as similar as possible to the ground-truth map: it should be able to not only preserve edges, but also provide the between-class discriminative information that is crucial to classification. However, the PCA adopted by PGF is an unsupervised feature-extraction approach which does not consider the discriminative information between classes. In contrast, we would consider generative classifiers, which can serve as supervised feature-extraction approaches and have the capability of exploiting the labelling information of the training samples, a capacity that PCA lacks. Moreover, to the discriminative classifiers like the SVM adopted by PGF, the generative classifiers can provide complementary discriminative strength, another capacity that PCA lacks; for the complementarity of these two types of classifiers and the advantages of combining them together, see [29, 30, 31, 32, 33].

Therefore, in this paper we propose a new filtering framework for classification, which enables the integration of a discriminative classifier, such as SVM, and a generative classifier, which can construct an complementary guidance image for GF to be used in classification. We call this new framework the discriminatively guided filtering (DGF). Specifically, we adopt linear discriminative analysis (LDA) as a generative classifier to demonstrate the effectiveness of the proposed DGF. For a C-class HSI classification problem, the LDA-based DGF can be implemented as follows: we first perform multi-class LDA to obtain $C - 1$ directions of projection, then we project the HSI onto the first leading direction to build the virtual greyscale guidance image (or onto the first three leading directions for the virtual colour guidance image). In this way, the obtained guidance image preserves the discriminative information among multiple classes. It is also worth noting that the proposed framework of DGF is a principled framework, which is flexible and can be adapted to the combinations of any discriminative classifiers and generative classifiers that suit the HSI classification. The combination of SVM and LDA adopted here serves more as a case study to demonstrate the use and effectiveness of DGF. Experimental results show that the proposed DGF outperforms the PCA-based GF of [28], and both of them improve the performance of SVM substantially for HSI classification.

2. Guided filtering

2.1. Greyscale guided filtering

For an input image p (e.g. in our case a classification map resulted from SVM), the guided filtering (GF) [27] is assumed to bridge its filtering output q and the guidance I by using a local linear model. For a greyscale image p, the output q is assumed to be a linear transformation of I in a local window ω_k centred at the pixel k:

$$q_i = a_k I_i + b_k, \ \forall i \in \omega_k, \tag{1}$$

where i indexes a pixel in ω_k such that q_i and I_i are the (scalar) values of pixel i in q and I, and (a_k, b_k) are coefficients to be estimated for ω_k. This model preserves $\nabla q = a \nabla I$, which ensures that q preserves the edges in I.

Although (1) is a simple linear regression model, the coefficients a_k and b_k are solved by a ridge regression model to minimise the following optimisation function:

$$E(a_k, b_k) = \sum_{i \in \omega_k} \left\{ (a_k I_i + b_k - p_i)^2 + \epsilon a_k^2 \right\}, \tag{2}$$

where ϵ is the smoothing parameter to penalise a_k, and p_i is the filtering input. The solution of (2) is given by

$$a_k = \frac{\frac{1}{|\omega_k|} \sum_{i \in \omega_k} I_i p_i - \mu_k \bar{p}_k}{\sigma_k^2 + \epsilon}, \ b_k = \bar{p}_k - a_k \mu_k, \tag{3}$$

where μ_k and σ_k^2 are the mean and variance of pixel values in ω_k of the guidance image I, $|\omega_k|$ is the total number of pixels in ω_k, and $\bar{p}_k = \frac{1}{|\omega_k|} \sum p_i$ denotes the mean in ω_k of the input image p . Once a_k and b_k are solved, the filtering output q_i in (1) can be obtained.

2.2. Multi-band guided filtering

The GF has been extended to a colour guidance image (i.e. with three bands) in [27]. Similarly, it has no difficulty to generalise to a multi-band guidance image of d bands by rewriting the local linear model (1) as

$$q_i = \mathbf{a}_k^T \mathbf{I}_i + b_k, \forall i \in \omega_k, \tag{4}$$

where $\mathbf{I}_i$ is a d-dimensional vector for pixel i, $\mathbf{a}_k$ is a d-dimensional coefficient vector, and q_i and b_k are still scalars. Then the GF with a multi-band guidance image becomes

$$\mathbf{a}_k = (\Sigma_k + \epsilon U)^{-1}\left(\frac{1}{|\omega_k|}\sum_{i\in\omega_k}\mathbf{I}_i p_i - \mu_k\bar{p}_k\right),\ b_k = \bar{p}_k - \mathbf{a}_k^T\mu_k, \tag{5}$$

where μ_k is a $d \times 1$ mean vector, and Σ_k is a $d \times d$ covariance matrix, of the multi-band guidance image $\mathbf{I}$ in window ω_k, and U is a $d \times d$ identity matrix.

3. Discriminatively guided filtering (DGF)

As pioneered by [28], the GF not only can be used as an edge-preserving smoothing operator, but also can help HSI classification. In this direction, we propose a new filtering framework called discriminatively guided image filtering (DGF), to combine a discriminative classifier and a generative classifier by the GF. Specifically, to incorporate the discriminative information from HSI into the GF, we use LDA as the generative classifier to construct the guidance image.

In 2D image processing, the guidance I_i in (1) (or $\mathbf{I}_i$ in (4)) is a greyscale value (or RGB values) of pixel i in a local window ω_k. However in the PGF [28] and the proposed DGF, the guidance I_i (or $\mathbf{I}_i$) is the projection of pixel i in the lower dimensional space induced by PCA and LDA, respectively.

3.1. Methodology of DGF

An HSI classification problem usually needs to address multiple classes (i.e. C classes with $C > 2$). An HSI of N pixels can be denoted by a $B \times N$ matrix $\mathbf{X}$ with each pixel being of B features, and $N > B$ as usual. For such a multi-class problem, the multi-class LDA seeks $C - 1$ directions (or say $C - 1$ linear combinations of the B features), the subspace spanned by which can best separate the classes [34].

These directions are the first $C - 1$ leading eigenvectors (corresponding to the largest eigenvalues) of

$$\mathbf{S}_W^{-1}\mathbf{S}_B, \tag{6}$$

where $\mathbf{S}_W$ is the pooled within-class scatter matrix over all C classes, and $\mathbf{S}_B$ is the between-class scatter matrix. We use $\mathbf{W}$ to denote a $B \times (C-1)$ matrix whose columns are these $B \times 1$ eigenvectors. Given $\mathbf{W}$, each class can have a unified multi-band guidance image for the GF. To align with the PCA-based GF [28], we also adopt the two strategies below.

3.1.1. DGF-g: DGF with a greyscale guidance image

We use the projection of the HSI on the first leading eigenvector ($\mathbf{w}_g$, a $B\times 1$ vector) in $\mathbf{W}$ as the greyscale guidance image:

$$\mathbf{I}_g = \mathbf{w}_g^T \mathbf{X}, \tag{7}$$

where $\mathbf{I}_g$ is a $1 \times N$ vector representing the greyscale guidance image of N pixels. The filtering output for each class is then obtained by the greyscale guided filtering in (1).

3.1.2. DGF-c: DGF with a colour guidance image

We use the projection of the HSI on the first three leading eigenvectors ($\mathbf{W}_c$, a $B \times 3$ matrix) in $\mathbf{W}$ as the colour guidance image:

$$\mathbf{I}_c = \mathbf{W}_c^T \mathbf{X}, \tag{8}$$

where $\mathbf{I}_c$ is a $3 \times N$ matrix representing the colour guidance image. The filtering output for each class is then obtained through the multi-band guided filtering in (4).

3.2. Classification algorithm based on DGF

The diagram of DGF for classification of an HSI is illustrated in Fig. 1, similar to the procedure in [28] but with a different filtering strategy. The whole HSI is firstly classified by SVM to obtain initial classification results called classification maps, one for each class, which contains the probability of the pixels belonging to the class, e.g. $\mathcal{C}_j$.

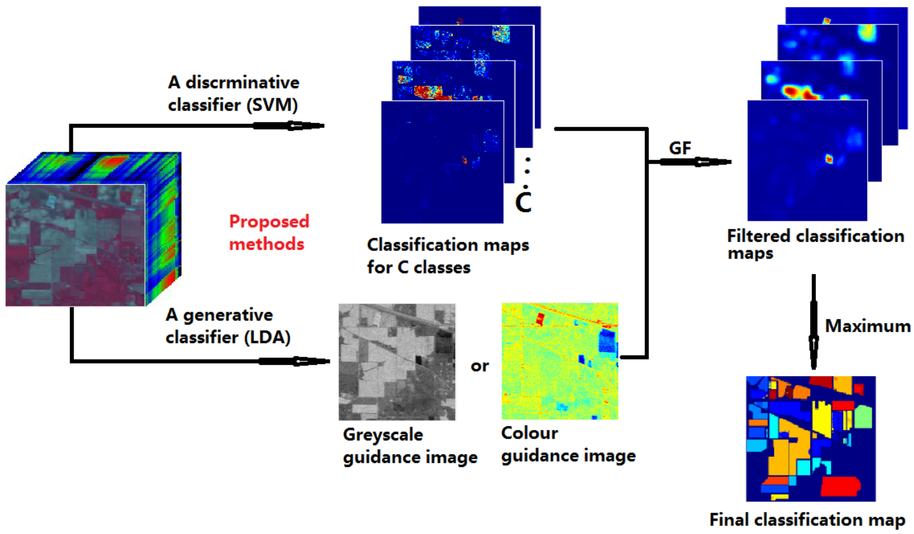


Figure 1: The proposed DGF methods for HSI classification.

Then to improve these initial spectrum-based classification results, the proposed DGF aims to incorporate spatial structure information, by using edge-preserving GF, and discriminative information, by using LDA to generate the guidance image.

Finally the C filtered classification maps are merged into a final classification map: The label of a test pixel $\mathbf{x}_{test}$ is

$$l(\mathbf{x}_{test}) = \operatorname*{argmax}_{j} f_j(\mathbf{x}_{test}), \text{ for } j = 1, \ldots, C, \tag{9}$$

where $f_j(\mathbf{x}_{test})$ is the filtered classification results of $\mathbf{x}_{test}$ for class $\mathcal{C}_j$, meaning that $\mathbf{x}_{test}$ is classified into class $\mathcal{C}_j$ if $f_j(\mathbf{x}_{test})$ has the highest value among all the C classes.

The DGF-based classification algorithm is in Algorithm 1.

Algorithm 1 Classification of an HSI based on **DGF**

Input:

A vectorised HSI $\mathbf{X} \in \mathbb{R}^{B\times N}$; training HSI pixels $\mathbf{X}^{train} \in \mathbb{R}^{B\times N_{train}}$ and their labels $\mathbf{y}^{train} \in \mathbb{R}^{N_{train}\times 1}$.

The radius of the local window size r and the smoothing parameter ϵ, for GF.

Output: A classification map $L(\mathbf{X})$, a $N \times 1$ vector.

Training phase:

- Train a classifier Φ (e.g. SVM) using $\left\{\mathbf{X}^{train}, \mathbf{y}^{train}\right\}$.
- Train a multi-class LDA model by (6).

Test phase:

- Classify $\mathbf{X}$ using the trained classifier Φ and obtain the initial classification maps $M_1, \ldots, M_C$.
- Obtain the DGF guidance image:
  - *either* $\mathbf{I}_g$ of $\mathbf{X}$ using DGF-g (7), *or*
  - $\mathbf{I}_c$ of $\mathbf{X}$ using DGF-c (8).
- Filter $M_1, \ldots, M_C$ by $\mathbf{I}_g$ as the greyscale guided filtering (1), or by $\mathbf{I}_c$ as the colour guided filtering (4).
- Classify each test pixel in $\mathbf{X}$ by (9) and obtain $L(\mathbf{X})$.

4. Experiments

4.1. Data and compared methods

The experiments are carried out on a real hyperspectral dataset: the AVIRIS Indian Pines dataset, which is publicly available [35] and has been widely used as a benchmark dataset for HSI classification including those involving GF [28, 36]. The dataset (Fig. 4(a)) consists of 145×145 pixels with 200 spectral bands after removing the water absorption bands. To make fair comparison with [28], for each of the 16 ground-truth classes (Fig. 4(b)), we randomly select the same number of labelled pixels as training samples and the rest as test samples which is utilised in [28], as listed in Table 1.

In our experiments, five methods are compared, including SVM, PGF-g [28], PGF-c [28], DGF-g and DGF-c. Among these methods, PGF-g represents the method using SVM as a spectral classifier and adopting the first principal component of the HSI as a virtual greyscale guidance image for GF (referred to as EPF-G-g in [28]). Similarly, PGF-c represents the method using the first three principal components as a virtual colour guidance image (EPF-G-c in [28]). Correspondingly, DGF-g and DGF-c represent our proposed methods in (7) and (8).

As with [28], the LIBSVM toolbox [37] is used to execute SVM. A dimension reduction toolbox (http://lvdmaaten.github.io/drtoolbox) is adopted to perform PCA and LDA. The Image Processing Toolbox of MATLAB is used to run GF. We also employ three standard performance measures: the overall accuracy (OA), the average accuracy (AA) and the κ coefficient [38] for evaluating the classification performances of the compared methods.

LDA has a limitation with the potential singularity of the pooled within-class scatter matrix $\mathbf{S}_W$ [39]. The singularity occurs when the data dimension B is larger than the number of training samples N_{train}. In our case, B is much smaller than N_{train}, and the singularity problem does not happen in our experiments. Nevertheless, we note that our approach is applicable when there are sufficient training samples. In the case of $B > N_{train}$, we can perform a preliminary dimension reduction before applying LDA to avoid the singularity issue.

4.2. Parameter settings

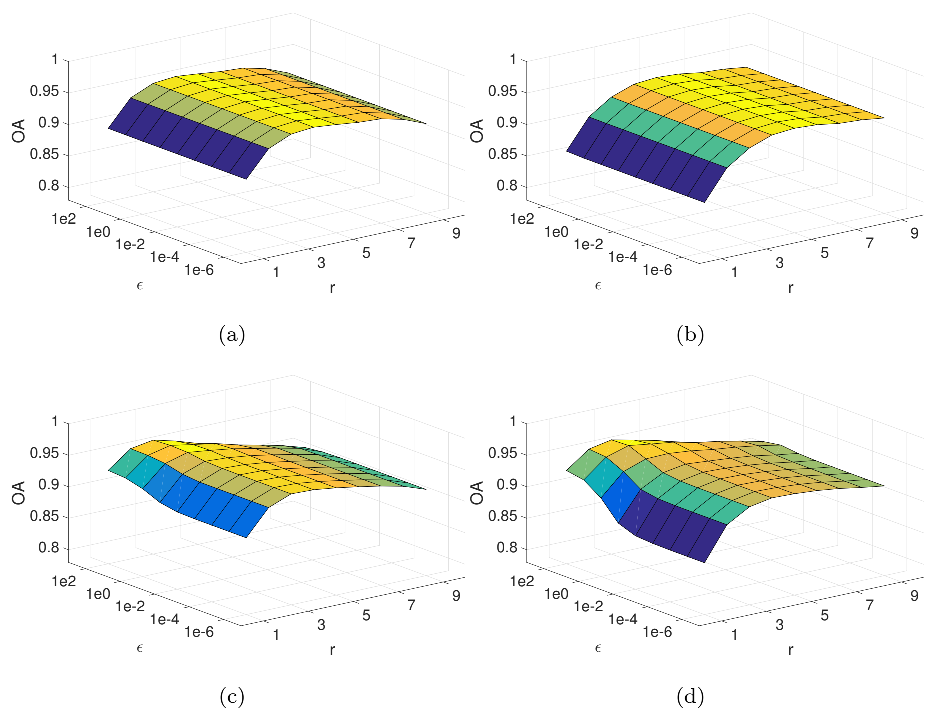


Figure 2: Overall classification accuracies over window radius r and smoothing parameter ϵ for (a) PGF-g [28], (b) PGF-c [28], (c) DGF-g, and (d) DGF-c.

There are parameters of SVM and GF. For SVM, we use the polynomial kernels with 20-fold cross validation to tune the parameters. The optimal values of parameters C and γ of the kernel function are tuned to be 5.66 and 0.16, respectively.

For GF, two parameters should be tuned: the radius r of the local window ω and the smoothing parameter ϵ. The influence of these two parameters on the overall classification performances of the compared methods are demonstrated in Fig. 2. The range of r is from 1 to 9 and the range of ϵ covers 10^{-6} to 10^{2}. We can observe that the optimal performances of PGF-g and PGF-c occur at about $r = 4$ (shown in Fig. 2(a)-2(b)). Hence we set $r = 4$ for PGF-g and PGF-c, and as with [28] we set $\epsilon = 0.01$ for them. For the proposed DGF-g and DGF-c, we set $r = 3$ and $\epsilon = 10$ as the optimal performances roughly occur there (shown in Fig.2(c)-2(d)).

4.3. Classification results

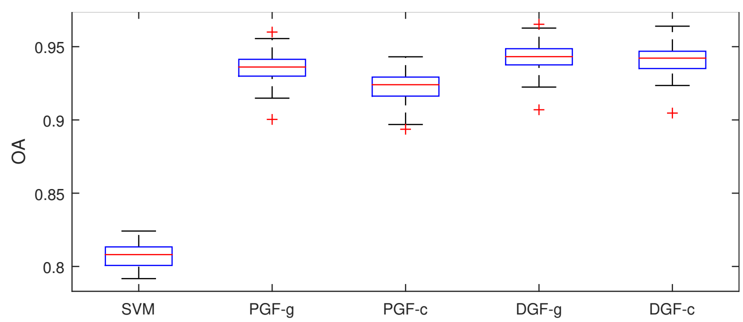


Figure 3: Boxplots of the overall classification accuracies on Indian Pines.

To have a reliable evaluation and fair comparison, we repeat the experiments under these parameter settings for 50 times through performing 50 random training-test splits while keeping the same numbers of samples for training and testing. For illustrative purposes, the classification results for one of the 50 experiments are given in Table 1 and depicted in Figs. 4(c)-4(g), respectively. Moreover, all of methods' overall classification accuracies are recorded and boxplotted in Fig. 3.

From these results, we can observe at least two clear patterns. Firstly, all PGF-g, PGF-c, DGF-g and DGF-c improve the performance of SVM substantially, which confirms that incorporating the guided filtering process can help the spectral-based classifier. Secondly, the proposed DGF-g and DGF-c outperform PGF-g and PGF-c in OA and κ. It indicates that the discriminative information provided by LDA (but unable by PCA) to GF can further improve classification performance.

Table 1: Indian Pines: Ground-truth label, training set, test set, and the classification accuracies (%) obtained by SVM, PGF-g [28], PGF-c [28], DGF-g and DGF-c. The best performance is in bold.

Class	Train	Test	SVM	PGF-g	PGF-c	DGF-g	DGF-c
1	25	21	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
2	83	1345	77.70	93.68	92.49	**95.46**	95.17
3	78	752	73.40	93.88	91.76	94.02	**94.68**
4	68	169	87.57	**100.00**	98.82	**100.00**	**100.00**
5	79	404	93.81	**98.27**	96.29	97.03	96.53
6	78	652	95.86	99.85	99.69	**100.00**	**100.00**
7	4	24	**4.17**	0.00	0.00	0.00	0.00
8	66	412	99.51	**100.00**	**100.00**	**100.00**	**100.00**
9	2	18	0.00	0.00	0.00	0.00	0.00
10	81	891	74.86	97.87	97.64	**98.77**	**98.77**
11	99	2356	66.47	92.19	90.32	**94.27**	94.10
12	73	520	90.00	**100.00**	99.23	**100.00**	99.81
13	70	135	99.26	99.26	**100.00**	**100.00**	99.26
14	90	1175	90.21	**98.13**	97.62	97.45	97.62
15	65	321	79.44	**99.07**	94.39	**99.07**	**99.07**
16	46	47	95.74	**97.87**	93.62	87.23	93.62
Total	1007	9242					
OA			79.81	95.55	94.31	**96.27**	96.25
AA			76.75	**85.63**	84.49	85.21	85.54
κ			0.770	0.949	0.945	**0.957**	0.957

Table 2: Overall classification accuracies (%). Methods with $*$ indicates that their OAs are obtained from [28] under their optimal parameters settings via 5-fold cross-validation.

WLS*	NC*	JBF-g*	PGF-g	DGF-g	JBF-c*	PGF-c	DGF-c
94.93	95.20	95.42	95.55	96.27	95.41	94.31	96.25

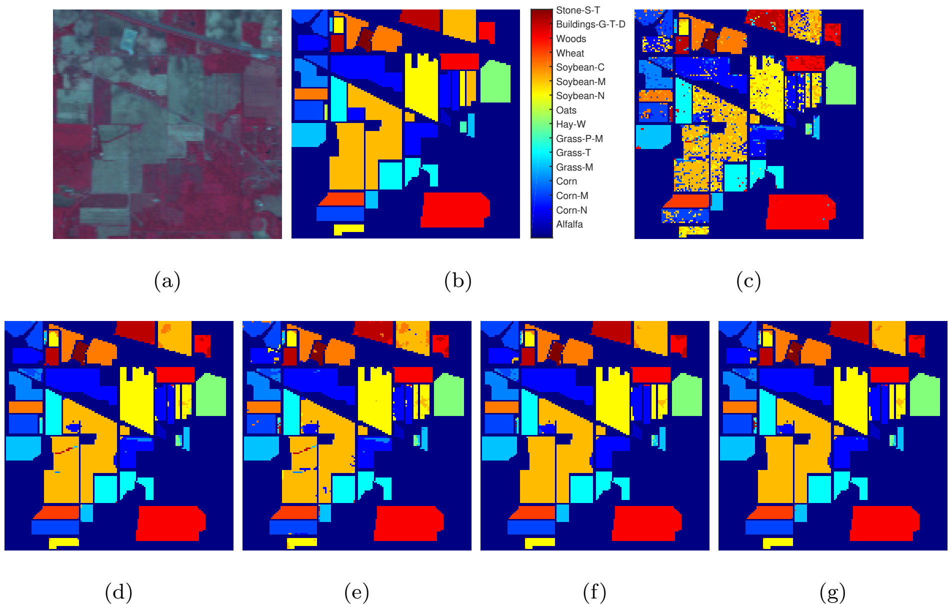


Figure 4: Indian Pines: (a) mean image shown in the false colour; (b) ground-truth labels. Classification maps (and OA) of (c) SVM (79.81%); (d) PGF-g (95.55%); (e) PGF-c (94.31%); (f) DGF-g (96.27%); (g) DGF-c (96.25%).

For further assess whether the difference of the performances of DGF and PGF is statistical significant, we also perform the Wilcoxon signed-rank test, a widely-used non-parametric statistical hypothesis test for paired samples. In this paper, the test is designed for testing whether the paired classification performances, i.e. 50 OAs (or 50 AAs) of PGF and that of DGF, differs at the 1% significance level. Specifically, for OAs (or AAs), we conduct two tests: one for PGF-g versus DGF-g and the other for PGF-c versus DGF-c. The obtained p-values of the two tests of OAs are 8.0e-10 and 7.5e-10; this indicates strong evidence that DGF is significantly better than PGF in terms of OA, confirming the observation from Fig. 3 and Table 1. The corresponding p-values for AAs are 0.011 and 7.6e-10; this provides strong evidence that DGF-c is significantly better than PGF-c in terms of AA, while no strong evidence that DGF-g is significantly worse than PGF-g, also in line with what is revealed by Table 1.

From Table 1, we also note that all GF-based methods (and SVM itself) fail for classes 7 and 9. This is due to the lack of training samples for these two class, which are only 4 for class 7 and 2 for class 9. Also, classes 7 and 9 cover a narrow region in the dataset, and the filtering of these two classes can be dominated by other classes adjacent and thus misclassified. The bad performance of the methods in identifying classes 7 and 9 make a big influence on AA but little influence on OA, because the number of test pixels of these two class are also small. This explains why OA are higher than AA for all the compared methods listed in Table 1.

For further evaluation, we also compare DGF with some other modern edge-preserving smoothing methods, such as the weighted least squares (WLS) method, the normalised convolution (NC) filter and the joint bilateral filter (JBF). These methods are also compared in [28]. Since our experiment settings are exactly the same as [28], we compare the performance of our proposed DGF directly with those reported in [28], as listed in Table 2. We can observe that our proposed method DGF-g and DGF-c still outperform the other compared edge-preserving methods, which shows the superiority of the proposed idea to combine the discriminative classifier and the generative classifier by GF for the

HSI classification.

5. Summary and future work

In this paper, a new filtering framework called discriminatively guided filtering (DGF) has been proposed, which integrates a discriminative classifier and a generative classifier by the guided filtering for hyperspectral image classification. The combination of SVM and LDA has been adopted illustrating the effectiveness of DGF, which also inspires us to investigate the performance of other generative-discriminative combinations as a direction of our future research.

Furthermore, there are other reports using the Indian Pines dataset for evaluation of classification performance, for example, our recent work called JSM-DKSVD which focuses on dictionary learning [10]. We shall note that it may not be fair to directly compare the proposed DGF to such a dictionary learning-based classification method, since the latter has an extra learning process to learn the dictionary. Nevertheless, we believe that the proposed GDF framework can be improved by incorporating such a learning process, which leads to another direction meriting our future research.

Acknowledgment

This work was partly supported by University College London's Security Science Doctoral Training Centre under Engineering and Physical Sciences Research Council (EPSRC) grant EP/G037264/1, by the Royal Society under Royal Society-Newton Mobility Grant IE161194, and by the National Natural Science Foundation of China under Grant 61711530239.

References

[1] G. Camps-Valls, D. Tuia, L. Bruzzone, J. A. Benediktsson, Advances in hyperspectral image classification: Earth monitoring with statistical learning methods, IEEE Signal Processing Magazine 31 (1) (2014) 45–54.

[2] F. Melgani, L. Bruzzone, Classification of hyperspectral remote sensing images with support vector machines, IEEE Transactions on Geoscience and Remote Sensing 42 (8) (2004) 1778–1790.

[3] X. Guo, X. Huang, L. Zhang, Three-dimensional wavelet texture feature extraction and classification for multi/hyperspectral imagery, IEEE Geoscience and Remote Sensing Letters 11 (12) (2014) 2183–2187.

[4] B. Du, L. Zhang, L. Zhang, T. Chen, K. Wu, A discriminative manifold learning based dimension reduction method for hyperspectral classification, International Journal of Fuzzy Systems 14 (2) (2012) 272–277.

[5] D. Lunga, S. Prasad, M. M. Crawford, O. Ersoy, Manifold-learning-based feature extraction for classification of hyperspectral data: a review of advances in manifold learning, IEEE Signal Processing Magazine 31 (1) (2014) 55–66.

[6] L. Zhang, X. Zhu, L. Zhang, B. Du, Multidomain subspace classification for hyperspectral images, IEEE Transactions on Geoscience and Remote Sensing 54 (10) (2016) 6138–6150.

[7] F. Luo, H. Huang, Z. Ma, J. Liu, Semisupervised sparse manifold discriminative analysis for feature extraction of hyperspectral images, IEEE Transactions on Geoscience and Remote Sensing 54 (10) (2016) 6197–6211.

[8] Y. Chen, N. M. Nasrabadi, T. D. Tran, Hyperspectral image classification using dictionary-based sparse representation, IEEE Transactions on Geoscience and Remote Sensing 49 (10) (2011) 3973–3985.

[9] P. Du, Z. Xue, J. Li, A. Plaza, Learning discriminative sparse representations for hyperspectral image classification, IEEE Journal of Selected Topics in Signal Processing 9 (6) (2015) 1089–1104.

[10] Z. Wang, J. Liu, J.-H. Xue, Joint sparse model-based discriminative K-SVD for hyperspectral image classification, Signal Processing 133 (2017) 144–155.

[11] J. Li, H. Zhang, Y. Huang, L. Zhang, Hyperspectral image classification by nonlocal joint collaborative representation with a locally adaptive dictionary, IEEE Transactions on Geoscience and Remote Sensing 52 (6) (2014) 3707–3719.

[12] Q. Shi, B. Du, L. Zhang, Spatial coherence-based batch-mode active learning for remote sensing image classification, IEEE Transactions on Image Processing 24 (7) (2015) 2037–2050.

[13] S. Sun, P. Zhong, H. Xiao, R. Wang, An MRF model-based active learning framework for the spectral-spatial classification of hyperspectral imagery, IEEE Journal of Selected Topics in Signal Processing 9 (6) (2015) 1074–1088.

[14] Z. Wang, B. Du, L. Zhang, L. Zhang, A batch-mode active learning framework by querying discriminative and representative samples for hyperspectral image classification, Neurocomputing 179 (2016) 88–100.

[15] J. Li, H. Zhang, L. Zhang, X. Huang, L. Zhang, Joint collaborative representation with multitask learning for hyperspectral image classification, IEEE Transactions on Geoscience and Remote Sensing 52 (9) (2014) 5923–5936.

[16] Q. Shi, B. Du, L. Zhang, Domain adaptation for remote sensing image classification: A low-rank reconstruction and instance weighting label propagation inspired algorithm, IEEE Transactions on Geoscience and Remote Sensing 53 (10) (2015) 5677–5689.

[17] Y. Zhong, B. Zhao, L. Zhang, Multiagent object-based classifier for high spatial resolution imagery, IEEE Transactions on Geoscience and Remote Sensing 52 (2) (2014) 841–857.

[18] L. Zhang, L. Zhang, B. Du, Deep learning for remote sensing data: A technical tutorial on the state of the art, IEEE Geoscience and Remote Sensing Magazine 4 (2) (2016) 22–40.

[19] Y. Tarabalka, M. Fauvel, J. Chanussot, J. A. Benediktsson, SVM-and MRF-based method for accurate classification of hyperspectral images, IEEE Geoscience and Remote Sensing Letters 7 (4) (2010) 736–740.

[20] Y. Zhao, J. Yang, J. C.-W. Chan, Hyperspectral imagery super-resolution by spatial–spectral joint nonlocal similarity, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7 (6) (2014) 2671–2679.

[21] T. Lu, S. Li, L. Fang, Y. Ma, J. A. Benediktsson, Spectral–spatial adaptive sparse representation for hyperspectral image denoising, IEEE Transactions on Geoscience and Remote Sensing 54 (1) (2016) 373–385.

[22] B. Du, R. Zhao, L. Zhang, L. Zhang, A spectral-spatial based local summation anomaly detection method for hyperspectral images, Signal Processing 124 (2016) 115–131.

[23] L. Zhang, L. Zhang, D. Tao, X. Huang, B. Du, Compression of hyperspectral remote sensing images by tensor approach, Neurocomputing 147 (2015) 358–363.

[24] F. Luo, H. Huang, J. Liu, Z. Ma, Fusion of graph embedding and sparse representation for feature extraction and classification of hyperspectral imagery, Photogrammetric Engineering & Remote Sensing 83 (1) (2017) 37–46.

[25] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, J. C. Tilton, Advances in spectral-spatial classification of hyperspectral images, Proceedings of the IEEE 101 (3) (2013) 566–569.

[26] X. Huang, L. Zhang, An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery, IEEE Transactions on Geoscience and Remote Sensing 51 (1) (2013) 257–272.

[27] K. He, J. Sun, X. Tang, Guided image filtering, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (6) (2013) 1397–1409.

[28] X. Kang, S. Li, J. A. Benediktsson, Spectral–spatial hyperspectral image classification with edge-preserving filtering, IEEE Transactions on Geoscience and Remote Sensing 52 (5) (2014) 2666–2677.

[29] A. Y. Ng, M. I. Jordan, On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes, in: Advances in Neural Information Processing Systems, 2002, pp. 841–848.

[30] R. Raina, Y. Shen, A. McCallum, A. Y. Ng, Classification with hybrid generative/discriminative models, in: Advances in Neural Information Processing Systems, 2004, pp. 545–552.

[31] J.-H. Xue, D. M. Titterington, Comment on on discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes, Neural Processing Letters 28 (3) (2008) 169–187.

[32] J.-H. Xue, D. M. Titterington, On the generative–discriminative tradeoff approach: Interpretation, asymptotic efficiency and classification performance, Computational Statistics & Data Analysis 54 (2) (2010) 438–451.

[33] J.-H. Xue, D. M. Titterington, Joint discriminative–generative modelling based on statistical tests for classification, Pattern Recognition Letters 31 (9) (2010) 1048–1055.

[34] C. M. Bishop, Pattern Recognition and Machine Learning, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[35] P. R. Foundation, A Freeware Multispectral Image Data Analysis System, `https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html`, [Online; accessed 22-July-2014] (2014).

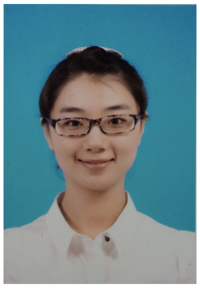
[36] J. Xia, L. Bombrun, T. Adalı, Y. Berthoumieu, C. Germain, Spectral–spatial classification of hyperspectral images using ICA and edge-preserving

filter via an ensemble strategy, IEEE Transactions on Geoscience and Remote Sensing 54 (8) (2016) 4971–4982.

[37] C.-C. Chang, C.-J. Lin, LIBSVM: A library for support vector machines, ACM Transactions on Intelligent Systems and Technology 2 (2011) 27:1–27:27, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[38] J. A. Richards, X. Jia, Remote Sensing Digital Image Analysis: An Introduction, New York: Springer-Verlag, 2006.

[39] X. Jia, B.-C. Kuo, M. M. Crawford, Feature mining for hyperspectral image classification, Proceedings of the IEEE 101 (3) (2013) 676–697.



Ziyu Wang received the M.Sc. degree in statistics in 2012, the M.Res. degree in security science in 2013 and the Ph.D. degree in security science and statistical science in 2017, all from University College London. Her research interests include hyperspectral image analysis, sparse representation and statistical classification.

Huafeng Hu received the B.S. degree in applied mathematics in 2015 from Xi'an Jiaotong-Liverpool University and the M.Sc. degree in statistics in 2016 from University College London. Her research interests include hyperspectral image classification and convolutional neural network for image classification.

Lefei Zhang received the B.S. and Ph.D. degrees from Wuhan University in 2008 and 2013, respectively. He is an Associate Professor with the School of Computer, Wuhan University, and also a Hong Kong Scholar with the Department of Computing, The Hong Kong Polytechnic University. His research interests include pattern recognition, image processing and remote sensing.

Jing-Hao Xue received the Dr.Eng. degree in signal and information processing from Tsinghua University in 1998 and the Ph.D. degree in statistics from the University of Glasgow in 2008. He is a Senior Lecturer in the Department of Statistical Science, University College London. His research interests include statistical classification, high-dimensional data analysis, pattern recognition and image analysis.