

IFGF Based Feature Extraction of Hyperspectral Images

Smitha K S¹, Saranya Sasidharan², Minu Thomas³

Federal Institute of Science and Technology (FISAT), Department of Computer Science

Abstract: *Hyperspectral sensors collect information as a set of images represented by different bands. Hyperspectral images are three-dimensional images with sometimes over 100 bands where as regular images have only three bands: red, green and blue. Each pixel has a hyperspectral signature that represents different materials. Hyperspectral images can be used for geology, forestry and agriculture mapping, land cover analysis and atmospheric analysis. Even though hyperspectral images can sometimes contain over 100 bands, relatively few bands can explain the vast majority of the information. Hyperspectral images contain rich and fine spectral information, an improvement of land use/cover classification accuracy is expected from the use of such images. However, the classification methods that have been successfully applied to multispectral data in the past are not as effective as to hyperspectral data. The major cause is that the size of training data set does not correspond to the increase of dimensionality of hyperspectral data. Actually, the problem of the "curse of dimensionality" emerges when a statistic based classification method is applied to the hyperspectral data. For such reason, hyperspectral images are mapped into a lower dimension while preserving the main features of the original data by a process called dimensional reduction. This can be done by feature extraction that a small number of salient features are extracted from the hyperspectral data when confronted with a limited set of training samples.*

Keywords: Hyperspectral images, Multispectral images, feature extraction.

1. Introduction

Remote sensing can be defined as collection and interpretation of information about an object, area or event without any physical contact with the object. Aircraft and satellites are the common platforms for remote sensing of earth and its natural resources. Aerial photography in visible portion of the electromagnetic wavelength was the original form of remote sensing but technological developments has enabled the acquisition of information at other wavelength including near infrared, thermal infrared and microwave. Collection of information over a large numbers of wavelength bands is Remote sensing image acquired by multispectral sensors such as the widely used Landsat Thematic Mapper (TM) sensor, have shown their usefulness in numerous earth observation (EO) applications. In general, the relatively small number of acquisition channels that characterizes multispectral sensors may be sufficient to discriminate among different land-cover classes (e.g., forestry, water, crops, urban areas, etc.). However, their discrimination capability is very limited when different types (or conditions) of the same species (e.g., different types of forest) are to be recognized. Hyperspectral sensors can be used to deal with this problem. These sensors are characterized by a very high spectral resolution that usually results in hundreds of observation channels. Thanks to these channels, it is possible to address various additional applications requiring very high discrimination capabilities in the spectral domain (including material quantification and target detection).

Hyperspectral remote sensors collect image data simultaneously in dozens or hundreds of narrow, adjacent spectral bands. These measurements make it possible to derive a continuous spectrum for each image cell. However, developing efficient methods to process hyperspectral images with more than 100 channels is a difficult objective. From a

methodological viewpoint, the automatic analysis of hyperspectral data is not a trivial task. In particular, it is made complex by many factors, such as: 1) the large spatial variability of the hyperspectral signature of each land-cover class; 2) atmospheric effects; and 3) the curse of dimensionality. In the context of supervised classification, one of the main difficulties is related to the small ratio between the number of available training samples and the number of features. This makes it impossible to obtain reasonable estimates of the class-conditional hyperdimensional probability density functions used in standard statistical classifiers. As a consequence, on increasing the number of features given as input to the classifier over a given threshold (which depends on the number of training samples and the kind of classifier adopted), the classification accuracy decreases. When performing supervised classification, it is important that the number of training/ground truth points scales with the number of bands. This is a presumption which is often hard to attain when using hyperspectral data. This phenomenon is known as the curse of dimensionality or Hughes effect.

Secondly, neighboring bands in hyperspectral data are generally strongly correlated. As a result, it is possible that no information is added by increasing the spectral resolution, as if the measurement of several neighboring bands is just a repeat. Both observations indicate that a high number of bands are generally not needed. At first sight, these comments seem to indicate that multispectral data is sufficient for most applications. This is in fact not true, since the required spectral bands for classification purposes should be specifically adapted to the problem at hand, and rarely correspond to the band settings of any multispectral sensor. Rather, a reduction of the dimensionality of hyperspectral data is pursued. Reduction of dimensionality can be achieved by making a selection of few existing bands (feature selection) or generate new features by taking (linear)

combinations of the bands (feature extraction). Sometimes the feature selection step is preceded by changing the representation (e.g. derivative or wavelet transform). In some applications domain knowledge can be used to define new features. A large number of feature extraction, feature reduction, and combination techniques have been proposed to address the high-dimensionality problem.

Recently, edge-preserving filtering has been applied successfully in many applications such as high dynamic imaging, stereo matching, image fusion (IF), dehazing, and denoising since it can smooth an image while preserving well, its edge structures. The transform domain recursive filter, guided filter etc can also be applied for the feature extraction of hyperspectral images.

The rest of the paper is organized as follows. In section 2, related-works of fusion methods is briefly described. Section 3 describes the framework of this method. Experimental results of proposed approach and its comparison with state-of-the-art is described in section 4, and finally, section 5 summarizes the conclusion of this paper.

2. Literature Survey

Due to advances in sensor technology, it is now possible to acquire hyperspectral data simultaneously in hundreds of bands. Algorithms that both reduce the dimensionality of the data sets and handle highly correlated bands are required to exploit the information in these data sets effectively. Classical techniques for features reduction that can be applied to the measured hyperspectral signatures mainly include features selection algorithms and features extraction algorithms. As compared to features extraction, features selection is a simpler and direct approach, and the resulting reduced set of features is easier to interpret. On the other hand, extraction methods can be expected to be more effective.

Feature selection for classification of hyperspectral data by using Support vector machine has been performed by Pal and Foody [1]. The study had principally focused on feature selection method by using SVM on the hyperspectral datasets. An attempt was made to address the key aspect of uncertainty over the sensitivity of the SVM and accuracy of classification of dataset to the dimensionality of the dataset. It was noticed that the accuracy of the SVM classification varied as a function of the number of features used and the size of the training set used. As the number of features were increased the accuracy of the SVM classification also increased. When a fixed size of training set were used the accuracy had initially rose when features to the peak were added but thereafter decreased with the addition of more features. However, the decrease in the accuracy was significant statistically. When small training sets were used, the curse of dimensionality reduction and the Hughes effect were observed with SVM classification. Finally, a conclusion was made that when larger training sets are used mostly the effect of the Hughes phenomenon could be reverted. Also as the features increases accuracy of the classification will be reduced. Burgers et al., [2] have performed a comparative

analysis of dimensionality reduction techniques aiming to evaluate the performance of the dimensionality reduction algorithms. Nonlinear methods had given comparably better results but had a major setback of taking very long runtimes. Thus increasing the cost of the processes run. When the high dimensional data sets were used, their runtimes was very high compared to linear methods resulting in increase of the computational cost. PCA had the least error rates in the processes and outperformed in all the tasks compared to other methods. A similar kind of comparative work has also been performed by Fong [3], where, different dimensionality reduction techniques like Principal Component Analysis, Fast ICA (Independent component Analysis), Laplacian Eigenmaps, Local Linear Embedding (LLE), Local Tangent Space Analysis (LTSA), Linear Local Tangent Space Analysis (LLTSA) and diffusion maps are compared for their performances.

A new feature weighting method for band selection[4] is presented by Rui Huang and Mingyi He in the paper Band Selection Based on Feature Weighting for Classification of Hyperspectral Data. Extraction of Spectral Channels from Hyperspectral Images for Classification Purposes [5] is a feature selection method proposed by Sebastiano B. Serpico and Gabriele Moser proposes a procedure to extract spectral channels of variable bandwidths and spectral positions from the hyperspectral image in such a way as to optimize the accuracy for a specific classification problem. It represents a special case of feature extraction that is expected to be more powerful than feature selection. The kind of transformation used allows the interpretability of the new features (i.e., the spectral bands) to be saved.

Principal Component Analysis for Hyperspectral Image Classification [6] proposed by Craig Rodarmel and Jie Shan is a preprocessing technique for the classification of hyperspectral images. The principal component analysis is based on the fact that neighboring bands of hyperspectral images are highly correlated and often convey almost the same information about the object. Classifications using the most significant PCA bands yield the same classification results as when entire hyperspectral data sets are used. Hyperspectral Image Classification With Independent Component Discriminant Analysis,[7] proposed by Alberto Villa, Jn Atli Benediktsson, Jocelyn Chanussot and Christian Jutten, uses the Independent Component (IC) Discriminant Analysis (ICDA) for remote sensing classification. It is a nonparametric method for discriminant analysis based on the application of a Bayesian classification rule on a signal composed by independent components (ICs) for the classification of hyperspectral images. ICA-based methods ensure that the transformed components are as independent as possible. As the components are more independent the misclassification rate decreases. But the computational burden can be increased due to the complexity of the ICA-based methods. It just processes each pixel independently without considering the spatial information.

Classification and Feature Extraction for Remote Sensing Images From Urban Areas Based on Morphological Transformations [8] proposed by Jon Atli Benediktsson, Martino Pesaresi, and Kolbeinn Arnason

investigates classification of panchromatic high-resolution data from urban areas using morphological and neural approaches. This approach is applied in experiments on high-resolution Indian Remote Sensing 1C (IRS-1C) and IKONOS remote sensing data from urban areas. It requires relatively few features are needed to achieve the same classification accuracies as in the original feature space. Classification of Hyperspectral Data From Urban Areas Based on Extended Morphological Proles[9] proposed by Jn Atli Benediktsson, Jn Aevor Palmason and Johannes R. Sveinsson presents a method based on mathematical morphology for preprocessing of the hyperspectral data. In certain applications, however, the integration of high spatial and spectral is mandatory to achieve sufficiently accurate mapping and/or detection results. For instance, urban area mapping requires sufficient spatial resolution to distinguish small spectral classes, such as trees in a park, or cars on a street. So we need to manage very high-dimensional data volumes in which the spatial correlation between spectral responses of neighboring pixels can be potentially high. As a result, there is a need to incorporate the spatial arrangement of the data in the development of robust analysis techniques. The performance is excellent because it mostly uses spatial information instead of the rich spectral information available in the hyperspectral data.

SpectralSpatial Hyperspectral Image Classification With Edge-Preserving Filtering,[10]proposed by Xudong kang, Shutao Li, and Jon Atli Benediktsson is a hyperspectral image classification method. It is novel spectral-spatial classification framework based on edge-preserving filtering. Feature Extraction of Hyperspectral Images with Image Fusion and Recursive Filtering [11] is a method proposed by Xudong Kang, Shutao Li, and Jn Atli Benediktsson. First, the hyperspectral image is partitioned into multiple subsets of adjacent hyperspectral bands. Then, the bands in each subset are fused together by averaging, which is one of the simplest image fusion methods. Finally, the fused bands are processed with transform domain recursive filtering to get the resulting features for classification. It combines the spectral spatial information for feature extraction.

In this paper we proposes a new feature extraction method for hyperspectral images by combining the image fusion and one of the most powerful edge preserving smoothing filter known as, guided filter[12].

3. Proposed Method

Feature extraction can be viewed as finding a set of vectors that represents an observation while reducing the dimensionality. In pattern recognition, it is desirable to extract features that are focused on discriminating between classes. Although a reduction in dimensionality is desirable, the error increment due to the reduction in dimension has to be without sacrificing the discriminative power of classifiers. The methods so far used for feature extraction found to be having certain common deficiencies. Some of them were using only spectral information available and merely avoiding the spatial information that is required for a better classification and vice versa.

So a new feature extraction methodology is required that make use of the spectral information, spatial information and at the same time extract the useful features with less time complexity [21].The basic method is based on two simple assumptions:

1. The adjacent bands of the hyperspectral image usually contain redundant information.
2. The neighboring pixels usually have quite strong correlations with each other.

The following steps explain the basic structure of the proposed method.

1. Partition the hyperspectral image into multiple subsets of adjacent bands.
2. Fuse the adjacent bands in each subset.
3. Perform guided filtering on the fused bands.
4. Perform classification on the filtered images.

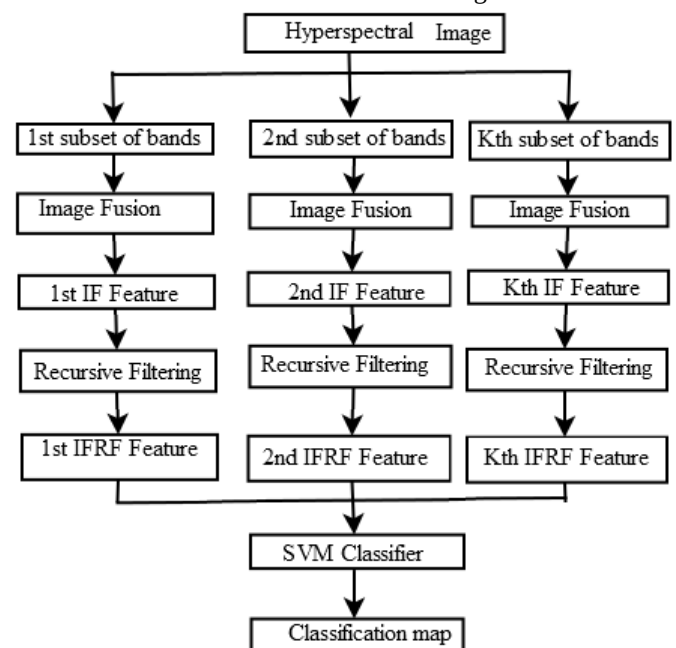


Figure 1: Schematic of the IFGF-based classification method

3.1 Band partitioning

The hyperspectral image is partitioned into K subsets of hyperspectral bands. The kth subset can be obtained as follows:

$$p^k = \begin{cases} (x_k, \dots, x_{k+\lfloor \frac{D}{K} \rfloor}), & \text{if } k + \frac{D}{K} \leq D \\ (x_k, \dots, x_D), & \text{if } k + \frac{D}{K} > D \end{cases} \quad (1)$$

Where $X = (X_1, \dots, X_D) \in \mathbb{R}^{D \times J}$ denotes the original hyperspectral image with D-dimensional feature vectors and J pixels, and $\lfloor D/K \rfloor$ represents the floor operation which calculates the largest integer not greater than D/K.

3.2 Image Fusion

The adjacent bands in the kth subset are fused by any image fusion method. Here one of the most simple fusion methods,

i.e., the averaging fusion method is used. Specifically, the k th fused band, i.e., the k th IF feature F_k , is calculated as follows:

$$Q^k = \frac{\sum_{n=1}^{N_k} P_n^k}{N_k} \quad (2)$$

Where P_n^k refers to the n th band of the k th subset of hyperspectral bands and N_k refers to the total number of bands in the k th subset. This step actually calculates the average image of each subset which aims at removing the noisy pixels and the redundant information for each subset.

3.3 Guided Filtering

Guided filtering [12] is performed on each fused band to obtain the k th feature

$$Q_i^k = G, F_{r,\varepsilon}(Q_i^k) \quad (3)$$

Where GF represents the guided filtering operation, r and ε are the parameters of the guided filter and $Q = (Q^1, \dots, Q^k \in R^{k \times J})$ is the resulting feature image obtained by IFGF.

3.4 Classification

The SVM classifier is utilized for the classification of the IFRF features. The SVM classifier is one of the most widely used pixel wise classifiers and has, in particular, shown a good performance in terms of classification accuracy. Furthermore, the SVM classifier has a major advantage, i.e., robust to the dimension of data sets.

4. Experimental Results

The performance of proposed IFGF based feature extraction method is compared with a feature extraction method that makes use of a recursive filter along with image fusion. The SVM classifier [15] is used to classify the hyperspectral images.

4.1 Experimental Setup

1)Data Sets: Three hyperspectral data sets, i.e., the Indian Pines image, the Salinas image, and the University of Pavia image, are utilized in our experiments. The Indian Pines has 220 bands of size 145 X 145, with a spatial resolution of 20 m per pixel and a spectral coverage ranging from 0.4 to 2.5 μ m. The Salinas image was captured by the AVIRIS sensor over Salinas Valley, CA, USA. The image has 224 bands and is of size 512 X 217, with a spatial resolution of 3.7 m per pixel. The University of Pavia image is from an urban area surrounding the University of Pavia, Pavia, Italy. It was recorded by Reflective Optics System Imaging Spectrometer with a spatial resolution of 1.3 m per pixel and a spectral coverage ranging from 0.43 to 0.86 μ m. The image has 115 bands of size 610 $\times$ 340.

2)Quality Indexes: In order to evaluate the performance of the proposed method, three widely used quality indexes, i.e., overall accuracy (OA), average accuracy (AA), and kappa coefficient, are adopted. Among the three metrics, OA

measures the percentage of pixels that are correctly classified. AA refers to the mean of the percentage of correctly classified pixels for each class. In order to make the measurement more objective, the kappa coefficient estimates the percentage of correctly classified pixels corrected by the number of agreements that would be expected purely by chance.

The classification results obtained by using IFRF method and IFGF method in Indian Pines image, are shown below :

Table 1: Indian Pines Image Classification Analysis on OA, AA, KAPPA

	IFRF	IFGF
OA	.9719	.9700
AA	.9702	.9653
KAPPA	.9700	.9654

The diagrammatic representations of the results are given below:



Figure 2: classification accuracies of IFRF and IFGF methods

4.2 Classification Results

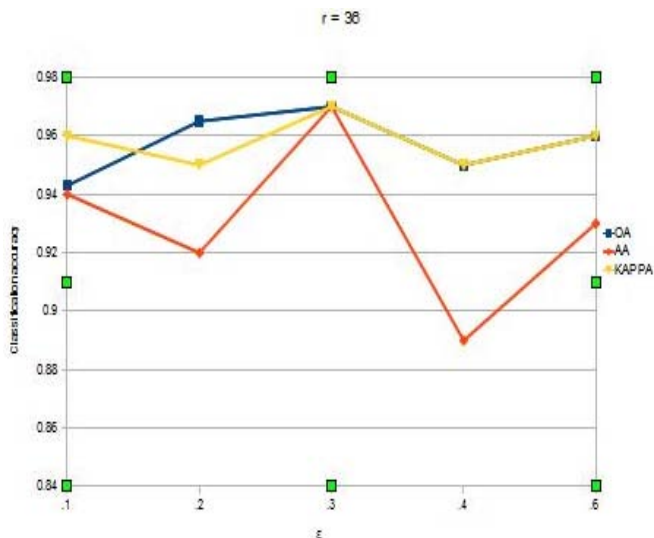
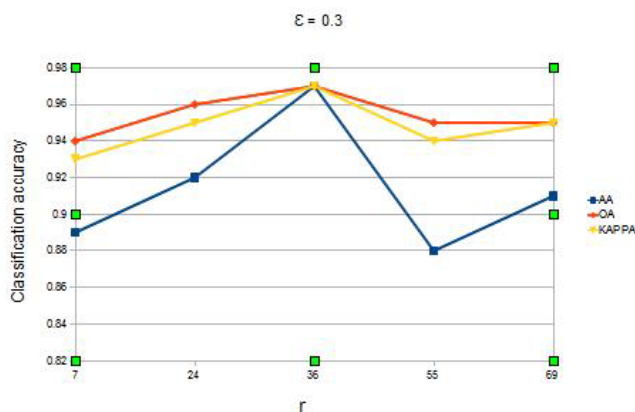
1) Analysis of the influence of parameters: The influence of the two parameters(r and ε) on the classification performance is analyzed in an experiment on the Indian Pines image. The training set which accounts for 10 of the ground truth was chosen randomly. The OA, AA, and kappa of the proposed method are measured with different parameter settings. When the influence of r is analyzed, ε fixed at .3. Similarly ε is analyzed in the same way with r fixed at 36. It was seen that when r is very small the classification accuracy was less, and the result was same when the value of r is very large. So in this IFGF methodology the value of r is taken as 36 to obtain a better result. And for this values of r the value of ε is taken as .3. These are the default values taken to obtain a better classification accuracy and time efficiency. The data values are shown in table II and III and the diagrammatic results are shown in figure 3 and figure 4.

Table 2: Classification accuracies when $r=36$

r	AA	OA	KAPPA
.1	.9478	.9661	.9610
.2	.9287	.9650	.9597
.3	.9691	.9695	.9650
.4	.8941	.9590	.9529
.6	.9360	.9569	.9620

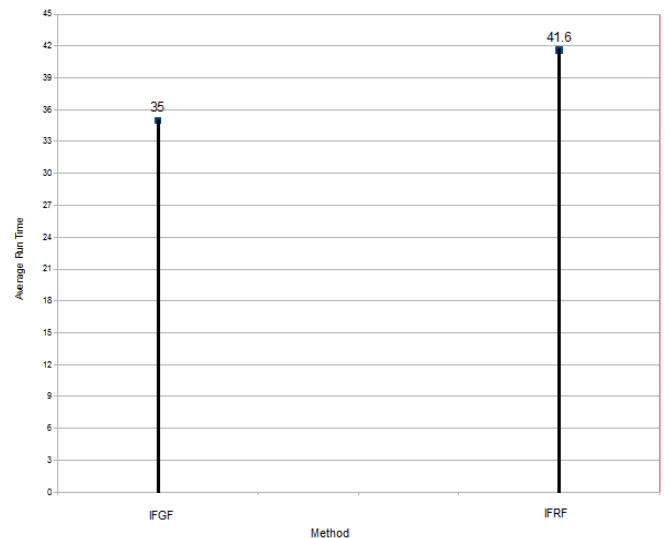
Table 3: Classification accuracies when $\varepsilon=0.3$

ε	AA	OA	KAPPA
7	.8925	.9436	.9353
24	.9225	.9627	.9527
36	.9691	.9695	.9650
55	.8874	.9502	.9425
69	.9197	.9572	.9508

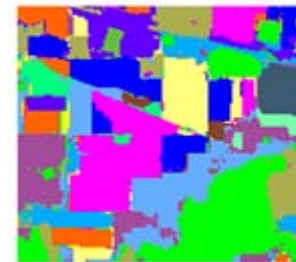
**Figure 3:** Classification accuracies when $r=36$ **Figure 4:** Classification accuracies when $\varepsilon=0.3$

2) Comparison with classification methods: Comparing to IFGF method IFRF performs well in terms of accuracy. But IFGF method provides almost same efficiency as IFRF method. For hyperspectral image classification more than accuracy time efficiency is an important concern. Comparing with IFRF method, IFGF method performs classification more faster. The average values of run times are shown below:

Runtime comparison

**Figure 5:** Runtime comparison of classification methods

The resulting classified image using the IFGF and IFRF methods are shown below.

**Figure 6:** Result of the IFRF-based classification method**Figure 7:** Result of the IFGF-based classification method

By considering the overall performance of IFGF method in terms of accuracy and run time efficiency it is clear that in cases where time is of first concern this method is more apt than in cases where accuracy is important. IFGF method provides a better improvement in run time and also provides better accuracy values in the context of classification of hyperspectral images.

5. Conclusion

Recently, edge-preserving filtering has been applied successfully in many applications such as high dynamic imaging, stereo matching, image fusion (IF), dehazing, and denoising since it can smooth an image while preserving well its edge structures. In some previous work, joint edge-

preserving filtering has been successfully applied for the post processing of support vector machine (SVM) classification.

The proposed approach is based on the application of IF to reduce the dimension of the data, the use of recursive filtering to combine spatial information into the resulting IFRF features. Experiments have been carried out on three different real hyperspectral images. The results of the experiments showed the effectiveness of the proposed method, which provided better results than those of the widely used pixel wise classifiers and the spectral spatial classifiers. Moreover, the proposed method has presented several other advantages: 1) the feature can well preserve the physical meaning of the hyperspectral data. In other words, the pixel values in the feature image still reflect the spectral response of a pixel in a specific spectral range; 2) it is time efficient since it is based on a very fast EPF; and 3) although the classification accuracy obtained by the IFGF is influenced by the parameters of the guided filter, these choices are not critical. The reason is that there is a large region around the optimal number of features for which the proposed method has similar results and outperforms other classification methods in terms of accuracy.

References

- [1] M. Pal and G. M. Foody, "Feature selection for classification of hyperspectral data by SVM," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 22972307, May 2010.
- [2] Pavan Kumar Kolluru "SVM Based dimensionality reduction and classification of hyperspectral data". The Netherlands April 2013.
- [3] Meiching Fong "Dimension Reduction on Hyperspectral Images", August 31, 2007.
- [4] R. Huang and M. He, "Band selection based on feature Weighting for classification of hyperspectral data," *IEEE Geosci. Remote Sens. Lett.* vol. 2, no. 2, pp. 156-159, Apr. 2005.
- [5] S. B. Serpico and G. Moser, "Extraction of spectral channels from hyperspectral images for classification purposes," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 2, pp. 484-495, Feb. 2007.
- [6] Craig Rodarmel and Jie Shan, "Principal Component Analysis for Hyperspectral Image Classification", *Surveying and Land Information Systems*, Vol. 62, No. 2, 2002, pp.115.
- [7] A. Villa, J. A. Benediktsson, J. Chanussot, and C. Jutten, "Hyperspectral image classification with independent component discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 4865-4876, Dec. 2011.
- [8] J. Benediktsson, M. Pesaresi, and K. Amason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Trans Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1940-1949, Sep. 2003.
- [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] X. Kang, S. Li, and J. A. Benediktsson, "Spectral spatial hyperspectral image classification with edge-preserving filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, Jun. 2014, to be published.
- [11] Xudong Kang, S. Li, and J. A. Benediktsson, "Feature Extraction of Hyperspectral Images With Image Fusion and Recursive Filtering", *IEEE Trans. Geosci. Remote Sens.*, to be published.
- [12] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397-1409, Jun. 2013.
- [13] E. S. L. Gastal and M.M. Oliveira, "Domain transform for edge-aware image and video processing," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 69:1-69:12, Jul. 2011.
- [14] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778-1790, Aug. 2004.
- [15] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1-27:27, Apr. 2011.
- [16] A. Villa, J. A. Benediktsson, J. Chanussot, and C. Jutten, "Hyperspectral image classification with independent component discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 4865-4876, Dec. 2011.
- [17] 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.*, vol. 113, no. S1, pp. S110-S122, Sep. 2009.
- [18] Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Multiple spectralspatial classification approach for hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4122-4132, Nov. 2010.
- [19] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 67:1-67:10, Aug. 2008.
- [20] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2864-2875, Jul. 2013.
- [21] R.C. Gonzalez and R.E. Woods, *Digital Image Processing*, second ed. Prentice Hall, 2002.
- [22] K. He, "Guided Image Filtering (Matlab Code)," <http://research.microsoft.com/en-us/um/people/kahe/>, 2010.