# Multikernel Learning using Relevance Vector Machine for Image Classification

Harshitha L
Post Graduate Scholar,MECS
DSCE,Bengaluru-78
Harshithahsm@gmail.com

Dr.V S Krushnasamy
Associate Professor
Dept.EIE,DSCE,Bengaluru
Krushnasamy.v@gmail.com

*Abstract*— **To enhance adaptability of kernel based machine learning, multiple kernel learning (MKL) strategies have been produced. The kernel learning technique utilize the kernel however it doesnot utilize feature extraction strategy as utilized in other machine learning strategies. The expansion of the separability in the kernel is the advancement in this technique. This strategy realizes a small within-class scatter large between class scatter . By uniting the basic kernels, an optimized combined kernel is found. MKL-FC is proposed from Fisher Criterion (FC) for finding the ideal projective direction. Classification of image is done using Relevance Vector Machine. A Matlab based implementation is done and the outcomes are tabulated. The parameters like position error, overall accuracy and run time are found.**

*Index Terms*— ***Multiple kernel Learning,Classification,Relevance Vector Machine.***

## I.INTRODUCTION

Remote Sensing is the marvel of acquiring the photos or scanned images of earth surface with the assistance of an aircraft or satellites. In other words, Remote Sensing is the way toward getting data in regards to the object not being in contact with it utilizing sensor, but rather from the electromagnetic radiation reflected from the objective. It is simply assembling of data comparing to the articles without physical contact. The source of remote sensing photo was taken from a ballon in 1858.

The principle point of remote sensing is to sort landcovers from the informations taken by the hyperspectral image and to build up an exact landcover maps. Later advances in hyperspectral remote sensing gives the utilization of different wavelength for each image pixel.This prompts precisely recognizing the materials of interest. Different imaging frameworks that makes utilization of images for various applications are recorded.

- Geographical Science: It is utilized to regain mineral properties such as structure.
- Hydrological Science: It is utilized to examine water quality, changes in wetland attributes.
- Environmental Science: It is utilized to evaluate biomass ,biodiversity or land cover changes.
- Agriculture:It is utilized to group rural classes and to separate nitrogen for agricultural precision.
- Military application:It is utilized for target discovery.

In remote sensing, large number of supervised methods have been created to handle the multispectral and hyperspectral data classification issue. In multispectral image classification . Artificial neural network and radial basis function neural networks are the fruitful approaches but they are not effective for larger spectral bands. Now-a-days, number of machine learning techniques are proposed for hyperspectral images. Among these strategies, Kernel learning turn out to be more appealing and mindful with its excellent execution using high-dimensional information.

Relevance vector machine (RVM) is a current advancement in kernel based machine learning approaches and can be utilized for both regression and characterization issues. RVM depends on a Bayesian formulation of a direct model.Relevance vector machines (RVM) have turned it's enthusiasm in different civil engineering applications. RVM can effectively be utilized for regression and characterization issues. Significant advantages of RVM are: 1. diminished sensitivity to the hyperparameter settings, 2. probabilistic yield with less relevance vectors for a given dataset.

In order to enhance the performance and interpretability, multiple kernels are used instead of using a single kernel

In current kernel strategies, the kernel function is resolved as Gaussian kernel, linear kernel, and polynomial kernel. There are three approaches to develop the fundamental kernels: 1) changing the parameters of the kernel with same input information; 2) changing the input information for every kernel with fixed kernel parameters; and 3) changing the information components and kernel parameters all the while. In MKL, we are consolidating multiple features or sources. Sparse MKL was used to take care of the issue of fundamental kernel[7]. Else ,optimization MKL approaches were proposed without strict restriction for determination of essential kernels[3]. One-step and two-step procedures were proposed to do optimization of kernels during the time of classification[5]. The most basic and direct combination of fundamental kernels depends on fixed rules considering direct summation or increase of the essential portions, similar to fundamental kernel with a similar weight coefficient . MKL strategies based on fixed rules does not require the weight coefficients for training, but they do not consider the qualities of different data. Keeping in mind the end goal to make full utilization of the Non-negative matrix factorization (NMF) and singular value decomposition (SVD) are utilized to make full use of nearby structure of data and the relationship between essential kernels. Low-rank NMF is utilized for the unsupervised learning and take an ideal combination of fundamental kernels in MKL. Its kernel extension, i.e., kernel NMF (KNMF), was additionally studied for unsupervised weight learning. In [3], gaining from essential kernels was planned as an issue of maximizing various projection, which was understood by SVD. The above strategies unravel the MKL model depending upon some specific attributes of the fundamental kernels, however they did not consider the relation between the MKL model and the classification task need to be solved.

In [2], the kernel parameters were chosen by amplifying the kernel arrangement with a fundamental kernel built with the vector of training output however without any ensures that the combination of kernels or the RVM classifier utilizing it would be the most discriminative.

Multiple Kernel Learning (MKL) is a method for consolidating multiple data features. One-step and two-step procedures were proposed to do improvisation of bulk of essential kernels during the classification process. In one-step MKL, simple MKL settles blend of essential kernel model and classifier parameter by seeking an optimal kernel combination in one step. In two-step strategy, the essential kernel model is solved in the initial step and the classifier parameters are obtained in the second step. Here we are utilizing a two-step technique. To start with locate an optimal projective direction, the most extreme class separability criteria is utilized. Every single essential kernels are projected into a consolidated combined kernel utilizing the projective direction, which diminishes the within-class scatter and expands the between-class scatter. The optimization of RVM with this consolidated kernel is done to acquire more effective classification.

SimpleMKL is one of the state-of-the-art algorithms used to comprehend MKL problem.This calculation uses gradient descent to iteratively optimize the objective function and discover a solution for the ideal optimization of basic kernels. A large portion of the MKL methods are not appropriate for remote sensing data due to the complexity included in the computation. In this manner, kernel alignment is used to enhance Simple MKL with computational efficiency and effectively executed MKL on distinguishing significant elements for remotely sensing data classification, including hyperspectral image. A novel sparse MKL algorithm is proposed based on a feature group choice technique that uses a group selection term implanted as a regularization term. The grouping choice term combines convex and concave terms and makes determination of kernels proportionate to selection of feature groups.

The contribution of this paper are the following:1)Maximizing distinctness in duplicating kernel to make the classes more divisible by finding an ideal projective direction;2)Constructing diverse essential kernels to acquire optimal kernel parameters;3)a strategy for higher computational efficiency since it has less support vectors.

The paper is structured such that the fundamentals of RVM classification is explained in the next section. In section III, the MKL methods are outlined. In section IV, the datasets are described before presenting the results in section V and concluding in section VI.

## II.RVM

Relevance Vector Machine (RVM) is a machine learning method that utilizes Bayesian interference to acquire stingy answers for regression and probabilistic classification.The RVM has an indistinguishable frame than the support vector machine, but gives probabilistic classification. It is proportional to a Gaussian procedure model.

Contrasted with that of support vector machines (SVM), the Bayesian formulation of the RVM opposes the arrangement of free parameters of the SVM (that often require cross-approval based post-optimizationss). However RVMs utilize a desire expansion (EM) - like learning technique and are consequently at risk of nearby minima. This is not at all like the standard sequential minimal optimization (SMO)- based calculations utilized by SVMs, which are ensured to locate a worldwide ideal (of the raised issue).

The Relevance Vector Machine (RVM) initially introduced by M. Tipping (2001), is a Bayesian learning model which gives best results comes out both regarding accuracy and sparsity by means of proper definition of various leveled priors, effectively constraining most of the model parameters around zero. In this manner, by expanding the marginal probability utilizing a type-II maximum likelihood (ML) strategy, we accomplish arrangements which use just a little subset of the original functions, named the relevance vectors. In spite of the fact that the Relevance Vector Machine gives fundamentally competitive outcomes as compared to the conventional Support Vector Machine, its adjustment to the multi-class setting has been hazardous, because of the terrible scaling of the type II ML technique with respect to the number of classes C. Now-a-days, two algorithms for calculation, mRVM1 and mRVM2 have been presented which expand the first Relevance Vector Machine to the multi-class multikernel setting [6]. These calculations accomplish sparsity without the requirement of having a binary class issue and give probabilistic yields to class membership rather than the hard parallel choices given by the conventional SVMs.

The RVM was originally designed for binary classification[6]. In a two class classification by RVM,the aim is to predict the posterior probability of membership for one of the classes (0 or 1) for a given input x. A case may then be allocated to the class with which it has the greatest likelihood of membership. For a 2-class problem with training data $X=(x_1,\ldots\ldots\ldots,x_n)$ having class labels $C=(c_1,\ldots\ldots\ldots,c_n)$ with $c_i \in(-1,1)$ . Based on Bernoulli distribution, the likelihood is represented by Eq.(1) as:

$$p(c/w)=\prod_{i=1}^{n} \sigma\{(y(x_i))\}^{ci}[1\sigma\{(y(x_i))\}]^{1-ci} \quad (1)$$

$\sigma(y)$ is the logistic sigmoid function represented by Eq(2):

$$\sigma(y(x))=\frac{1}{1+\exp(-y(x))} \quad (2)$$

To obtain p(c/w), an iterative method has to be used. Let $\alpha_i^*$ denotes the maximum a posteriori estimate of the hyperparameter $\alpha_i$ . The maximum posteriori estimate of the weights ( $W_m$ ) can be obtained by maximizing the following objective function represented by Eq(3):

$$f(w_1,w_2,\ldots,w_n)=\sum_{i=1}^{n} \log p(c_i/w_i)+\sum_{i=1}^{n} \log p(w_i/\alpha_i^*) \quad (3)$$

where the first summation term corresponds to the likelihood of the class labels, and the second term corresponds to the prior on the parameters $w_i$. In the resulting solution, the gradient of f with respect to w is calculated and only those training data having nonzero coefficients $w_i$ (called relevance vectors) will contribute to the decision function.

The posterior is approximated around $W_m$ by a Gaussian approximation as in Eq (4) & (5) with

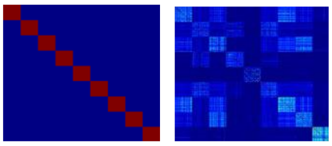
$$\text{covariance} \quad \Sigma = -(H|w_i)^{-1} \quad (4)$$

$$\text{mean} \quad \mu=\Sigma\Phi^T Bc \quad (5)$$

where H is the Hessian of f, matrix Φ has elements $\varphi_{ij}=K(x_i,x_j)$ and B is a diagonal matrix with elements defined by $\sigma(y(x_i))[1-\sigma(y(x_i))]$ .

An iterative analysis is followed to find the set of weights that maximizes the function in which the hyperparameters $\alpha_i$,associated with each weight are updated.

## III.MKL method



(a) (b)

Fig.1. Different kernel matrices of Pavia University using 9 classes with 50 training samples per class.(a)Ideal kernel (b)Simple MKL

The ideal kernel is computed from the inner product of data samples and the corresponding labels. The ideal kernel is shown in Fig1(a). We can observe that the within-class scatter is 0 and the between-class scatter is infinity for the ideal kernel. When utilizing comparability between data points, there is fluctuation in each class and likeliness between different classes. As an outcome, the kernel acquired along these way strongly varies from the ideal one. Simple MKL is acquired by taking the average of sigma values.Simple MKL is shown in Fig1(b). Now, we need to take an account of base kernels that decreases the within-classes scatter and expands the between-classes scatter in the combined kernel.M kernel matrices are obtained from candidate basic kernels $K_o=\{K_m, m=1,2,\ldots,M, K_m \in R^{N\times N}\}$.A 3-D data cube of size (N ×N ×M) is generated by the series of kernel matrices. In order to facilitate the subsequent operations, the 3-D data cube is converted to a 2-D matrix by using a vectorization operator represented by Eq.(6)

$$k_m= vec(K_m), \; m=1,2,\ldots,M, \quad (6)$$

where vec(·) is the vectorization operator which converts a matrix into a vector. The vectored set of kernels is $P = [vec(k_1),vec(k_2),\ldots,vec(k_M)]^T =[p_1,p_2,\ldots,p_{N2}]\in R^{M\times N2}$. Once P has been generated, two classes are extracted, denoted as c1 and c2. The elements of c1 were constituted by the diagonal elements of the basic kernels. These elements correspond to those training samples belonging to the same class. The remaining elements of the basic kernels constitute c2, corresponded to the kernel values between points of different classes. Two scalars (nc1 and nc2) are defined as the sum of elements of classes c1 and c2 formulated as $nc_1=\sum_{i=1}^{c} n_i^2$, $nc_2=N^2-\sum_{i=1}^{c} n_i^2$ respectively.

Then, we can calculate the mean vectors of each class represented by Eq.(7)

$$m^{c1} = \frac{1}{nc1}\sum_{j=1}^{nc1} k_j^{c1} \in R^{M\times 1} \qquad m^{c2} = \frac{1}{nc2}\sum_{j=1}^{nc2} k_j^{c2} \in R^{M\times 1} \quad (7)$$

The within-class scatter matrix Si,i=1 ,2 and between-class scatter matrix Sb are defined represented by Eq(8) & (9)

$$S_i=\sum_{j\in ci}\{(k_j^{ci} - m^{ci})(k_j^{ci} - m^{ci})^T\}, i=1,2 \quad (8)$$

$$Sb=(m^{c1}-m^{c2})(m^{c1}-m^{c2})^T \quad (9)$$

where the T superscript denotes the transpose. The total within- class scatter matrix is defined as St = S1 + S2. We have to find an M ×1 projective direction w∗ that decreases the within-class scatter and increases the between-class scatter in mapped 1-D subspace projected by $y = (w^*)^T P$. This means finding a projection where samples of the same class are close to each other and far from those of other classes [10]. To find the projective direction, we propose a variant based on the Fisher criterion (FC)

FC

Using the FC, the projective direction w can be calculated by solving the following maximization problem, as in linear discriminative analysis represented by Eq.(10)

$$w^*=\arg\max \{(w^T S_b w)/(w^T S_t w)\} \quad (10)$$

In order to ensure the stability of solution, the within-class scatter matrix is regularized by adding a constant value to its diagonal elements.After solving the projection direction, the DMKL model is determined represented by Eq.(11):

$$K^*=\sum_{m=1}^{M} d_m k_m = \sum_{m=1}^{M} w_m^* k_m \quad (11)$$

The whole procedure is summarized in MKL algorithm Shown below:

DMKL Algorithm

i)Initializing the kernel scale values

ii)Compute the basic kernel matrices

iii)Solve the projective direction corresponding to that of MKL-FC.

iv)Use projective direction w* to project the basic kernels to a combined kernel k*.

v)Utilize the combined kernel k* to solve the classification problem based on RVM.

RVM operates in two modes for proposed MKL: Training mode(Fig.2) and testing mode(Fig.3)

Training mode: During this mode, we are going to train the RVM by assigning the classes for various features of the image. This is done by naming different classes.

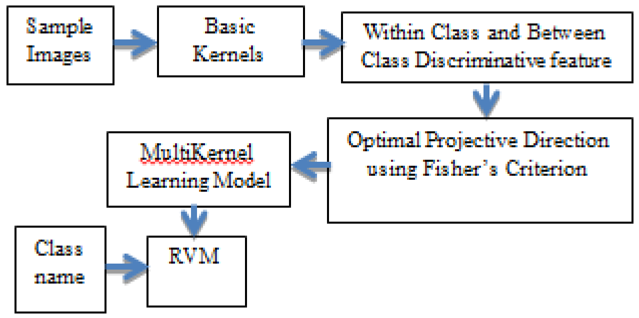



Fig.2.Training mode of proposed MKL

Testing mode: During this mode, new images are taken to classify their features. They are going to classify as the RVM are trained i.e., if the feature in the new image are trained initially, then it will classify that feature of the image by using training data.

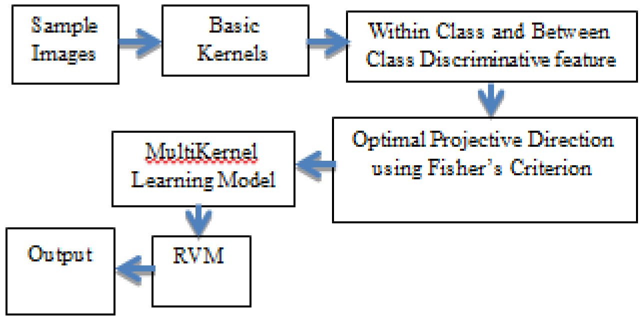
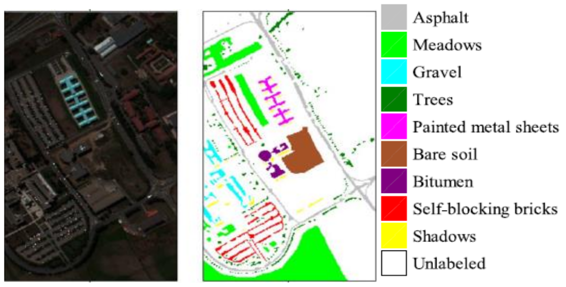



Fig.3.Testing mode of proposed MKL

## IV.DATA DESCRIPTION

The data sets considered are real multispectral and hyperspectral remote sensing images. They are detailed as follows.

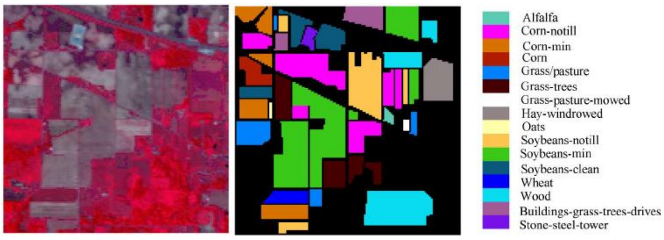
1)ROSIS PaviaU Data Set: The second data set was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS-03) optical sensor over an urban area surrounding the University of Pavia, Italy, on July 8, 2002. There are 42776 labeled samples in total and 9 classes of interest. The false-color composite image and class information are shown in Fig.4.




(a) (b)

Fig.4.Pavia University data set.(a) RGB composite image of three bands.(b)Ground truth map.

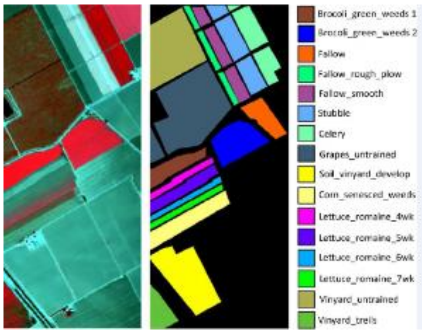
2)Indian Pines Sample Data Set: The second data set was acquired by the Indian Pines test site in North Western India.The scene consists of two-thirds agriculture,one-third forest and other perennial vegetation.The available ground truth consists of 16 classes and are not mutually exclusively. The false-color composite image and class information are shown in Fig.5.




(a) (b)

Fig5.Indian Pines data set.(a) False-color composite image.(b)Ground truth image

3)Salinas Data Set: The third data set was collected by the 224-band AVIRIS sensor over Salinas Valley, California, and is characterized by high spatial resolution (3.7-meter pixels). The area covered comprises 512 lines by 217 samples.This image was available only as at-sensor radiance data. It includes vegetables, bare soils, and vineyard fields. Salinas groundtruth contains 16 classes.



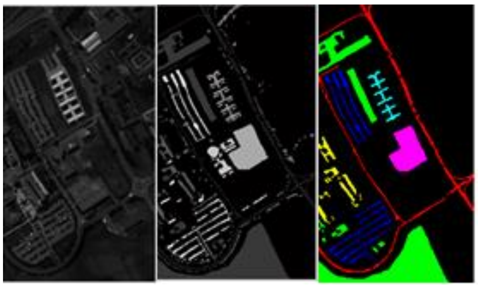
(a) (b)

Fig6.Salinas data set.(a)Sample set.(b)Groundtruth set

## V.RESULTS AND DISCUSSION

Matlab based simulation of the proposed algorithm is carried out on three sets of data. One of the image from the database is taken and Multiple Kernel Learning algorithm is implemented.After applying MKL,the effective classification is done using RVM.. The overall accuracy, position error and run time were tabulated for RVM.

Initially, Pavia University dataset is taken for experimental purpose.Downloading of image with different variations is done using MATLAB coding. The corresponding Simulation result is as shown in Fig.7.(a). The groundtruth image for the corresponding downloaded image is as shown in Fig.7.(b). 6 classes with 50 training samples for each classes were defined. Thus 6 different features/objects were detected and classified as shown in Fig.7.(c).Thus objects in an image are detected using the proposed MKL algorithm and classification of the detected object in an image is done using Relevance Vector Machine.



(a) (b) (c)
Fig.7.Pavia university dataset results

Likewise, identification and classification of objects in an image is carried out for the remaining two datasets i.e.,Indian pines and Salinas data set. Accuracy, position error and run time were tabulated.

Table I. Accuracy, Position error and Run time for RVM classifier

| | Pavia | Indian pines | Salinas |
|---|---|---|---|
| Accuracy(%) | 97.21 | 97.64 | 98.08 |
| Position Error(%) | 2.79 | 2 .36 | 1.92 |
| Run time(Sec) | 497.6 | 69.75 | 287.6 |

## VI.CONCLUSION

In this paper, multiple kernel learning (MKL) algorithm is proposed. In MKL, we have found a discriminative projective direction and have built the multiple kernel by projecting the base kernels.MKL-FC is proposed and applied to very high spatial resolution spectral image classification. Experiments were carried out. RVM is used for classification. The Matlab based implementation were carried out and the results were satisfactory. The position error, accuracy and runtime were calculated and tabulated.

Future work is to increase the classification accuracy and applying the proposed methods in several application,such as target detection.

## REFERENCES

[1] F. Bach, G. Lanckriet, and M. Jordan, “Multiple kernel learning, conic duality, and the SMO algorithm,” in Proc. Int. Conf. Mach. Learn., 2004, pp. 41–48.
[2] D. Tuia, G. Camps-Valls, G. Matasci, and M. Kanevski, “Learning relevant image features with multiple-kernel classification,” IEEE Trans. Geosci. Remote Sens., vol. 48, no. 10, pp. 3780–3791, Oct. 2010.
[3] Y. Gu et al., “Representative multiple-kernel learning for classification of hyperspectral imagery,” IEEE Trans. Geosci. Remote Sens., vol. 50, no. 7, pp. 2852–2865, Jul. 2012.
[4] 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.
[5] Y. Gu, Q. Wang, H. Wang, D. You, and Y. Zhang, “Multiple kernel learning via low-rank nonnegative matrix factorization for classification of hyperspectral imagery,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 99, no. 6, pp. 1–13, Oct. 2014
[6]Michael E. Tipping “Sparse Bayesian Learning and the Relevance Vector Machine,” MicrosoftResearch St George House, 1 Guildhall Street Cambridge CB2 3NH, U.K.
[7] A. Rakotomamonjy, F. R. Bach, S. Canu, and Y. Grandvalet, “Simple- MKL,” J. Mach. Learn. Res., vol. 9, pp. 2491–2521, Nov. 2008.
[8] 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.
[9] A. Rakotomamonjy, F. R. Bach, S. Canu, and Y. Grandvalet, “Simple- MKL,” J. Mach. Learn. Res., vol. 9, pp. 2491–2521, Nov. 2008.
[10] H. L. Yang and M. M. Crawford, “Spectral and spatial proximity-based manifold alignment for multitemporal hyperspectral image classifica- tion,” IEEE Trans. Geosci. Remote Sens., vol. 54, no. 1, pp. 51–64, Jan. 2016.
[11] L. Fang, S. Li, W. Duan, J. Ren, and J. Atli Benediktsson, “Classification of hyperspectral images by exploiting spectral–spatial information of superpixel via multiple kernels,” IEEE Trans. Geosci. Remote Sens., vol. 53, no. 12, pp. 6663–6674, Dec. 2015.
[12]Mahesh Pal, “Kernel Methods in Remote Sensing: A review”Department of Civil Engineering NIT Kurukshetra, 136119.
[13] G. Camps-Valls and L. Bruzzone, “Kernel-based methods for hyperspectral image classification,” IEEE Trans. Geosci. Remote Sens., vol. 43, no. 6, pp. 1351–1362, Jun. 2005.
[14] H. Zhou, Z. Mao, and D. Wang, “Classification of coastal areas by air- borne hyperspectral image,” in Proc. SPIE Opt. Technol. Atmos., Ocean, Environ. Stud., 2005, pp. 471–476.
[15] Y. Kim and Y. Kim, “Improved classification accuracy based on the output-level fusion of high-resolution satellite images and airborne LiDAR data in urban area,” IEEE Trans. Geosci. Remote Sens. Papers., vol. 11, no. 3, pp. 636–640, Mar. 2014.
[16] Y. Zhong, J. Zhao, and L. Zhang, “A hybrid object-oriented conditional random field classification framework for high spatial resolution remote sensing imagery,” IEEE Trans. Geosci. Remote Sens., vol. 52, no. 11, pp. 7023–7037, Nov. 2014.
[17] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. Atli Benediktsson, “Advances in hyperspectral image classification: Earth monitoring with statistical learning methods,” IEEE Signal Process, vol. 31, 45–54, Jan. 2014.
[18] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, “Advances in spectral–spatial classification of hyperspectral images,” Proc. IEEE, vol. 101,no.3,pp.652–675,March2013.