A Novel MKL Model of Integrating LiDAR Data and MSI for Urban Area Classification

Yanfeng Gu, *Member, IEEE*, Qingwang Wang, *Student Member, IEEE*, Xiuping Jia, *Senior Member, IEEE*, and Jón Atli Benediktsson, *Fellow, IEEE*

***Abstract*—A novel multiple-kernel learning (MKL) model is proposed for urban classification to integrate heterogeneous features (HF-MKL) from two data sources, i.e., spectral images and LiDAR data. The features include spectral, spatial, and elevation attributes of urban objects from the two data sources. With these heterogeneous features (HFs), the new MKL model is designed to carry out feature fusion that is embedded in classification. First, Gaussian kernels with different bandwidths are used to measure the similarity of samples on each feature at different scales. Then, these multiscale kernels with different features are integrated using a linear combination. In the combination, the weights of the kernels with different features are determined by finding a projection based on the maximum variance. This way, the discriminative ability of the HFs is exploited at different scales and is also integrated to generate an optimal combined kernel. Finally, the optimization of the conventional support vector machine with this kernel is performed to construct a more effective classifier. Experiments are conducted on two real data sets, and the experimental results show that the HF-MKL model achieves the best performance in terms of classification accuracies in integrating the HFs for classification when compared with several state-of-the-art algorithms.**

***Index Terms*—Classification, heterogeneous features (HF), light detection and ranging (LiDAR), multiple-kernel learning (MKL), multispectral images.**

I. Introduction

RECENTLY, multispectral images (MSIs) and/or hyperspectral images (HSIs) have been widely used in land cover mapping [1]–[3]. However, when using such data, it can be difficult to distinguish objects built by similar materials, e.g., between streets and roofs of buildings. As a result, it is necessary to exploit the joint use of other data sources. Light detection and ranging (LiDAR) data, from which the altitude of a surface with respect to the sensor can be extracted [4]–[7], provide complementary information to spectral images. In other words, MSI/HSI provide information related to the spectral domain, and LiDAR data give information related to the spatial domain (vertical distribution). On the other hand, multispectral and hyperspectral sensors belong to passive sensors, and data provided by them might be influenced by the surrounding environment or climate. As LiDAR is an active senor, data provided by it are less affected by the surrounding environment or climate, when compared with passive MSI/HIS imaging devices. Because of the complementary properties of these data, the joint use of MSI/HSI and LiDAR data for classification has attracted a considerable attention in recent years [8]–[17].

To make full use of the information provided by both spectral images and LiDAR data, different strategies for feature extraction are widely adopted. A natural and simple strategy is to stack elevation and intensity data of LiDAR as additional channels to spectral bands [8], [9]. Dalponte *et al.* [8] proposed a system based on the joint use of hyperspectral and LiDAR data for classification of complex forest areas, using elevation and intensity channels of LiDAR data along with spectral bands as features. Moreover, they found that the elevation channel of the first LiDAR return data played the most important role for increasing the discriminability of different classes with similar spectral signatures. Sugumaran and Voss [9] combined hyperspectral data with LiDAR for identification of tree species, which outperformed the classification of using spectral images only. In [10], Puttonen *et al.* used 34 LiDAR-derived features related to the height distribution of trees along with all 123 spectral bands of hyperspectral images to accomplish classification of tree species in an urban garden. Dinuls *et al.* [11] developed a method for identification of tree species using LiDAR and multispectral data. In their work, LiDAR data were used to detect tops of individual trees by means of a so-called single-tree method. In addition to identification of trees, previous research has been also done on joint classification of LiDAR and hyperspectral data for urban areas. For example, in [12], Lemp and Weidner generated a normalized digital surface model (nDSM) from LiDAR data and applied the eCognition software to accomplish classification of roof surface materials with hyperspectral and laser scanning data. In [13], Elshehaby and Taha extracted nDSM from LiDAR data and the normalized difference vegetation index (NDVI) from spectral information and used these two features as additional channels with spectral data for classification of an urban area (buildings, roads, and vegetation). In [14], experiments were carried out on LiDAR data along with either hyperspectral or multispectral images collected on a rurally urban area. In that research, extended

Manuscript received May 5, 2014; revised September 29, 2014 and February 6, 2015; accepted April 6, 2015. Date of publication May 7, 2015; date of current version June 10, 2015. This work was supported in part by the Natural Science Foundation of China under Grant 61371180 and in part by the Fundamental Research Funds for the Central Universities under Grant HIT.NSRIF.2010095.

Y. Gu and Q. Wang are with the School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001, China (e-mail: guyf@hit.edu.cn; 786120585@qq.com).

X. Jia is with the School of Engineering and Information Technology, The University of New South Wales, Canberra, A.C.T. 2600, Australia, and also with Harbin Engineering University, Harbin 150001, China (e-mail: x.jia@adfa.edu.au).

J. A. Benediktsson is with the Department of Electrical and Computer Engineering, University of Iceland, 107 Reykjavik, Iceland (e-mail: benedikt@hi.is).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2015.2421051

attribute profiles were extracted from two data sources as joint features to better exploit spatial information. In [15], García *et al.* presented a method of mapping fuel types in a nature park using LiDAR and multispectral data. In their work, several spectral indexes were derived from spectral information, and the height above the ground of each vegetation point was computed with LiDAR data, and these features were used with spectral bands. To summarize, a commonly applied approach to make use of features collected from spectral and LiDAR data is that all features are normalized into 0 to 1 and then stacked together as an input vector for classifiers [8], [10], [13], [15].

Considering classification techniques, the support vector machine (SVM), which has some significant advantages over other classification approaches, such as robustness and less sensitivity to the increase in dimensionality, is a widely used method. It has been successfully applied in joint use of spectral images and LiDAR data [8], [10], [14], [17] and has shown superior performance [8], [14] in terms of classification accuracies when compared with other widely used classifiers, such as the Gaussian maximum-likelihood method with the leave-one-out covariance algorithm and the random forest (RF) classifier. Decision tree classifiers were adopted in some published papers for joint classification of spectral images and LiDAR data. In [15], the joint classification of fuels was divided into two phases: SVM was adopted in the first phase to classify the main groups of fuels. In the second phase, a set of decision rules was applied according to the Prometheus classification system. Koetz *et al.* [16] combined multispectral and LiDAR data to map fuel types for forest fire management via a three-level classification tree. The first and second levels represent functional categories and land cover classes that are relevant for fuel type discrimination. The third level is thematically divided by surface properties and could be extended to include different vegetation functional types or species.

In addition to these aforementioned investigations, which are directly related to the classification with the spectral images and LiDAR data, an effective kernel-based framework was proposed by Camps-Valls *et al.* [18] to integrate heterogeneous information from multitemporal and multisource remote sensing data for classification and change detection. Four ways to form composite kernels were given in their work, including *stack, direct summation, weighted summation, and cross information*. The results of their study indicate that direct summation composite kernels yield better results and achieves higher efficiency in the particular application domain of urban monitoring, outperforming the traditional stacked-vector approach in real-scenario cases. Their work was not involved in the joint classification of the spectral images and LiDAR data, but the direct-summation composite kernel method can be used for this task.

While recent research has shown promising performance of joint classification of spectral images and LiDAR data, there are still some challenging problems that need to be addressed.

First, most existing methods either only considered a few features or only used a subset of these features for a specific classification task, such as tree species recognition. There is still no general categorization of the features that can be extracted from the spectral images and LiDAR data.

Second, regarding the use of the heterogeneous features (HFs), existing methods often simply stack all features together and form an extended feature vector. Features from different sources obviously have different meanings, dimension units, and statistical significance. Therefore, they may play different roles in classification. Consequently, the stacked-vector approach is not a good choice for the joint classification.

Third, regarding classification procedures, SVM and decision tree are mainly adopted in existing methods for joint classification of spectral images and LiDAR data. However, SVM cannot weight HFs because it treats all features equally. Decision-tree-based classifiers may achieve a good performance in terms of accuracies, but its design is a time-consuming and challenging task.

In summary, the increasing number of features from different sources brings a big challenge on how to integrate them, as well as precludes the use of simple classifiers.

Recently, MKL based on the SVM scheme has been developed and has become one of the hot topics in pattern analysis and machine learning [19]–[22]. There are two main issues for most MKL methods, namely, which kernels should be selected and how to accommodate their significance in the kernel combination. Each MKL method has its way to construct and solve an optimization problem to address these two issues. In terms of remote sensing applications, it has been shown that MKL is a highly useful approach [23]. It has been also demonstrated in the literature that MKL is an effective tool for combining HFs [24].

In this paper, an HF-MKL model is proposed to integrate effective features from the spectral images and the LiDAR data for urban classification. The main contributions of this paper are threefold.

1) An effective feature extracting strategy is presented. All the features are categorized into three different attribute types, i.e., spectral, spatial, and elevation attributes; and discriminative information from these three different feature types is extracted. This strategy will help the development of a new multiple-kernel learning (MKL) model with HFs from hyperspectral images and LiDAR data.
2) In the HF-MKL model, basis kernels are constructed to account for individual features and individual scales. Then, the MKL framework is implemented at both the scale and feature levels, and a joint classification with different kinds of HFs is achieved.
3) For the HF-MKL algorithm, we carry out MKL separately at the scale and feature levels in order to find the optimal subspace projection, which maximizes the variance for basis kernels with reduced complexity in optimization. The automatically selected weights for basis kernels can be expected to preserve most information in both the scale and feature domains. By doing this, the discriminative ability of HFs is exploited at different scales and is integrated to generate an optimal combined kernel. The optimization routine of the conventional SVM with this multiple kernel is performed to construct a more effective classifier.

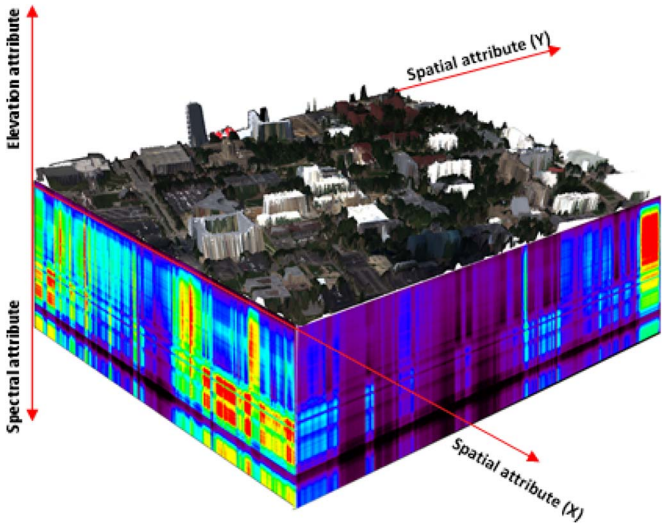

Fig. 1. Categorization of multidimensional attributes from the spectral images and LiDAR data.

The rest of this paper is organized as follows. In Section II, the feature-extracting strategy used in our work is described. In Section III, we first provide a brief introduction to the conventional SVM, which provides the theoretical basis for the proposed algorithm. Then, several state-of-the-art MKL algorithms are overviewed. The HF-MKL model for joint classification of spectral images and LiDAR data is presented in details in Section IV. Section V presents the experimental results on two real data sets and compares the HF-MKL method with other classifiers in terms of classification accuracies, the Kappa coefficient, Z-test, and thematic maps of classification results. Finally, conclusions are drawn in Section VI.

II. Features From Spectral Images and LiDAR Data

High-dimensional MSI/HSI data contain huge information for classification. However, not all the measurements are important and useful for a specific application. Therefore, feature mining is necessary in most cases [26].

Here, a scheme for extracting features from spectral images and LiDAR data is described. The features include original spectra of urban objects and morphological profiles (MPs) from spectral images, nDSM, and MPs from first LiDAR return data.

According to the difference of attributes represented by various HFs, we first categorize the features that can be extracted from spectral images and LiDAR data into three different types, i.e., spectral, spatial, and elevation attributes. This categorization indicates how exactly the different HFs from the spectral images and LiDAR data represent. Fig. 1 demonstrates the categorization of object attributes, based on the spectral images and LiDAR data.

The features that correspond to the spectral attribute usually include the spectral band, the NDVI, and the intensity image of LiDAR data. The features that correspond to the spatial attribute include MPs, Gabor texture, and the gray-level cooccurrence matrix. The features that correspond to elevation attribute include DSM from different LiDAR returns, nDSM, and local features derived from point cloud, such as gradient, shape, and normal vector.

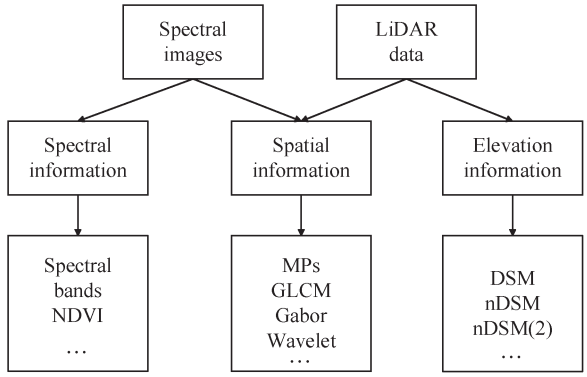

Fig. 2. Features extracted from spectral images and LiDAR data.

Fig. 2 shows the organization of the HFs extracted from the spectral images and LiDAR data according to the categorization. Selected features in this study are discussed in detail in the following.

A. Spectral Bands

For multispectral images, the number of bands is not as high as for hyperspectral images, and the correlation between bands is small. Each channel of the original multispectral image is treated as a feature.

B. MPs

A morphological filter is used to extract structural information. MPs have been shown to have good discrimination capability for remote sensing classification [24], [25]. In morphological transform, opening and closing are commonly used operations. The opening operations can remove objects smaller than the structuring element (SE) defined and isolate the fractions. On the contrary, closing operations can fill small holes and connect adjacent objects. The SE is used to detect the shape and size of geometrical structures and control the smoothness of morphological operations. The MP of an image $f(x, y)$, i.e., MP$(f(x, y))$, is defined as the combination of the original image $f(x, y)$, the opening profiles OP$(f(x, y))$, and the closing profiles CP$(f(x, y))$ with SEs of multiple scales. Here, MPs are computed for the first principal component of the spectral bands and the nDSM, as described in the next subsection.

C. nDSM

The height information is the most important information provided by LiDAR. The elevation channel of a LiDAR image measures the height of land cover objects, and the corresponding image is called a digital surface model (DSM) in the literature. To remove the influence of terrain, pixels of elevation channel are divided into ground (soil, grass, road, etc.) and nonground (buildings, trees, etc.) pixels based on the first return. A digital terrain model (DTM) can be generated using the ground pixels. Then, an nDSM can be computed by subtracting DTM from DSM. The nDSM represents the height of land objects with respect to the ground.

The aforementioned features extracted from spectral images and LiDAR data are used to classify land covers from different domains. Spectral bands respond to materials of land covers. The nDSM provides information related to vertical spatial domain, whereas MPs provide information related to the horizontal spatial domain.

III. MKL

A. From Single-Kernel Learning to MKL

Kernel-based methods, in general, and SVM, in particular, have been investigated for remote sensing image classification. We can achieve the single-kernel learning (SK) by embedding kernel method in SVM.

1) Single-Kernel Learning Algorithm (SK): The main idea of the nonlinear SVM is to obtain a classification hyperplane in the mapped higher dimensional feature space with a maximum margin by solving a convex quadratic programming as follows:

$$\max\left\{Q(\alpha_i,\alpha_j)=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j K(\mathbf{x}_i,\mathbf{x}_j)\right\}$$

$$\text{s.t.}\begin{cases}\sum_{i=1}^{N}\alpha_i y_i=0\\ \alpha_i,\alpha_j\in[0,C] \qquad \forall_{i,j}=1,2,\ldots,N\end{cases} \tag{1}$$

where α_i and α_j are Lagrange multipliers, and if α_i is nonzero, the corresponding $\mathbf{x}_i$ is called a support vector, which determines the decision hyperplane.

We use kernel alignment (KA) to determine suboptimal single kernel in the single-kernel learning algorithm. Given two kernels matrices $\mathbf{K}_1$ and $\mathbf{K}_2$ computed on the same training set $\mathbf{X}$, the KA between the two kernels is defined as follows [24]:

$$\mathrm{KA}(\mathbf{K}_1,\mathbf{K}_2)=\frac{\langle\mathbf{K}_1,\mathbf{K}_2\rangle_F}{\sqrt{\langle\mathbf{K}_1,\mathbf{K}_1\rangle_F\langle\mathbf{K}_2,\mathbf{K}_2\rangle_F}} \tag{2}$$

where $\langle\bullet,\bullet\rangle$ is the Frobenius norm between the two matrices and defined as $\langle\mathbf{K}_1,\mathbf{K}_2\rangle_F=\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{K}_1(\mathbf{x}_i,\mathbf{x}_j)\mathbf{K}_2(\mathbf{x}_i,\mathbf{x}_j)$.

If a mapping function exists between data samples and the corresponding labels, the mapped space will be constituted of class labels. The samples are linearly separable in the mapped space because of distinct labels for different classes. The corresponding kernel is called the ideal kernel, which can be computed by inner product of labels for binary classifier, i.e., $K(\mathbf{x}_i,\mathbf{x}_j)=y_i\cdot y_j$. Then, the Gram matrix of the ideal kernel $\mathbf{K_1}$, is represented as $\mathbf{K_1}=\mathbf{yy^T}$, where $\mathbf{y}=[\mathrm{y}_1,\mathrm{y}_2,\ldots,\mathrm{y}_N]^T$ is the vector of labels. For multiclass classification, the values of the ideal kernel are computed as $K(\mathbf{x}_i,\mathbf{x}_j)=\begin{cases}0, & y_i\neq y_j\\ 1, & y_i=y_j\end{cases}$. Class labels are the target of classification task; thus, the ideal kernel can be viewed as a target benchmark for selection of discriminative kernels. We choose the kernel $\mathbf{K}_2$, which makes $\mathrm{KA}(\mathbf{K}_1,\mathbf{K}_2)$ to become maximum. The detailed process of choosing the suboptimal kernel can be found in [24].

2) MKL: It has been well documented that using a single kernel is seriously limited for pattern recognition and classification [24]. A possible solution is to extract discriminative information by using multiple basis kernels. It is expected that an optimal kernel is learned by building a weighted linear combination of M basis kernels. The combined kernel is generally expressed as follows:

$$K(\mathbf{x}_i,\mathbf{x}_j)=\sum_{m=1}^{M}d_m K_m(\mathbf{x}_i,\mathbf{x}_j)$$
$$\text{s.t. } d_m\geq 0, \text{ and } \sum_{m=1}^{M}d_m=1 \tag{3}$$

where d_m is the weight for the mth basis kernel $K_m(\bullet,\bullet)$.

The multiple-kernel version of SVM dual problem can be represented as

$$\max\left\{Q(\alpha_i,\alpha_j)=\sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j\sum_{m=1}^{M}d_m K_m(\mathbf{x}_i,\mathbf{x}_j)\right\}$$

$$\text{s.t.}\begin{cases}\sum_{i=1}^{N}\alpha_i y_i=0\\ \alpha_i,\alpha_j\in[0,C], \quad \forall_{i,j}=1,2,\ldots,N\\ d_m\geq 0, \quad \text{and } \sum_{m=1}^{M}d_m=1.\end{cases} \tag{4}$$

This intuitive idea leads to the development of MKL methods. Let $\mathbf{D}$ be a vector that is composed of $[d_1,d_2,\ldots,d_M]^T$. In the MKL framework, the combinational weights D and the Lagrange multipliers for each support vector in SVM are optimized simultaneously. However, solving the MKL problem becomes rapidly intractable with the increase of training examples and number of kernels. This fact seriously limits the application of the conventional MKL algorithm. Recently, some algorithms have been proposed to reformulate the original version of the MKL and a new way to solve the optimization [23].

B. Related Work of MKL

The brief introduction to the state-of-the-art MKL algorithms based on SVM is given in the following, which are used as baseline algorithms in this paper.

1) Simple MKL: To solve the MKL problem, the Simple MKL algorithm utilizes the gradient descent (GD) method to optimize the decision function of SVM. The MKL based on the SVM optimization problem is represented as

$$\min\left\{J(\mathbf{w},\xi_i,b)=\frac{1}{2}\sum_{m=1}^{M}\frac{1}{d_m}\|\mathbf{w}_m\|^2+C\sum_{i=1}^{N}\xi-i\right\}$$

$$\text{s.t.}\begin{cases}y_i\left(\sum_{m=1}^{M}\langle\mathbf{w}_m,\Phi_m(x_i)\rangle+b\right)\geq 1-\xi_i\\ \xi_i\geq 0, \quad \forall_i=1,2,\ldots,N\\ d_m\geq 0, \quad \text{and } \sum_{m=1}^{M}d_m=1\end{cases} \tag{5}$$

where $\mathbf{w}_m$ is the weighting vector of the mth decision hyperplane with the nonlinear mapping function $\Phi_m(\mathbf{x}_i)$ corresponding to the basis kernel K_m.

The objective function $Q(\alpha_i, \alpha_j)$ in (4) is convex and differentiable. Therefore, the combination weight d_m can be optimized by updating it on the GD direction of Q. The gradient of objective function is computed by the derivatives of Q as

$$\frac{\partial Q}{\partial d_m} = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j K_m(\mathbf{x}_i, \mathbf{x}_j), \qquad m = 1, 2, \ldots, M. \tag{6}$$

We search the descent direction D of gradients and update weights as

$$\mathbf{d} = \mathbf{d} + \lambda \mathbf{D} \tag{7}$$

where λ is the step length, and $\mathbf{d} = [d_1, d_2, \ldots, d_M]^T$ is the combination coefficients vector.

2) RMKL: Representative MKL (RMKL) tries to determine the set of coefficients $\{d_m\}_{m=1}^{M}$ within the learning process of the decision function as well. Given Gram matrices of M basis kernels $\mathbf{K}_0 = \{\mathbf{K}_m, m = 1, 2, \ldots, M, \mathbf{K}_m \in \mathbb{R}^{N\times N}\}$. This series of matrices construct a *3-D* data cube, which needs to be converted into a *2-D* matrix so that we can conveniently deal with vector or matrix operation. After vectorization, we can get a new form of $\mathbf{K}_0$ expressed as $\mathbf{L} = [\mathbf{k}_1, \mathbf{k}_2, \ldots, \mathbf{k}_M]^T = [\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_{N^2}] \in \mathbb{R}^{M\times N^2}$. Thus, $\mathbf{l}_i$ is an M-dimensional column vector, $i = 1, 2, \ldots, N^2$. We find a low-dimensional linear subspace and build a corresponding loss function as follows:

$$\Gamma(\mathbf{W}, \mathbf{Z}) = \|\mathbf{L} - \mathbf{W}\mathbf{Z}\|_F^2 = \sum_{i=1}^{N^2}\sum_{j=1}^{N^2}\left(l_{j,i} - \sum_{r=1}^{p} w_{j,r} z_{r,i}\right)^2 \tag{8}$$

where $\mathbf{W} \in \mathbb{R}^{M\times p}$ is the projection matrix whose columns $\{w_r\}_{r=1}^{p}$ are the bases of p-dimensional linear subspace, and $\mathbf{Z} \in \mathbb{R}^{p\times N^2}$ is the projected matrix onto the linear subspace spanned by W. $\|\bullet\|_F$ denotes the Frobenius norm of the matrix.

Minimizing the loss function $\Gamma(\mathbf{W}, \mathbf{Z})$ is only determined by $\mathbf{Z} = \mathbf{W}^{\mathrm{T}}\mathbf{L}$. Thus, the minimization for $\Gamma(\mathbf{W}, \mathbf{Z})$ is represented as

$$\arg\max_{\mathbf{W}} \left\|\mathbf{W}^{\mathrm{T}} \sum\nolimits_{\mathbf{L}} \mathbf{W}\right\|_F = \arg\max_{\mathbf{W}} \left\|\mathbf{W}^{\mathrm{T}}\mathbf{L}\right\|_F$$
$$\text{s.t.} \quad \mathbf{W}^{\mathrm{T}}\mathbf{W} = \mathbf{I}_p \tag{9}$$

where $\sum_{\mathbf{L}} = \mathbf{L}\mathbf{L}^{\mathrm{T}}$, and $\mathbf{I}_p$ is the $p \times p$ identify matrix.

For more information on the subspace kind of MKL, please see our previous work [23].

3) Mean MKL: Mean MKL has another name in [27], i.e., RBMKL, which denotes rule-based MKL. Mean MKL trains an SVM with the mean of the combined kernels, which means that the weights of each kernel are the same, i.e.,

$$K(\mathbf{x}_i, \mathbf{x}_j) = \sum_{m=1}^{M}\frac{1}{M}K_m(\mathbf{x}_i, \mathbf{x}_j) \tag{10}$$

where M is the number of basis kernels.

IV. HF-MKL Model for Joint Classification

According to the principle of MKL, it can be seen that MKL provides a higher capacity in learning samples or features with a combined kernel. More important, these basis kernels not only can share same features but also integrate HFs for a learning task, such as classification or regression. As far as the joint classification of the spectral images and LiDAR data is concerned, a novel MKL model is designed to integrate the heterogeneous spectral, spatial, and elevation features and measure the similarity of these features at different kernel scales. Fig. 3 shows the framework of the proposed MKL model. It can be found that there are two levels of MKL in column and row, respectively.

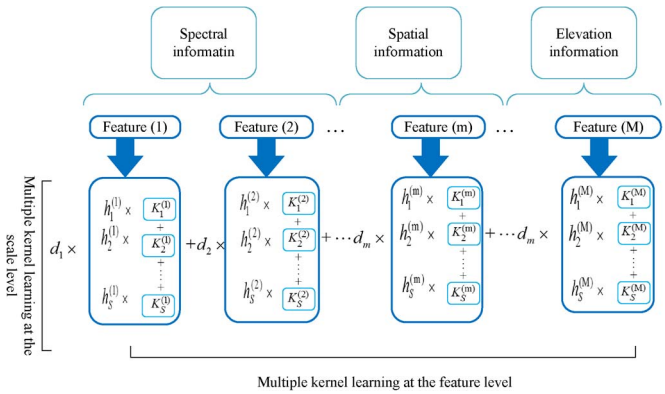

Fig. 3. HF-MKL model for joint classification of spectral images and LiDAR data.

A. Multiscale MKL on Each Feature

In the HF-MKL model, the HFs are first used to separately measure similarity of samples by using multiscale basis kernels. This is the "column" MKL. Each column of MKL model corresponds to each feature but different basis kernels. Here, Gaussian kernels with different bandwidths are used for the basis kernels. This way, the similarity of samples with each feature in a column of the model can be captured at different scale levels. The Gaussian kernel is the most widely used kernel for remote sensing images. It can be shown as follows:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right) \tag{11}$$

where σ is a radius parameter, which is called width or bandwidth. The parameter σ can be regarded as a scale; the range of the scale to use controls the kernel resolution and the discriminative ability of the kernels.

The goal of integrating kernels at different scales is to simultaneously obtain a higher sensitivity to the degree of similarities between samples and the kernel matrix and a better generalization capability. Given a data set $\mathbf{X} = \{\mathbf{x}_i, i = 1, 2, \ldots, N\}$, where $\mathbf{x}_i \in \mathbb{R}^M$ is an input training vector with M elements (M is the number of the HFs). The basis kernel corresponding to the mth feature with scale s, i.e., the basic element of the column MKL, can be then constructed as

$$K_s^{(m)}(x_i^{(m)}, x_j^{(m)}) = \exp\left(-\frac{\|x_i^{(m)} - x_j^{(m)}\|^2}{2\sigma_s^2}\right) \tag{12}$$

where m is an index of the HFs, and $m = 1, 2, 3, \ldots, M$; s is an index of scales, $s = 1, 2, 3, \ldots, S$; and S is the total number of scales; $x_i^{(m)}$ is the ith sample of the mth feature.

To integrate different similarity scales, a series of basis kernels that account for different values of the kernel parameters are then linearly combined, i.e.,

$$K_{\text{Col}}^{(m)}\left(x_i^{(m)}, x_j^{(m)}\right) = \sum_{s=1}^{S} h_s^{(m)} K_s^{(m)}\left(x_i^{(m)}, x_j^{(m)}\right)$$
$$\text{s.t. } h_s^{(m)} \geq 0, \text{ and } \sum_{s=1}^{S} h_s^{(m)} = 1 \tag{13}$$

where $h_s^{(m)}$ is the weight for the sth basis kernel on the mth feature.

B. MKL Model on HFs

The following procedure of the HF-MKL model is to integrate the HFs with spectral, spatial, and elevation attributes. This is the "row" MKL, as shown in Fig. 3. When an MKL-based classifier confronts features from different sources (e.g., spectral and LiDAR data), constructing basis kernels that only account for scales is not enough. It is not a good choice that all these features are simply transformed using same kernel function and apply the same set of weights for different scales. Therefore, the row MKL is used to integrate the HFs via a linear combination of the basis kernels selected for individual features. By building a weighted linear combination of M basis kernels corresponding to each feature, an optimal kernel will be learned as

$$K_{\text{Row}}(x_i, x_j) = \sum_{m=1}^{M} d_m K_{\text{Col}}^{(m)}\left(x_i^{(m)}, x_j^{(m)}\right)$$
$$\text{s.t. } d_m \geq 0, \text{ and } \sum_{m=1}^{M} d_m = 1 \tag{14}$$

where d_m is the weight for the mth basis kernel $K_{\text{Col}}^{(m)}(\bullet, \bullet)$. It is worth noting that d_m can be used to indicate the importance of the mth feature.

C. Learning With HF-MKL Model

With (14), we can rewrite the optimization function based on SVM as

$$\max_{\alpha_i, d_m} \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \sum_{m=1}^{M} d_m K_{\text{Col}}^{(m)}\left(x_i^{(m)}, x_j^{(m)}\right)$$
$$\text{s.t.} \begin{cases} \sum_{i=1}^{N} \alpha_i y_i = 0, \quad \alpha_i, \alpha_j \in [0, C] \\ d_m \geq 0, \quad \text{and } \sum_{m=1}^{M} d_m = 1 \end{cases} \tag{15}$$

$$\max_{\alpha_i, d_m} \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \sum_{m=1}^{M} d_m \times \sum_{s=1}^{S} h_s^{(m)} K_s^{(m)}\left(x_i^{(m)}, x_j^{(m)}\right)$$
$$\text{s.t.} \begin{cases} \sum_{i=1}^{N} \alpha_i y_i = 0, \quad \alpha_i, \alpha_j \in [0, C] \\ h_n^m \geq 0, \quad \text{and } \sum_{n=1}^{N} h_n^m = 1 \\ d_m \geq 0, \quad \text{and } \sum_{m=1}^{M} d_m = 1 \end{cases} \tag{16}$$

where α_i is the Lagrange multiplier, and C controls the generalization capabilities.

Equation (15) is a typical MKL optimization problem. Equation (13) can be further extended to (16). Here, a subspace learning approach proposed in our previous work [23] is adopted. A low-dimensional subspace is solved by a maximum L2-norm variance problem. An optimal linear combined kernel is then obtained via the subspace learning. This kernel is then used within the optimization routine of the conventional SVM, and the final multiple-kernel classifier can be derived.

The subspace learning problem can be expressed as the following:

$$\min_{\mathbf{w}^T\mathbf{w}} \left\|\mathbf{K}^* - \mathbf{w}^T\mathbf{K}\right\|_2^2 \tag{17}$$

where $\mathbf{K} = [(\mathbf{K}_1^{(1)})_{1D} \cdots (\mathbf{K}_S^{(1)})_{1D} \cdots (\mathbf{K}_1^{(m)})_{1D} \cdots (\mathbf{K}_S^{(m)})_{1D} \cdots (\mathbf{K}_1^{(M)})_{1D} \cdots (\mathbf{K}_S^{(M)})_{1D}]^T$ $\mathbf{K} \in \mathbb{R}^{P \times N^2}$, $P = M \times S$, P is the total number of the basis kernels, $(\mathbf{K}_s^{(m)})_{1D}$ is the 1-D form of the column basis kernel matrix corresponding to the mth feature and the sth scale, and $\mathbf{w} \in \mathbb{R}^{P \times 1}$ is the projection vector, which is the base of linear subspace corresponding to the maximum variance projection. $\mathbf{K}^*$ is the 1-D form of the optimal kernel matrix, which is used in the following SVM optimization.

According to the HF-MKL model in Fig. 3, the columns and the rows in the multiple-kernel models have different meanings and play different roles. The multiscale kernels in the columns share the same features, and the basis kernels in the rows are used to integrate the similarity measure on the HFs. Hence, it is a feasible way to separately carry out MKL, namely, to find the optimally subspace projection on the column kernels and the row kernels one after another. Additionally, to solve (17), a fact is that there would be $M \times S$ basis kernels, and the size of each basis kernel is $N \times N$. N is the number of training samples for classification. Such a large amount of data may bring challenges for computer storage and optimization efficiency. To tackle this problem, a strategy is that two series of kernel weights, i.e., d_i and h_s, corresponding to features and scales, respectively, are optimized separately. Two sequential optimizations can be respectively expressed as the following:

$$\min_{\mathbf{w}_{\text{Col}^T}\mathbf{w}_{\text{Col}}} \left\|\mathbf{K}_{\text{Col}}^{(m)} - \mathbf{w}_{\text{Col}}^T \mathbf{K}_{\text{Col}}^{(m)}\right\|_2^2 \tag{18}$$

where $\mathbf{K}_{\text{Col}}^{(m)} = [(\mathbf{K}_1^{(m)})_{1D} \cdots (\mathbf{K}_S^{m})_{1D}]^T$, $\mathbf{K} \epsilon \mathbb{R}^{S \times N^2}$, $m = 1, 2, \ldots, M$, and

$$\min_{\mathbf{w}_{\text{Row}}^T\mathbf{w}_{\text{Row}}} \left\|\mathbf{K}^* - \mathbf{w}_{\text{Row}}^T \mathbf{K}_{\text{Row}}\right\|_2^2 \tag{19}$$

where $\mathbf{K}_{\text{Row}} = [(\mathbf{K}_{\text{Col}}^{(1)})_{1D} \cdots (\mathbf{K}_{\text{Col}}^{(M)})_{1D}]^T, \mathbf{K} \in \mathbb{R}^{M \times N^2}$.

Thus, this algorithm carries out subspace learning twice on column multiscale basis kernels and row multifeature basis kernels, respectively. Such a way leads to a more effective way for joint classification of features from different data sources. More details about subspace MKL and how to solve (18) and (19) can be found in [23].

To further enhance efficiency, we select the suboptimal single kernel for each feature by KA instead of MKL in the scale level. After getting the suboptimal single kernel for each feature, MKL is carried out in the feature level. We use the aforementioned subspace learning approach to carry out MKL in the feature level. In this paper, we denote the method as *KA-MKL*, which is regarded as a comparison method.

D. Procedure for KA-MKL and HF-MKL

The procedure for KA-MKL [a Gaussian radial basis function (RBF) kernel is used here] is given as follows.

Algorithm: KA-MKL

i) Initialize the range of kernel scale values $[\sigma_{\min}, \sigma_{\max}]$
ii) Sample N scales within the preceding range, i.e., $\sigma_{\min} = \sigma_1 \sigma_2 \cdots < \sigma_N = \sigma_{\max}$
iii) For each feature, select the suboptimal single kernel for each feature by KA.
iv) Among all the features, find the coefficient d_m for each feature's kernel via solving the optimization problem (19).
v) Utilize d_m to solve MKL classification problem based on SVM.

The procedure for *HF-MKL* (a Gaussian RBF kernel is used here) is given as follows.

Algorithm: HF-MKL

i) Initialize the range of kernel scale values $[\sigma_{\min}, \sigma_{\max}]$
ii) Sample S scales within the preceding range, i.e., $\sigma_{\min} = \sigma_1 < \sigma_2 \cdots < \sigma_S = \sigma_{\max}$
iii) For each feature, find the kernel coefficient $h^s_{(m)}$ for each scale by solving the optimization problem (18).
iv) Among all the features, find the coefficient d_m for each feature's kernel via solving the optimization problem (19).
v) Utilize $h^s_{(m)}$ and d_m to solve MKL classification problem based on SVM.

V. Experimental Results and Analysis

A. Data Description

The experiments were carried out on two groups of real data sets. The data sets are the combination of multispectral images and LiDAR data. The two data sets are classified as different classes according to heights and materials of land covers. In the following part, description of the data sets is first given, and the main characteristics of the data sets are summarized in Table I.

These two data sets are from two subregions of a whole scene around downtown area of San Francisco, USA. One is located at a factory named "Recology," and the other is located at a park named "Bayview Park." The data come from 2012 IEEE GRSS Data Fusion Contest and contain multispectral images (eight bands in the wavelength range of 400–1040 nm) and the corresponding LiDAR data. The multispectral images were acquired by WorldView2 on October 9, 2011, and the LiDAR data were acquired in June 2010. The two data sets have a spatial resolution of 1.8 m. Figs. 4 and 5 show the false RGB composition and information of the labeled classes for the two selected study areas, respectively.

TABLE I
Main Characteristics of the Two Data Sets

Data set	Bayview Park	Recology
Image size	300×200	200×250
Labeled pixels	19537	11811
Background pixels	40463	38189
Spatial resolution	1.8m	1.8 m
Classes	7	11

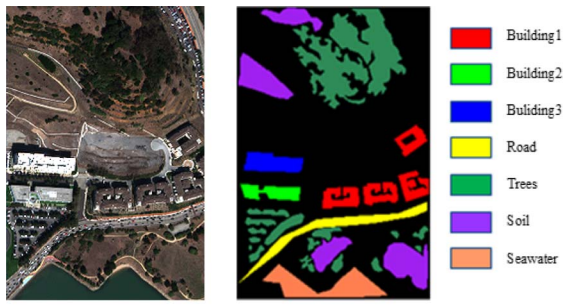

Fig. 4. Bayview Park data set. (a) RGB composite image of bands. (b) Ground truth map.

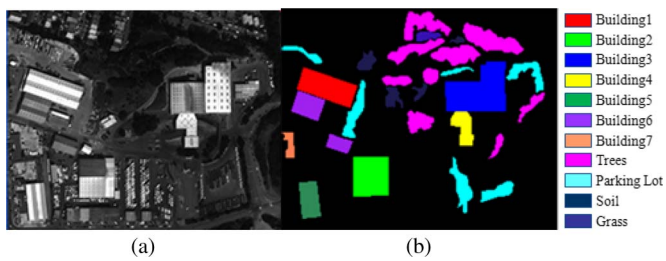

Fig. 5. Recology data set. (a) RGB composite image of bands. (b) Ground truth map.

The land cover classes in the Bayview Park data set are "Building1," "Building2," "Building3," "Road," "Trees," "Soil," and "Seawater," which were identified by visual inspection with the help of Google Earth.

The reference land cover classes in the Recology data set, which were also identified by visual inspection with the help of Google Earth, are "Building1," "Building2," "Building3," "Building4," "Building5," "Building6," "Building7," "Trees," "Parking lot," "Soil," and "Grass."

Data collected by multispectral sensors are a series of images corresponding to different spectral channels, whereas the original LiDAR data are usually organized as point cloud with LAS format. To enable LiDAR data to be available together with spectral information, the LiDAR point clouds should be interpolated and rasterized at the same spatial resolution as the spectral images. That is accomplished by assigning to each pixel the mean value of points within the associated area on

TABLE II
NUMBER OF FEATURES FOR THE TWO DATA SETS

Data set	Bayview Park	Recology
Spectral features	8	8
MPs	8	8
nDSM	1	1

the ground. In other words, the LiDAR point cloud data are converted into a series of raster images, corresponding to elevation and intensity channels of each laser pulse return. For pixels with missing data, linear interpolation is adopted. The obtained LIDAR images are then coregistered to the spectral images, and a polynomial transformation of second order and a nearest neighbor resampling of the pixels are used. It is worth noting that the singular values points of elevation should be deleted.

B. Experimental Setting

The numbers of features extracted for these two groups of data sets are given in Table II and described as follows.

Spectral features were constructed by using the original eight spectral bands. MPs were computed on two windows of sizes 3×3 and 5×5 pixels. The dimension of the morphological features is 8 (with 2 opening and 2 closing for the first principal component of spectral bands and the same number for the LiDAR data). Then, nDSM was calculated with the method in Section II. There were 17 features used for classification.

To validate the proposed HF-MKL model, we compare it with several state-of-the-art methods. They are single-kernel *SVM* (*SK* for short), *RKML* [23], *Mean MKL* [27], and *Simple MKL* [28]. At the same time, KA-MKL is proposed as a method of comparison, which is introduced in Section IV. In those kernel-based SVM classifiers, it is very crucial to tune key parameters, including kernel scales, model parameters of SVM, and the combination of the coefficients, which were solved by MKL algorithms. Gaussian RBF was adopted as kernel function in all the methods. Hence, the kernel parameter to be determined is the scale σ. For all the classifiers, the range of σ was set to [0.05, 2], and uniform sampling that selects scales from the interval with a fixed step size of 0.05 was used to select 40 scales within the given range, which may not be exactly optimal for the classifiers, but are considered reasonable after a study of different combinations of scales.

For the model parameters of SVM, which were used in all SVM-based classifiers, the penalization parameter C is selected by cross-validation (CV) in the range of $[10^{-1}, \ldots, 10^{7}]$, and the slack variable $\xi = 10^{-7}$ also is selected by CV.

In all experiments, the labeled training samples were randomly selected. The number of training samples for each class was set to 10, 15, 20, 30, 40, 50, and 100. The rest of the samples in each data set were treated as test samples. In order to avoid biased conclusions, experiments were conducted with ten trials, each time with randomly selected training samples, and the average results were reported. All classifiers were compared in terms of overall accuracy (OA, the percentage of pixels correctly assigned), Kappa statistic, Z-test, computational cost, and a thematic map of classification results.

TABLE III
POSSIBLE RESULTS OF TWO ALGORITHMS

	Algorithm A failed	Algorithm A succeeded
Algorithm B failed	N_{ff}	N_{sf}
Algorithm B succeeded	N_{fs}	N_{ss}

1) Kappa Statistic: Kappa statistic is based on the comparison of the predicted and actual class labels for each case in the testing set and has been widely used for evaluating image classification in remote sensing [29]. It can be calculated as $kappa = (P_o - P_c)/(1 - P_c)$, where P_o is the proportion of cases in agreement (i.e., correctly assigned), and P_c is the proportion of agreement that is expected by chance. When the Kappa coefficient is in the range 0–0.2, the classification accuracy is considered "poor"; for values in the range 0.2–0.4, the classification accuracy is considered to be "normal"; in the range 0.4–0.6, the classification accuracy is considered to be "good"; in the range 0.6–0.8, the classification accuracy is considered "very good"; in the range 0.8–1.0, the classification accuracy is considered to be "the best".

2) Z-test: According to Z-test, two algorithms can have four possible outcomes arranged in a 2×2 contingency table [30], as shown in Table III.

N_{ff} denotes the number of times (instances) when both algorithms failed, and N_{ss} denotes success for both algorithms. These two cases do not give much information about the performances of algorithms as they do not indicate how their performances differ. However, the other two parameters (N_{sf} and N_{fs}) show cases where one of the algorithms failed and the other succeeded, indicating the performance discrepancies.

In order to quantify the differences of these two algorithms, Z-test is defined as

$$Z\text{-test} = \frac{N_{sf} - N_{fs}}{\sqrt{N_{sf} + N_{fs}}}.$$

Z-test scores are interpreted as follows: When Z - test $= 0$, the two algorithms are said to show a similar performance. As this value diverges from 0 in the positive direction, this indicates that their performance differs significantly. Furthermore, if $|Z - \text{test}| > 1.96$, the performance of two algorithms differs at 5% level of significance.

C. Experimental Results

In order to examine the effectiveness of data fusion, four different experiments were designed, which include the use of spectral images and the LiDAR data, respectively (these two cases are not HF), and joint use of spectral images and LiDAR data for classification.

1) Classification Using Spectrum and nDSM, respectively: First, we only used the features of spectrum that come from MSI directly for classification on the two data sets. The histograms of OA and Kappa coefficients with the standard deviation of ten trials, which vary along with the number of the training samples per class, are revealed for the two data sets in Figs. 6 and 7, respectively.

Second, we used nDSM, which comes from LiDAR directly for the classification on the two data sets. The histograms of

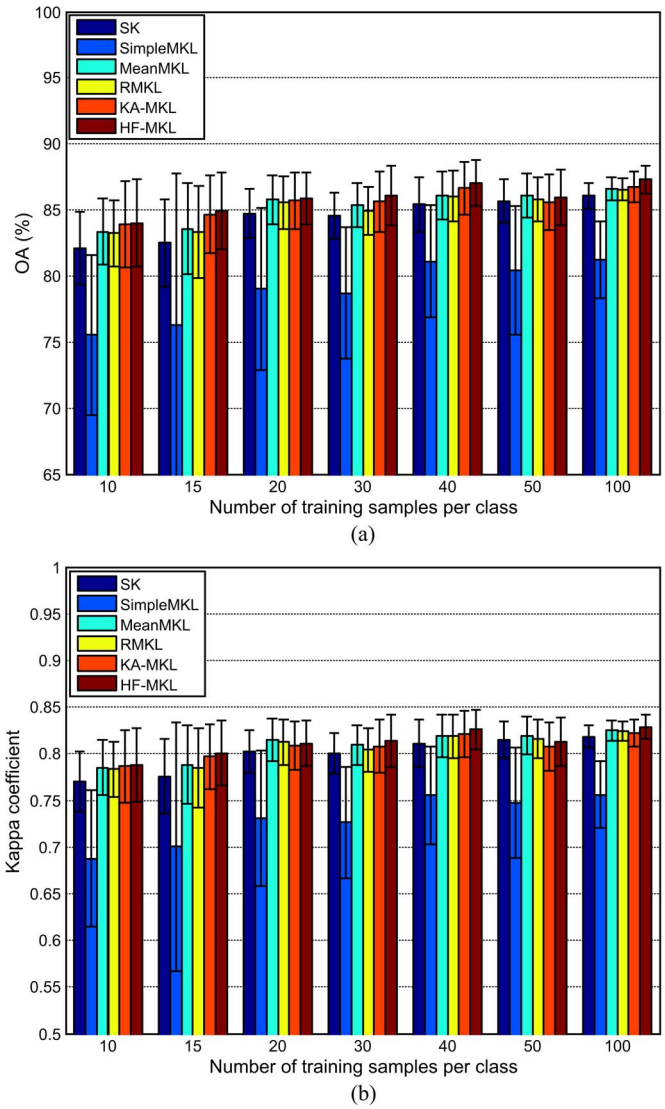

Fig. 6. (a) OA and (b) Kappa coefficient of the Bayview Park data set.

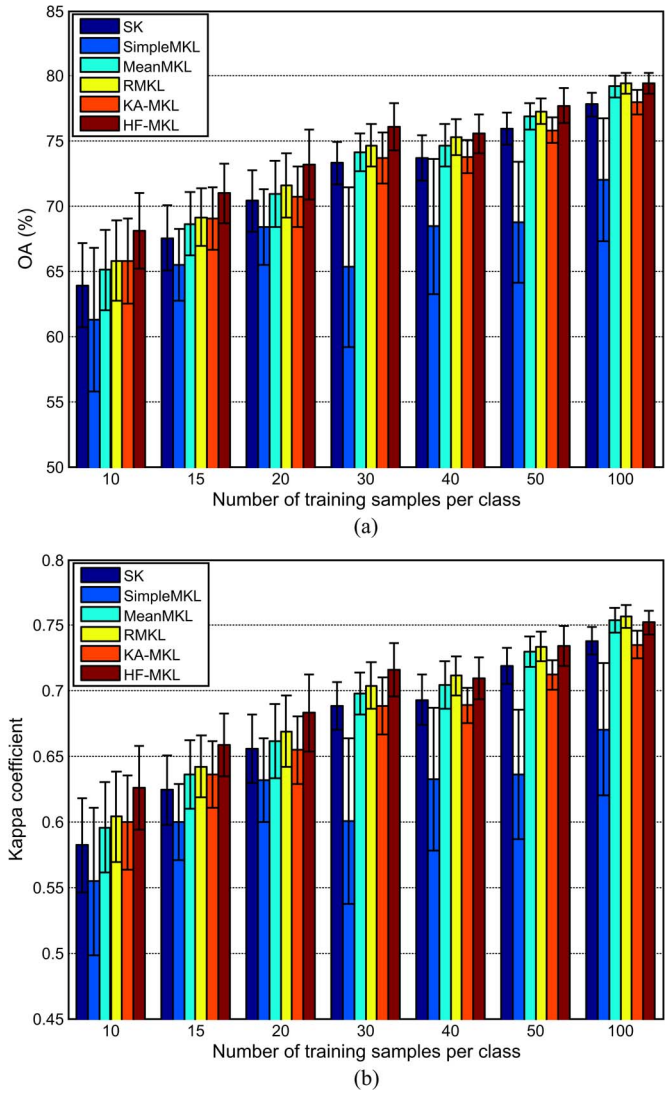

Fig. 7. (a) OA and (b) Kappa coefficient of the Recology data set.

OA with the standard deviation of ten trials, which vary along with the number of the training samples per class, are shown in Figs. 8 and 9.

The results observed from Figs. 6–9 can be summarized as follows.

1) The MKL methods (RMKL, Mean MKL, KA-MKL, and HF-MKL) outperform the single-kernel method (SK) in terms of accuracies.
2) HF-MKL performs best among the MKL approaches in terms of accuracies.
3) The classification accuracy increases with the increase of training samples.
4) In classification, LiDAR and MSI provide different information; thus, it is significant to integrate LiDAR data and MSI.

2) Simple Joint Classification With Spectrum and nDSM: Joint classification with spectrum and nDSM, which is the simplest features setting to joint use of spectral images and LiDAR data, was carried out. The average results for ten trials are shown for the two data sets in Figs. 10 and 11.

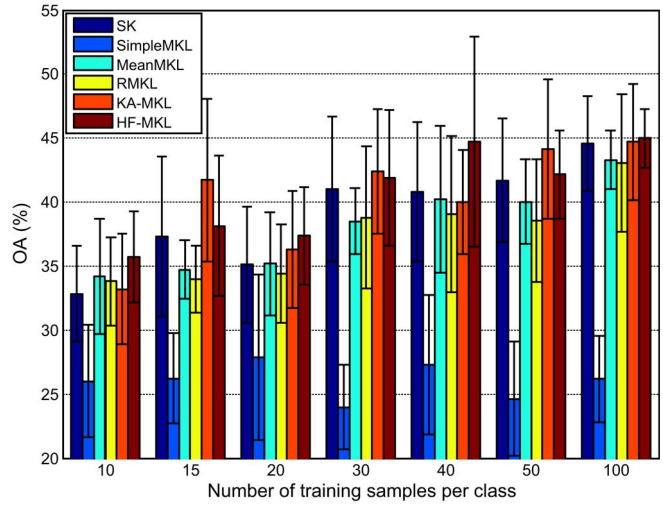

Fig. 8. OA of the Bayview Park data set.

From experiments 1 and 2, we can see that joint use of spectral images and LiDAR data is better than only using the LiDAR data or the spectral images for classification. Experiment 2

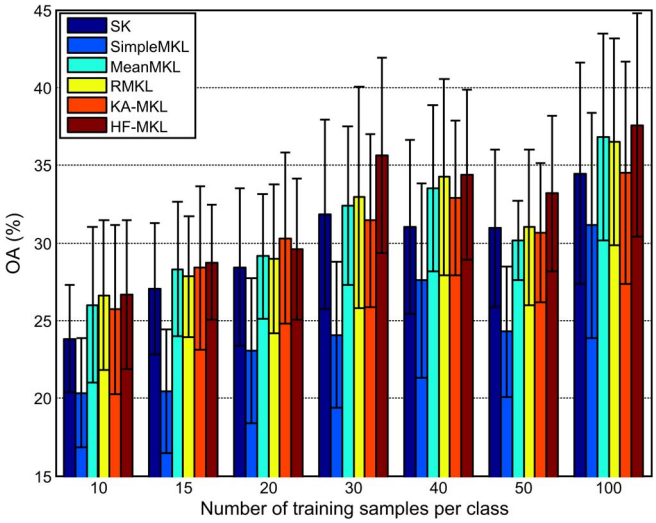

Fig. 9. OA of the Recology data set.

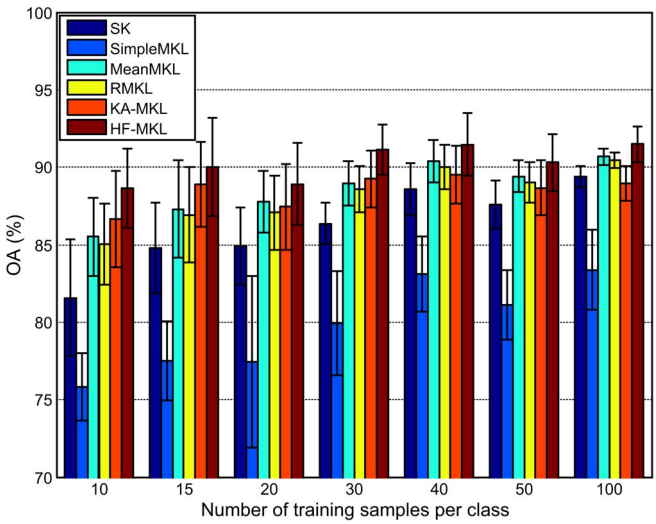

Fig. 10. OA of the Bayview Park data set.

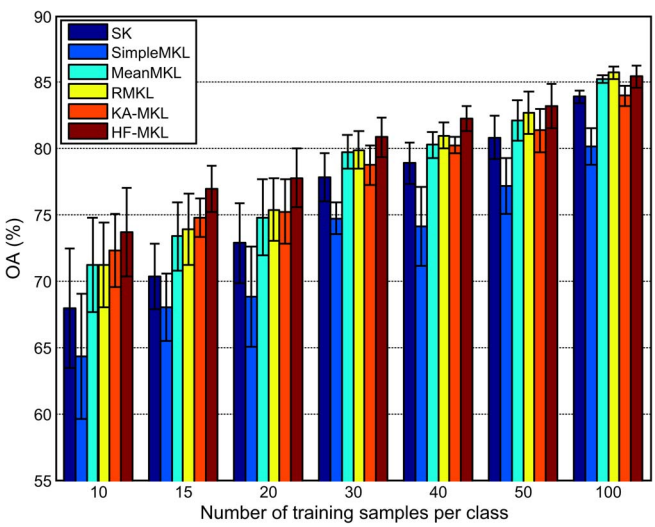

Fig. 11. OA of the Recology data set.

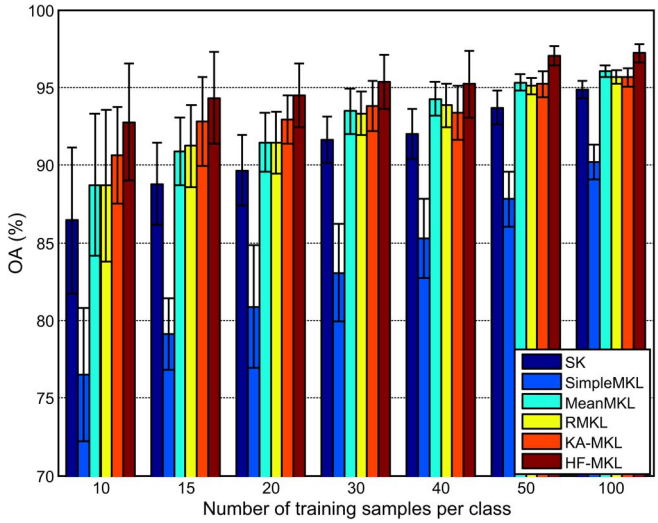

Fig. 12. OA of the Bayview Park data set.

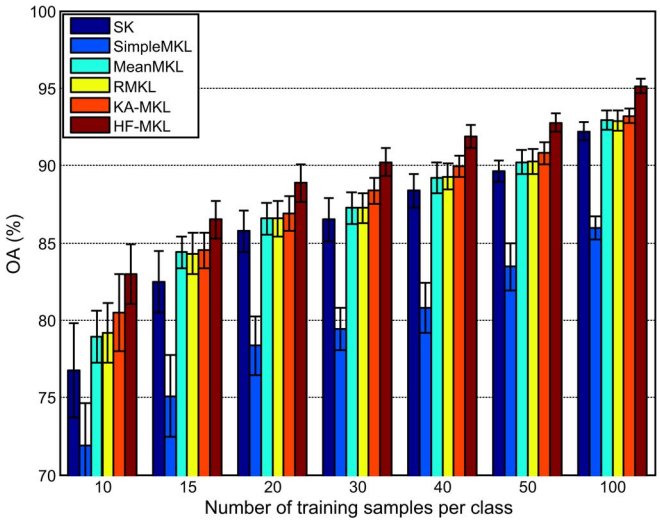

Fig. 13. OA of Recology data set.

shows that the LiDAR data and spectral images can provide complementary information for classification.

3) Joint Classification With Spectral, Spatial, and Elevation Features: Joint classification with the spectral features (spectral attributes), nDSM (elevation attributes), and MPs (spatial attributes) was carried out, and the results of classification for the two data sets are shown as follows.

1) The histograms of OA with ten trials are shown in Figs. 12 and 13.

From Figs. 12 and 13, we can see that HF-MKL achieves more than 95% classification accuracy by using the spectral features (spectral attributes), the nDSM (elevation attributes), and the MPs (spatial attributes). In the united use of HFs, HF-MKL is best in all the classifiers, with the improvement in the 2%–3% level compared with the other multiple-kernel classifiers.

2) The Z-tests between HF-MKL and other classifiers are shown in Tables IV and V.

From the results of the Z-test, we can know that the performance of HF-MKL and other classifiers differs at 5% level of

TABLE IV
Z-TEST OF THE BAYVIEW PARK DATA SET

Classifier	HF-MKL						
	10	15	20	30	40	50	100
SK	**20.23**	**17.77**	**14.86**	**9.63**	**7.39**	**8.49**	**2.62**
Simple MKL	**47.53**	**45.85**	**41.91**	**39.85**	**32.87**	**32.11**	**25.02**
Mean MKL	**10.73**	**8.48**	**5.69**	**4.98**	**3.55**	**2.29**	-6.15
RMKL	**11.26**	**7.03**	**6.27**	**4.14**	**3.06**	0.53	-3.47
KA-MKL	**7.94**	**4.23**	**5.10**	**4.42**	**6.65**	**7.04**	**4.99**

TABLE V
Z-TEST OF THE RECOLOGY DATA SET

Classifier	HF-MKL						
	10	15	20	30	40	50	100
SK	**16.10**	**10.34**	**8.14**	**10.27**	**10.00**	**9.18**	**8.76**
Simple MKL	**28.80**	**29.84**	**28.13**	**29.61**	**30.52**	**27.01**	**26.57**
Mean MKL	**11.16**	**4.88**	**5.56**	**8.34**	**7.46**	**7.22**	**5.97**
RMKL	**10.55**	**5.24**	**5.61**	**8.42**	**7.09**	**7.04**	**6.03**
KA-MKL	**5.49**	**4.84**	**5.07**	**4.67**	**5.28**	**5.43**	**5.59**

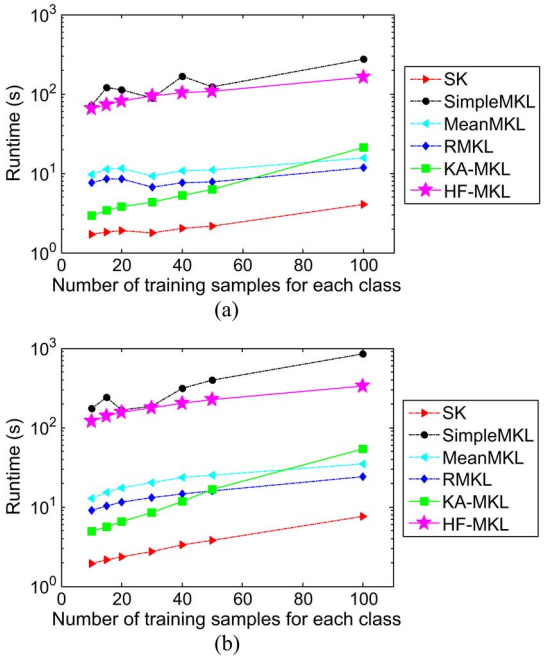

Fig. 14. Comparison of the runtime for each classifier. (a) Runtime of classifiers for the Bayview Park data set. (b) Runtime of classifiers for the Recology data set.

significance, with almost all the values of Z-test being greater than 1.96.

3) The analysis of computational load

In order to illustrate classification efficiency, the computational time of different methods for the two data sets is shown in Fig. 14. Compared with SK, RMKL, and Mean MKL, the proposed HF-MKL increases the computational burden, due to the need to compute a large number of kernel matrices. However, the computational time of the proposed method HF-MKL is still less than that of the traditional one-step method Simple MKL.

4) The thematic maps of the classification results

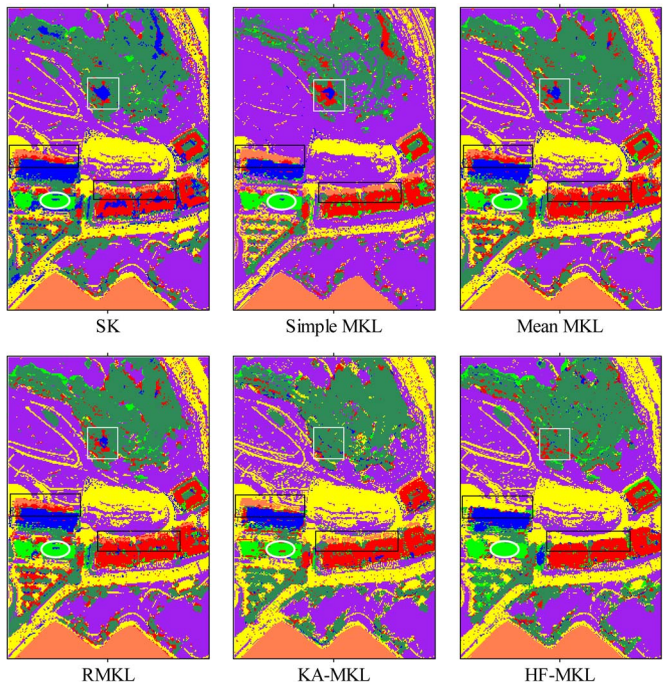

Fig. 15. Thematic map of the Bayview Park data set.

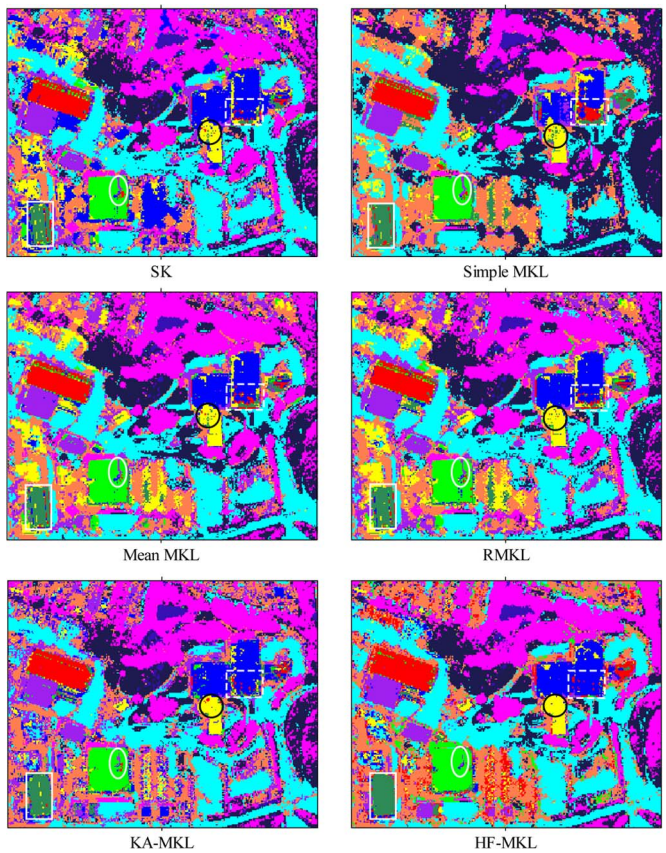

Fig. 16. Thematic map of the Recology data set.

The thematic maps of the full images with 50 training samples per class are shown in Figs. 15 and 16 for the two images, respectively. From the thematic maps, we can visually see that HF-MKL has the best performance compared with the other

TABLE VI
CLASSIFICATION ACCURACY PER CLASS

	Method	Mean MKL	KA-MKL	HF-MKL
Accuracy per class (%)	Building1	73.52	78.12	83.61
	Building2	68.24	74.72	84.13
	Building3	73.85	69.93	67.14
	Roads	63.07	55.36	66.53
	Trees	84.47	87.09	84.26
	Soil	91.51	92.91	95.22
	Seawater	98.87	99.96	1
OA (%)		83.51	84.25	85.91

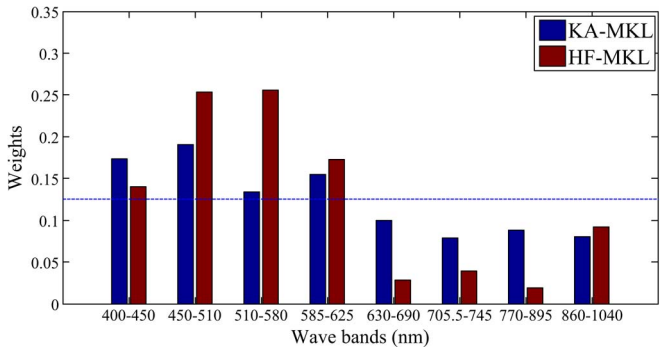

Fig. 17. Weights of different methods (only using spectrum).

TABLE VII
CLASSIFICATION ACCURACY PER CLASS

	Method	Mean MKL	KA-MKL	HF-MKL
Accuracy per class (%)	Building1	88.46	89.39	90.05
	Building2	93.13	93.99	92.99
	Building3	85.74	83.90	91.08
	Roads	96.31	96.57	95.47
	Trees	93.40	92.82	95.12
	Soil	94.70	94.53	94.9
	Seawater	1	1	1
OA (%)		93.62	93.42	94.74

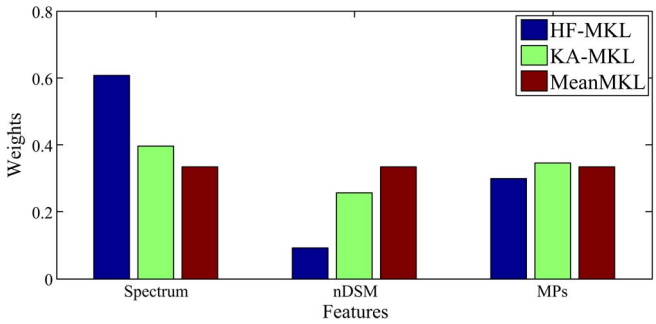

Fig. 18. Weights of the spectral feature, nDSM, and MPs.

classifiers. The different buildings are correctly separated by using the information extracted from LiDAR data. Specifically, the areas marked by rectangles and ellipses in the thematic maps are significantly improved by HF-MKL compared with the other classifiers. From the thematic maps of the Bayview Park data set, we can see that the tree area marked with white rectangle is misclassified into Building 1 and Building 3 by the traditional methods, and the proposed KA-MKL and HF-MKL significantly improve the classification result. The shadow areas induced by buildings are misclassified into seawater by the control methods, and only the proposed method can reduce the influence of shadow and correctly discriminate the target. The ellipse shows more significantly improved areas by HF-MKL compared with the other methods. Looking at the classification maps of the Recology data set, HF-MKL significantly improves the classification accuracy for Building 2, Building 3, Building 4, and Building 5, as marked by rectangles and ellipses.

4) Analysis for the Weights of Different Features: In this part, we analyze the relation between weights and accuracy per class by analyzing the relationship between reflectivity of different wave bands and different land covers. We set the training number per class to 20 in the experiments.

a) Bayview park data set: The accuracy per class is shown in Table VI, in which we only used the spectral features and each band corresponds to a feature.

The weights corresponding to spectral features are shown in Fig. 17. The dashed line represents the mean value of weight or the weights of Mean MKL.

All of the spectral bands are regarded as a characteristic, MPs as a feature, and nDSM as a characteristic; and the accuracy per class is shown in Table VII.

TABLE VIII
CLASSIFICATION ACCURACY PER CLASS

	Method	Mean MKL	KA-MKL	HF-MKL
Accuracy per class (%)	Building1	72.26	67.83	75.00
	Building2	49.37	48.30	51.43
	Building3	51.28	50.57	57.85
	Building4	45.39	46.72	50.85
	Building5	59.17	60.49	66.35
	Building6	67.19	54.64	58.82
	Building7	59.18	61.22	60.54
	Trees	91.88	92.82	93.64
	Parking Lot	60.81	70.05	75.61
	Soil	89.87	90.57	92.05
	Grass	92.02	93.02	95.37
OA (%)		69.80	69.99	73.58

The weights corresponding to spectrum, nDSM, and MPs are shown in Fig. 18.

b) Recology data set: We only used the spectral features, and each band corresponds to a feature, and the classification accuracy per class is shown in Table VIII.

The weights corresponding to spectral features are shown in Fig. 19. All of the spectral bands are regarded as a characteristic, MPs as a feature, and nDSM as a characteristic; and the joint classification accuracy per class is shown in Table IX.

The weights corresponding to the spectrum, nDSM, and MPs are shown in Fig. 20.

From the view of HF-MKL and KA-MKL, the results observed from Fig. 18 can be summarized as follows.

1) In the classification with the Bayview Park data set, spectral information provides more information for classification than LiDAR data.

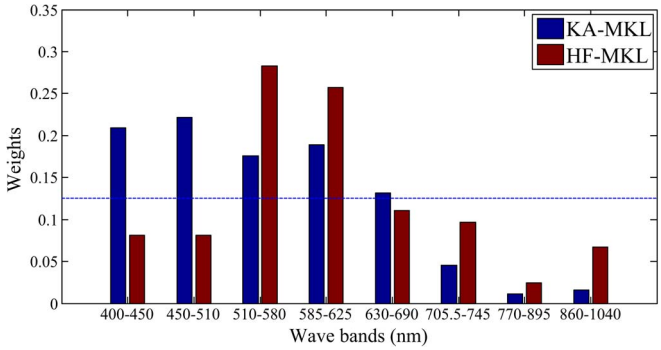

Fig. 19. Weights of different methods (only using the spectral feature).

TABLE IX
CLASSIFICATION ACCURACY PER CLASS

	Method	Mean MKL	KA-MKL	HF-MKL
Accuracy per class (%)	Building1	83.40	83.49	82.36
	Building2	80.38	83.06	83.51
	Building3	58.12	59.03	62.00
	Building4	76.16	74.21	80.78
	Building5	71.46	69.81	71.83
	Building6	82.35	84.29	84.58
	Building7	96.60	98.64	95.92
	Trees	91.61	92.20	94.21
	Parking Lot	95.80	95.69	96.71
	Soil	96.12	96.12	97.04
	Grass	96.90	96.9	97.67
	OA (%)	83.38	84.39	85.45

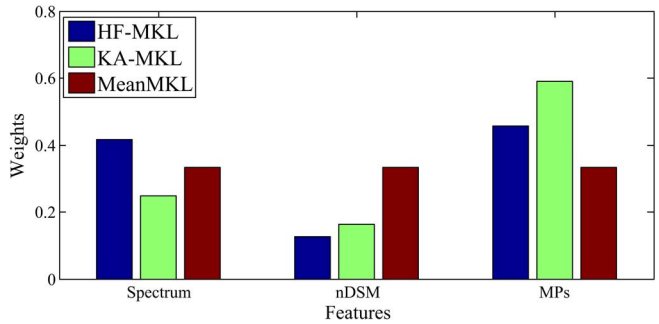

Fig. 20. Weights of the spectral feature, nDSM, and MPs.

2) From the viewpoint of spectral, spatial, and elevation attributes, the maximum effect is spectral information, then spatial information, and the minimum is elevation information.

From Fig. 20, we can observe that, in the classification with the Recology data set, spatial attributes provide more information than spectral attributes. The maximum effect is spatial information, then spectral information, and the minimum is elevation information. These conclusions are consistent with the results of only using the features of spectrum, MPs, and nDSM, respectively. By comparing Fig. 18 with Fig. 20, we observe that spectral attributes provide the most information in the Bayview Park data set and spatial attributes provide the most information in the Recology data set. We find that the Recology data set provides more spatial information than the Bayview Park by observing the two sets of data.

From Table VI, we can see that seawater gets the highest classification accuracy. We know that water reflectivity increases with the increase of wavelength at 0–550 nm and reduces with the increase of wavelength that is more than 550 nm. In other words, water reaches the maximum reflectance at about 550 nm. From Fig. 17, we can see that HF-MKL and KA-MKL reach greater weight in the band around 550 nm used for corresponding features. Therefore, the two methods make full use of the physical properties of water and consequently obtain higher classification accuracies than Mean MKL.

From Table VIII, we can observe that the classification accuracy of trees, grass, and soil using HF-MKL is higher than that using KA-MKL and Mean MKL. We know that the reflectivity of trees, grass, and soil roughly increases with the increase of wavelength. From Fig. 19, we find that HF-MKL's weights distribution biased in favor of long wavelength, whereas the weights distribution of KA-MKL tends to short wavelength. Therefore, HF-MKL makes full use of the physical characteristics of trees, grass, and soil with the large long-wave reflectivity, thus getting high classification accuracy.

Due to the complexity of building materials, the reflectance for the different wavelengths of light is difficult to measure. Thus, it is not easy to analyze the relationship between weights distribution and classification accuracy.

D. Discussion

From the analysis of the obtained results in experiment 1, it can be seen that, generally speaking, classification accuracy obtained only using features from spectral images is higher than using only the one from LiDAR data. It can be concluded that, for classification, spectral images play a more important role compared with LiDAR data. Features from spectral images are good at identifying ground truth such as trees, grass, and soil; whereas features from LiDAR data perform better for classes with flat surfaces such as buildings. From experiments 1 and 2, it is obvious that significant improvements are achieved for all the cases when both features from spectral images and LiDAR data are used together. This point confirms the complementary relationship between spectral images and LiDAR data for classification.

From experiments 1–3, it can be seen that HF-MKL in our work reaches the best classification result among the compared methods with different feature sets. This fact shows the effectiveness of our features-extracting framework for joint use of spectral images and LiDAR data. In other words, more useful information has been extracted compared with the other methods. It can be seen that the HF-MKL classification method outperforms the SVM-based classifier presented in [14] in every case with changes in number of training samples. Significantly improved classification accuracy can be achieved, although the number of labeled training samples is small. This result indicates the excellent ability of our proposed method to handle a complex features set by integrating multiscale information with individual features and properly arranging the combination of different features instead of treating them equally and in a stacked input vector for the kernels. In addition, our method

is proved to be an effective tool for combined use of spectral images and LiDAR data for classification.

VI. Conclusion

In this paper, urban area classification using spectral images and LiDAR data has been investigated. In greater detail, a feature-extracting strategy and an MKL classifier based on two levels are proposed to improve the joint effect of the aforementioned two data sources. From an analysis on the experimental results on two real data sets, some conclusions can be made.

1) The combined use of spectral images and LiDAR data can significantly improve the classification accuracy for the reason that these two data sources have complementary relationship.
2) The feature-extraction strategy used in our work can extract more discriminative information from these sources, and it is more effective than the other tested methods.
3) As an advanced classification technique, the proposed MKL framework can arrange weights for different features and integrate kernels with different scales. For these two reasons, the HF-MKL-based SVM method outperforms the state-of-the art classification methods in the field of joint use of spectral information and LiDAR data.

In terms of future work, we are planning to do some analysis on the determination of scales range for basis kernels and develop a supervised features ranking method in the framework of MKL by introducing sparse theory.

Acknowledgment

The authors would like to thank Digital Globe and the IEEE for providing the data sets that cover the Bayview Park and Recology areas data sets used in this study and the IEEE GRSS Data Fusion Technical Committee for organizing the 2012 Data Fusion Contest.

References

[1] 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.
[2] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.
[3] G. Shaw and D. Manolakis, "Signal processing for hyperspectral image exploitation," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 12–16, Jan. 2002.
[4] N. Haala and C. Brenner, "Extraction of buildings and trees in urban environments," *ISPRS J. Photogramm. Remote Sens.*, vol. 54, no. 2, pp. 130–137, Jul. 1999.
[5] S. Filin, "Surface clustering from airborne laser scanning data," in *Proc. ISPRS Commission III Symp. Photogramm. Comput. Vis.*, 2002, pp. 119–124.
[6] S. Filin and N. Pfeifer, "Segmentation of airborne laser scanning data using a slope adaptive neighborhood," *ISPRS J. Photogramm. Remote Sens.*, vol. 60, no. 2, pp. 71–80, Apr. 2006.
[7] I. Lee, and T. Schenk, "3-D perceptual organization of laser altimetry data," *Int. Archives Photogramm. Remote Sens.*, vol. 34. pp. 57–65, 2001.
[8] M. Dalponte, L. Bruzzone, and D. Gianelle, "Fusion of hyperspectral and LIDAR remote sensing data for classification of complex forest areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1416–1427, May 2008.
[9] R. Sugumaran and M. Voss, "Object-oriented classification of LIDAR fused hyperspectral imagery for tree species identification in an urban environment," in *Proc. Urban Remote Sens. Joint Event*, Paris, France, Apr. 11–13, 2007, pp. 1–6.
[10] E. Puttonen, A. Jaakkola, P. Litkey, and J. Hyyppä, "Tree classification with fused mobile laser scanning and hyperspectral data," *Sensors*, vol. 11, no. 5, pp. 5158–5182, 2011.
[11] R. Dinuls, G. Erins, A. Lorencs, I. Mednieks, and J. Sinica-Sinavskis, "Tree species identification in mixed baltic forest using LIDAR and multispectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 594–603, Apr. 2012.
[12] D. Lemp and U. Weidner, "Improvements of roof surface classification using hyperspectral and laser scanning data," in *Proc. ISPRS Joint Conf.: 3rd Int. Symp. Remote Sens. Data Fusion Over Urban Areas (URBAN), 5th Int. Symp. Remote Sens. Urban Areas (URS)*, Tempe, AZ, USA, Mar. 14–16, 2005, pp. 1–6.
[13] A. R. Elshehaby and L. G. Taha, "A new expert system module for building detection in urban areas using spectral information and LiDAR data," *Appl. Geomatics.* vol. 1, no. 4, pp. 97–110, Dec. 2009.
[14] M. Pedergnana, P. R. Marpu, M. D. Mura, J. A. Benediktsson, and L. Bruzzone, "Classification of remote sensing optical and LiDAR data using extended attribute profiles," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 7, pp. 856–865, Nov. 2012.
[15] M. García, D. Riaño, E. Chuvieco, J. Salas, and F. M. Danson, "Multispectral and LiDAR data fusion for fuel type mapping using support vector machine and decision rules," *Remote Sens. Environ.* vol. 115, no. 6, pp. 1369–1379, Jun. 2011.
[16] B. Koetz, F. Morsdorf, S. van der Linden, T. Curt, and B. Allgower, "Multi-source land cover classification for forest fire management based on imaging spectrometry and LiDAR data," *Forest Ecology Manage.*, vol. 256, no. 3, pp. 263–271, Jul. 2008.
[17] L. Liu, Y. Pang, W. Fan, Z. Li, and M. Li, "Fusion of airborne hyperspectral and LiDAR data for tree species classification in the temperate forest of northeast China," in *Proc. 19th Int. Conf. Geoinformatics*, 2011, pp. 1–5.
[18] G. Camps-Valls, L. Gómez-Chova, J. Muñoz-Marí, J. Luis Rojo-Álvarez, and M. Martínez-Ramón, "Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection," *IEEE Tran. Geosci. Remote Sens.*, vol. 46, no. 6, pp. 1822–1835, Jun. 2008.
[19] G. R. G. Lanckriet, N. Cristianini, P. Bartlett, L. E. Ghaoui, and M. I. Jordan, "Learning the kernel matrix with semidefinite programming," *J. Mach. Learn. Res.*, vol. 5, pp. 27–72, Dec. 2004.
[20] G. Lanckriet, T. De Bie, N. Cristianini, M. Jordan, and W. Noble, "A statistical framework for genomic data fusion," *Bioinformatics*, vol. 20, no. 16, pp. 2626–2635, Nov. 2004.
[21] 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.
[22] S. Sonnenburg, C. S. G. Rätsch, and B. Schölkopf, "Large scale multiple kernel learning," *J. Mach. Learn. Res.*, vol. 7, no. 1, pp. 1531–1565, Jul. 2006.
[23] Y. Gu *et al.*, "Representative multiple-kernel learning for classification of hyperspectral imagery," *IEEE Tran. Geosci. Remote Sens.*, vol. 50, no. 7, pp. 2852–2865, Jul. 2012.
[24] 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.
[25] 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. 42, no. 3, pp. 480–491, Mar. 2005.
[26] X. Jia, B.-C. Kuo, and M. M. Crawford, "Feature mining for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 3, pp. 676–697, Mar. 2013.
[27] M. Gönen and E. Alpaydin, "Multiple kernel learning algorithms," *J. Mach. Learn. Res.*, vol. 12, pp. 2211–2268, Jul. 2011.
[28] A. Rakotomamonjy, F. R. Bach, S. Canu, and Y. Grandvalet, "SimpleMKL," *J. Mach. Learn. Res.*, vol. 9, pp. 2491–2521, Nov. 2008.
[29] G. M. Foody, "Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 5, pp. 627–633, May 2004.
[30] B. Bostanci and E. Bostanci, "An evaluation of classification algorithms using Mc Nemar's test," in *Proc. 7th Int. Conf. BIC-TA*, pp. 15–26, Jan. 2013.

Yanfeng Gu (M'06) received the Ph.D. degree in information and communication engineering from the Harbin Institute of Technology (HIT), Harbin, China, in 2005.

He joined, as a Lecturer, the School of Electronics and Information Engineering, HIT, where he was appointed as an Associate Professor in 2006 and is currently a Professor with the Department of Information Engineering. He was enrolled in the first Outstanding Young Teacher Training Program at HIT. From 2011 to 2012, he was a Visiting Scholar with the Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA, USA. He has authored or coauthored over 60 peer-reviewed papers and four book chapters. He is the inventor or the coinventor of seven patents. His research interests include image processing in remote sensing, machine learning and pattern analysis, and multiscale geometric analysis.

Dr. Gu is a Peer Reviewer for several international journals such as the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, and the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS and the *IET Electronics Letter*.

Qingwang Wang (S'14) was born in Zhaotong, China, in 1990. He received the B.E. degree in electronics and information engineering from the Harbin Institute of Technology, Harbin, China, in 2014. He is currently working toward the Ph.D. degree in information and communication engineering at the Harbin Institute of Technology.

His research interests include machine learning and its application to remote sensing data analysis. In particular, his studies currently focus on kernel methods and classification of hyperspectral image and joint classification for hyperspectral image and LiDAR data.

Xiuping Jia (M'93–SM'03) received the B.Eng. degree from Beijing University of Posts and Telecommunications, Beijing, China, in 1982, and the Ph.D. degree in electrical engineering from The University of New South Wales, Canberra, Australia, in 1996.

Since 1988, she has been with the School of Information Technology and Electrical Engineering, The University of New South Wales, where she is currently a Senior Lecturer. She is also a Guest Professor with Harbin Engineering University, Harbin, China. She has coauthored the remote sensing textbook *Remote Sensing Digital Image Analysis* [Springer-Verlag, 3rd (1999) and 4th eds. (2006)]. Her research interests include remote sensing and imaging spectrometry.

Dr. Jia is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and CPGIS Annals of GIS.

Jón Atli Benediktsson (S'84–M'90–SM'99–F'04) received the Cand.Sci. degree in electrical engineering from the University of Iceland, Reykjavik, Iceland, in 1984, and the M.S.E.E. and Ph.D. degrees from Purdue University, West Lafayette, IN, USA, in 1987 and 1990, respectively.

He is currently a Pro Rector for Academic Affairs and a Professor of electrical and computer engineering with the University of Iceland. His research interests are in remote sensing, biomedical analysis of signals, pattern recognition, image processing, and signal processing. He has published extensively in those fields.

Prof. Benediktsson is the 2011 President of the IEEE Geoscience and Remote Sensing Society (GRSS), and he has been on the GRSS Administrative Committee since 2000. He was an Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING from 2003 to 2008. He was recipient of the Stevan J. Kristof Award from Purdue University in 1991 as outstanding graduate student in remote sensing. In 1997, he was the recipient of the Icelandic Research Council's Outstanding Young Researcher Award; in 2000, he was a recipient of the IEEE Third Millennium Medal; in 2004, he was a corecipient of the University of Iceland's Technology Innovation Award; in 2006, he was a recipient of the yearly research award from the Engineering Research Institute of the University of Iceland; in 2007, he was a recipient of the Outstanding Service Award from the IEEE GRSS. He is also a member of Societas Scinetiarum Islandica and Tau Beta Pi.