A Novel Feature Fusion Approach for VHR Remote Sensing Image Classification

Sicong Liu, *Member, IEEE*, Yongjie Zheng, Qian Du, *Fellow, IEEE*, Alim Samat, *Member, IEEE*, Xiaohua Tong, *Senior Member, IEEE*, and Michele Dalponte, *Senior Member, IEEE*

***Abstract*—This article develops a robust feature fusion approach to enhance the classification performance of very high resolution (VHR) remote sensing images. Specifically, a novel two-stage multiple feature fusion (TsF) approach is proposed, which includes an intragroup and an intergroup feature fusion stages. In the first fusion stage, multiple features are grouped by clustering, where redundant information between different types of features is eliminated within each group. Then, features are pairwisely fused in an intergroup fusion model based on the guided filtering method. Finally, the fused feature set is imported into a classifier to generate the classification map. In this work, the original VHR spectral bands and their attribute profiles are taken as examples as input spectral and spatial features, respectively, in order to test the performance of the proposed TsF approach. Experimental results obtained on two QuickBird datasets covering complex urban scenarios demonstrate the effectiveness of the proposed approach in terms of generation of more discriminative fusion features and enhancing classification performance. More importantly, the fused feature dimensionality is limited at a certain level; thus, the computational cost will not be significantly increased even if multiple features are considered.**

***Index Terms*—Classification, feature fusion, guided filtering (GF), spectral-spatial features, very high resolution (VHR) image.**

I. Introduction

THE current development of very high resolution (VHR) remote sensing satellites allows the acquisition of submeter extremely high spatial resolution images. This provides great opportunities to enhance the earth's surface mapping at a very detailed level, as land-cover classification can benefit from this in many different remote sensing application fields, e.g., urban, agriculture, disaster mapping, and forestry [1]–[6]. Unlike the traditional moderate resolution multispectral images, VHR images are characterized by a higher spatial detail of land objects, so context coherence and spatial patterns are as important as spectral information in the classification process in order to produce an accurate land-cover thematic map. In this article, many advanced techniques have been proposed to utilize multiple features, especially spectral-spatial features and to improve the classification or detection performance, such as the morphological reconstruction [7], the attribute profiles (AP) [8], the edge-preserving filtering [9], the superpixel segmentation [10], and the convolutional neural networks [11]–[14]. Despite their effectiveness in extracting multiscale spectral-spatial features, many of them do not address the feature fusion problem due to increasing feature complexity and computational cost, which may limit their utilization in practical applications.

Information fusion plays a very important role in remote sensing processing and application, either from the data, feature, or the decision level [15]–[18]. For feature-level fusion, a simple fusion strategy is the direct feature stacking [1], [19], [20], [21]. However, this may lead to information redundancy and to a great increase of computational cost. In some sense, features are cascaded, not really "fused" intrinsically. Several other works were carried out to design advanced fusion models. In [22], a fusion method based on feature transformation was proposed, where the spectral bands, the Gabor features, and the pixel-shaped features were unified, and then, manifold learning was applied to extract a low-dimensional representation of the stacking features. In [23], a multifeature fusion framework was developed to combine the original spectral features with extended morphological profile features based on the multiscale spatial and spectral kernels. In [24], a probabilistic weighted strategy was proposed for spectral-spatial feature fusion using multiple classifiers on different features. Although the manifold learning and the multikernel learning methods have improved the nonlinear discriminability, they did not consider the physical meaning of features and may lose a part of the original information. Moreover, the processing time may significantly increase, especially in the commonly used kernel-based methods, where the dimensionality of the data space is still high and the selection of proper kernels is also difficult [8], [25]. This is the first problem that needs to be properly addressed.

Taking into account the advantage of the spatial features, edge-preserving filtering methods become more noticeable. Among them, the guided filtering (GF) shows a good performance for edge-preserving that has been widely used in image

Manuscript received August 31, 2020; revised November 13, 2020; accepted November 22, 2020. Date of publication December 2, 2020; date of current version January 6, 2021. This work was supported in part by the National Key R&D Program of China under Grant 2018YFB0505000, and in part by the Natural Science Foundation of China under Grant 42071324, Grant 41601354, Grant 42071424, and Grant 42001387. *(Corresponding authors: Qian Du; Xiaohua Tong.)*

Sicong Liu, Yongjie Zheng, and Xiaohua Tong are with the College of Surveying and Geoinformatics, Tongji University, Shanghai 200092, China (e-mail: sicong.liu@tongji.edu.cn; yongjie@tongji.edu.cn; xhtong@tongji.edu.cn).

Qian Du is with the College of Surveying and Geoinformatics, Tongji University, Shanghai 200092, China, and also with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS 39762 USA (e-mail: du@ece.msstate.edu).

Alim Samat is with the Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China (e-mail: alim.smt@gmail.com).

Michele Dalponte is with the Department of Sustainable Agro-Ecosystems and Bioresources, Research and Innovation Centre, Fondazione E. Mach, 38010 San Michele all'Adige, Italy (e-mail: michele.dalponte@fmach.it).

Digital Object Identifier 10.1109/JSTARS.2020.3041868

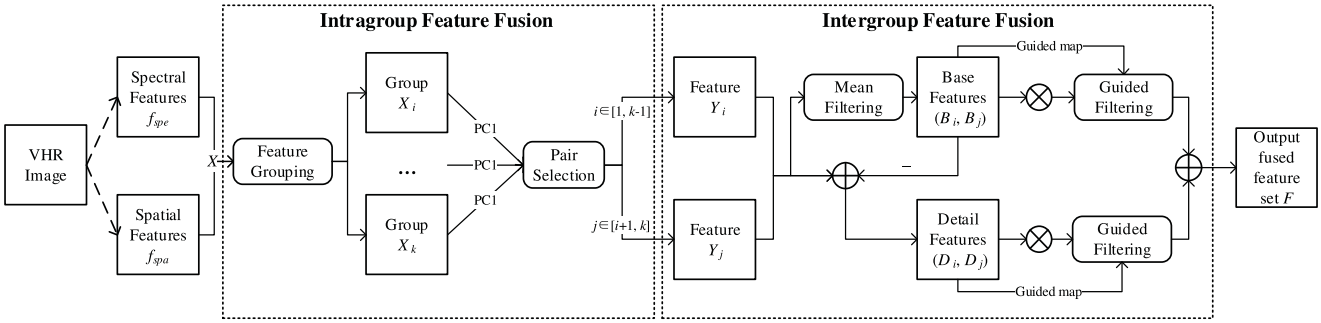

Fig. 1. Block scheme of the proposed TsF approach.

fusion. With the help of the guidance image (e.g., the geometric and spatial features of image objects), GF can make the filtering output more structured and smoother than the input [26]. In practical applications, GF can be directly used as a fusion algorithm, or as a part of fusion models to further improve the fusion performance. For example, in [27], a guided filtering-based weighted average fusion method (GFF) was proposed based on double-scale decomposition of images for fusing multifocus, multimodal, and multiexposure pictures. It can make full use of the spatial consistency to fuse the base layers and the detail layers. Note that in [27], the considered images are quite similar to each other that are imported to GFF, since they are multifocus images. However, for multispectral VHR remote sensing images, different kinds of features are often characterized in significantly different representations. Therefore, the selection of proper features as guidance and input image in GF-based methods will lead to very different fusion results, which usually are more similar to the input image. This is the second open issue in the current development of GF-based fusion methods [28], [29].

To solve the aforementioned problems, a novel two-stage multiple feature fusion (TsF) approach for VHR image classification is proposed in this article. The proposed approach aims to take full advantage of the information representation in multiple features. Experimental results obtained on two QuickBird (QB) VHR datasets covering complex urban scenarios demonstrated its effectiveness, compared with the state-of-the-art methods. The main novelty and contributions of this article can be summarized as follows.

1) The proposed TsF approach can preserve discriminative information via the intragroup fusion step, and capture more significant information via the intergroup fusion step, respectively. Therefore, fusion results are not selection-driven but feature-driven.
2) The dimensionality of the fused feature set is controlled at a certain level according to the fusion step, and this resulted in a limited increase in dimensionality with respect to other multifeature-based methods. From the computational cost point of view, the proposed approach is also competitive.
3) The proposed fusion model is not feature-specific; thus, theoretically, it is suitable for fusing any kinds of features.

The remainder of the article is organized as follows. The proposed TsF feature fusion method is described in Section II. Experimental results and analysis are presented in Section III. Finally, Section IV draws the conclusions.

II. Proposed Two-Stage Multiple Feature Fusion Approach

Based on the considered multiple features, the proposed TsF approach mainly consists of the following two processing steps: 1) intragroup feature fusion; and 2) intergroup feature fusion. Its block scheme is shown in Fig. 1, and details are provided as follows.

A. Step 1: Intragroup Feature Fusion

In order to model the feature integration problem, in this article, we take spatial and spectral features as an example to generate a multiple feature set X:

$$X = \{f_{\mathrm{spe}}, f_{\mathrm{spa}}\}, X \in \mathbb{R}^{h \times w \times n} \tag{1}$$

where $h \times w$ represents the feature size, and n is the number of features, and f_{spe} and f_{spa} denote the spectral and spatial features, respectively. For f_{spe}, raw bands are directly used. We considered the AP as f_{spa}, since the morphological features are capable to preserve spatial information [30]. In particular, the f_{spa} is built based on the attribute: the length of the diagonal of the bounding box (d). As a result, based on the original spectral feature f_{spe}, its corresponding f_{spa} is calculated as

$$f_{\mathrm{spa}} = \{\gamma_{\lambda_d}(f_{\mathrm{spe}}), \varphi_{\lambda_d}(f_{\mathrm{spe}})\} \tag{2}$$

where γ and φ represent attribute thinning and thickening operations, respectively, and λ_d is the predefined threshold for d with a value of 200 in this work. Note that the dimensionality of AP is usually high due to the fact that the multiscale and the multiattribute are jointly modeled.

Efficient extraction of the low-redundancy and low-dimensional spatial information features is preferable when classifying VHR imagery, especially for a large study area [1]. To this end, k-means clustering is first applied to divide the whole feature set X into k groups, resulting in

$$X = \{X_1, X_2, \ldots, X_k\} \tag{3}$$

where $X_i \in \mathbb{R}^{h \times w \times n_i}$, $\sum_{i=1}^{k} n_i = n$.

In literature, the mutual information (MI) has been widely applied for feature selection in remote sensing image processing [31], [32]. It represents the maximal relevance criterion between

feature and class label. MI can effectively measure the interdependence between two given features (e.g., x_i, x_j) in X from the entropy point of view. Therefore, the values of MI and original pixels are selected as the input for feature grouping.

$$MI(x_i, x_j) = H(x_i) + H(x_j) - H(x_i, x_j) \tag{4}$$

where $H(x_i)$ and $H(x_j)$ represent the entropy of feature x_i and x_j, respectively, and $H(x_i, x_j)$ is the joint entropy of x_i and x_j.

In order to avoid the random selection of an initial class center X_{c1}, we manually assign it according to the equidistance principle [see (5)]. In this work, the MI matrix computed by different features is used as the input for k-means clustering to generate a more accurate class center X_{c2}. The final grouping feature $\{X_1, X_2, \ldots, X_k\}$ is then clustered by the pixel values of X and the class center X_{c2}

$$X_{c1} = \{x_s, x_{2s}, \ldots, x_{ks}\}, s = \left|\frac{n}{k}\right|. \tag{5}$$

After that, the principal component analysis (PCA) is adopted to reduce the dimensionality of each intragroup feature subset by retaining the first principal component (PC1). Then, the intragroup fusion feature set Y can be built as

$$Y = \{\mathrm{PC1}(X_1), \mathrm{PC1}(X_2), \ldots, \mathrm{PC1}(X_k)\}. \tag{6}$$

It is worth noting that the intragroup feature fusion step in the proposed TsF approach can retain more distinctive information represented in different features, while reducing the input feature dimensionality.

B. Step 2: Intergroup Feature Fusion

This step aims to further combine the intragroup fusion result and integrate significant information representations in different types of features. To this end, the intragroup fusion feature sets Y are pairwisely fused in the intergroup fusion step based on the GF algorithm. Following are the details.

1) *Double-scale feature decomposition:* In order to make full use of the features complementarity in different layers, the input feature pair [Y_i, Y_j] are first decomposed into base layers [B_i, B_j] and detail layers [D_i, D_j] as follows:

$$\begin{cases} B_i = Y_i * z, & B_j = Y_j * z \\ D_i = Y_i - B_i, & D_j = Y_j - B_j \end{cases} \tag{7}$$

where $i \in [1, k-1]$, $j \in [i+1, k]$, B and D represent the base layer and detail layer of the input feature, respectively, z is the moving window size of the mean filter, and '$*$' represents the convolution.

2) *Double-scale feature fusion with GF:* GF assumes that the filtering output O is a linear transformation of the guidance image I in a local window ω_j with the center pixel j.

$$O_i = a_j G_i + b_j, \forall i \in \omega_j \tag{8}$$

where ω is equal to $(2r+1)^2$, r is the window parameter that needs to be defined in advance, and O_i (G_i) represents the value in pixel i. Then, to satisfy the minimum difference between O_i and the input image, the optimal values of two linear coefficients a_j and b_j in each ω_j can be calculated by minimizing the following cost function as:

$$E(a_j, b_j) = \sum_{i \in \omega_j} \left[(a_j G_i + b_j - I_i)^2 + \delta {a_j}^2 \right] \tag{9}$$

where δ represents the regularization parameter. Hence, the final O_i can be calculated as

$$O_i = \bar{a}_i G_i + \bar{b}_i \tag{10}$$

where $\bar{a}_i$ and $\bar{b}_i$ are the average values of a_j and b_j in all windows overlapping i, respectively.

To retain the saliency information inherited from different features and layers, the input maps of GF is generated from the base (detail) layers based on the multiple fusion. Then, the base (detail) layers are selected as the guidance maps in GF. Finally, the GF fusion features can be calculated as

$$\begin{cases} \overline{B} = G_{r,\delta}(B_i \times B_j, B_i) + G_{r,\delta}(B_i \times B_j, B_j) \\ \overline{D} = G_{r,\delta}(D_i \times D_j, D_i) + G_{r,\delta}(D_i \times D_j, D_j) \end{cases} \tag{11}$$

where $\overline{B}$ and $\overline{D}$ represent the fusion results of the base and detail average features, respectively, $G_{r,\delta}(P, I)$ represents the GF algorithm with two parts: the input map $P = \{B_i \times B_j, D_i \times D_j\}$ and the guided map $I = \{B_i, B_j, D_i, D_j\}$.

3) *Double-scale feature reconstruction:* In this last step, both fusion results of the base layer $\overline{B}$ and detail layer $\overline{D}$ average features are summed to generate the final fusion feature F.

$$F = \overline{B} + \overline{D} \tag{12}$$

where $F \in \mathbb{R}^{h \times w \times m}$. Note that $m = k(k-1)/2$ is the total number of fused features.

III. Experimental Results and Analysis

A. Datasets Description

Experiments were conducted on two multispectral VHR remote sensing images acquired by the QB satellite over urban areas of the city of Zurich, Switzerland. Two scene images, including Zurich 1 (with the pixel sizes of 1295 × 1364) and Zurich 2 (with the pixel sizes of 833 × 881) were considered in the experiments, which are denoted in the rest of the article as ZH1 and ZH2, respectively. Note that the original four multispectral bands (blue, green, red, and near-infrared) were fused with the panchromatic band by using the Gram–Schmidt (G-S) algorithm to generate the pan-sharpened results having an approximate resolution of 0.62 meter. Figs. 2 and 3 present the false color composite image (a) and the reference map (b) of the two multispectral VHR images. The ZH1 scene contains seven land-cover classes, including roads, buildings, trees, grass, bare soil, railways, and swimming pools, as shown in Fig. 2(b). The ZH2 dataset contains four land-cover classes (i.e., roads, buildings, trees, and grass), as shown in Fig. 3(b). Details of the reference samples for each class in two datasets are provided in Table I.

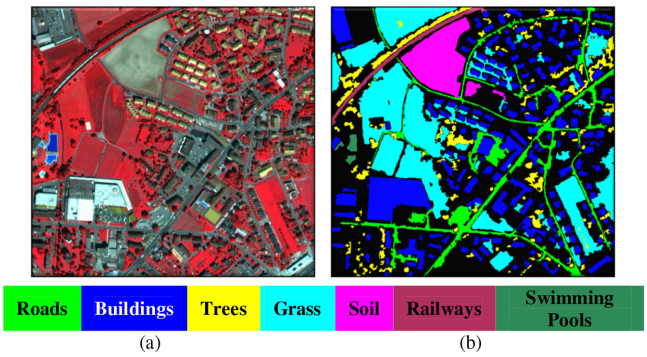

Fig. 2. ZH1 scene of QB dataset. (a) False color composite image (RGB: near-infrared, red, and green bands). (b) Reference sample map.

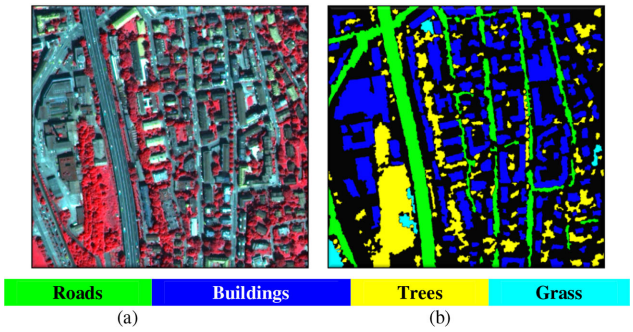

Fig. 3. ZH2 scene of QB dataset. (a) False color composite image (RGB: near-infrared, red, and green bands). (b) Reference sample map.

TABLE I
DETAILED REFERENCE SAMPLE INFORMATION IN TWO DATASETS

Classes	Number of samples (pixels)	
	ZH1	ZH2
Roads	120532	86551
Buildings	251222	154164
Trees	79153	81033
Grass	311833	7201
Bare soil	72429	-
Railways	16043	-
Swimming pools	5070	-

B. Parameter Settings

The proposed feature fusion TsF approach was evaluated according to the classification performance by using a support vector machine (SVM) classifier, where the radial basis function (RBF) was selected as the kernel function.

In the intragroup fusion step, the feature grouping was constructed by assigning the group numbers k. Therefore, a detailed quantitative analysis and performance evaluation based on the overall accuracy (OA) and computational time cost (T) indices was conducted under different k values. In order to obtain a reliable conclusion under different input conditions, the classification performances were compared by changing the number of training samples, which were selected as 20, 50, 100, 200, 500, 1000, and 2000 pixels for each class. The results and comparison are shown in Figs. 4 and 5 for ZH1 and ZH2 scene datasets, respectively. It is worth noting that the OA achieves a rapid increase with k in the range of [1, 6], and tends to

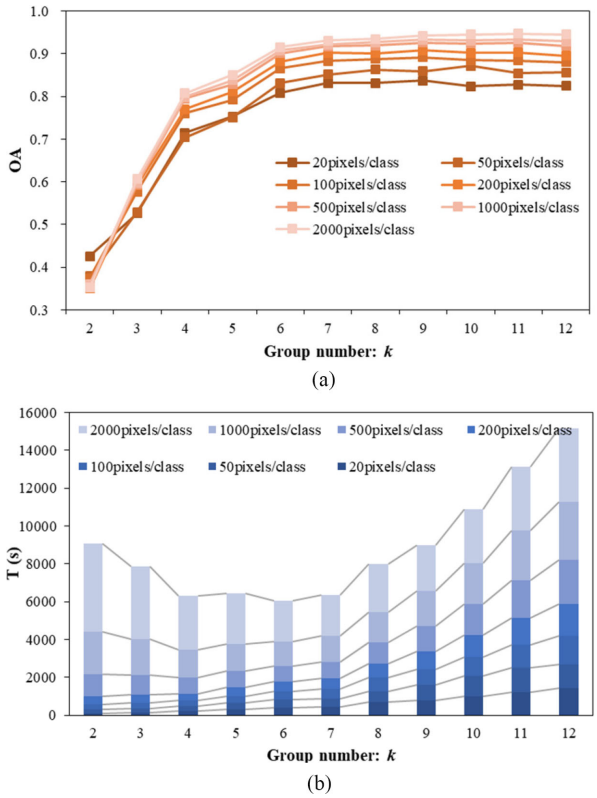

Fig. 4. Comparison of (a) OA and (b) T obtained by different group numbers k on ZH1 scene dataset.

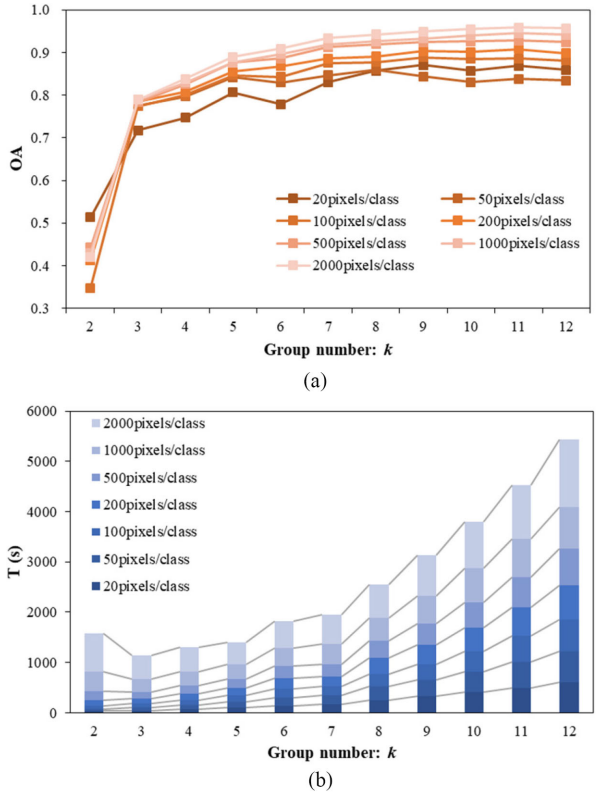

Fig. 5. Comparison of (a) OA and (b) T obtained by different group numbers k on ZH2 scene dataset.

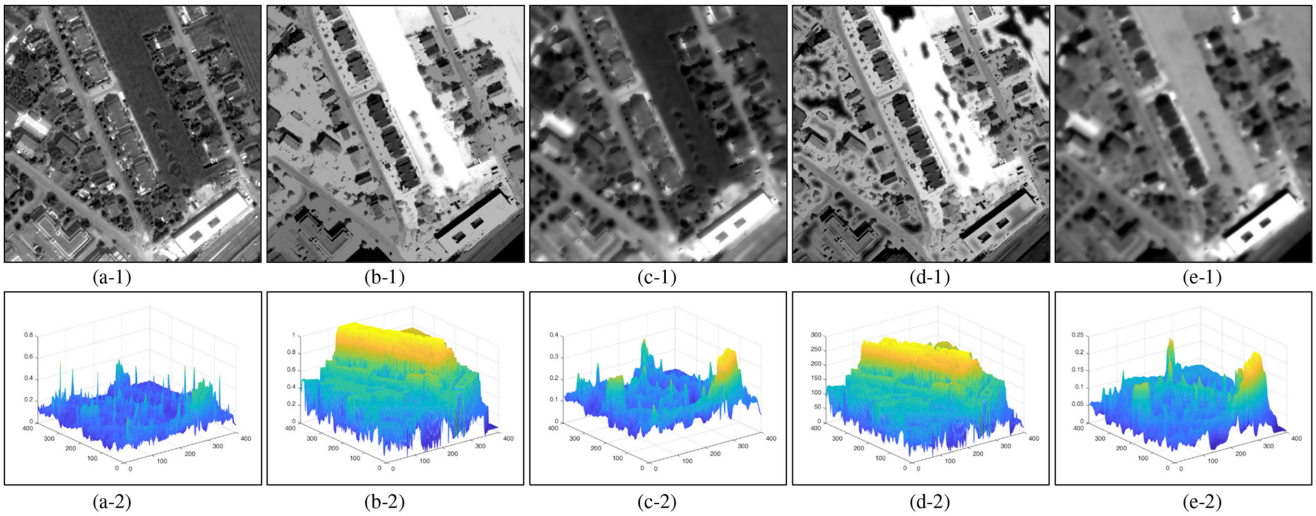

Fig. 6. Visual comparison between different features obtained on the ZH1 scene subset. (a) Spectral feature. (b) Spatial feature. (c) GF feature. (d) GFF feature. (e) TsF feature. Row 2 shows the 3-D visualization corresponds to the features in row 1.

be stable when k exceeds 7. Meanwhile, time cost T is at a relatively low level when k is defined in the range of [3, 7], whereas when k is larger than 7, T increased dramatically. This behavior is similar in all training samples' conditions, and it accentuates when the number of training samples is large (e.g., 2000 pixels/class). Accordingly, searching for a compromise between the classification accuracy and the computational cost, k was set to 7 in the experiments.

In the intergroup fusion step, by taking into account the distinctive features at different fusion scales, the size of the moving window z and r were set to [3, 10] to search more significant feature information. For the compared GF algorithm, the parameter δ was manually fixed to 0.02 due to the fact that it has less influence on the classification results. For the GFF algorithm, it performs guiding with different input parameters in two kinds of layers [27], in order to better distinguish the information of the base layer and the detail layer. Therefore, based on multiple trials, two window parameters (r_1 and r_2) and two regularization parameters (δ_1 and δ_2) were defined in GFF as follows: $r_1 = 45$, $\delta_1 = 0.3$; $r_2 = 7$, and $\delta_2 = 10^{-6}$.

C. Qualitative Analysis and Comparison of Different Fusion Features

To visually evaluate the fusion performance of the proposed TsF approach, qualitative analysis was made by comparing our results with the original GF [26] and its improved version GFF [27]. In particular, the first spectral band (blue band) was selected as the spectral feature input (f_{spe}) and the AP feature based on the attribute thinning of the fourth spectral band (near-infrared band) was selected as the spatial feature input (f_{spa}). Note that in order to have a fair comparison, only the second intergroup feature fusion step is performed in the TsF.

In Fig. 6(a-1) and (b-1), spectral and spatial features for a subset in ZH1 scene dataset are shown. Three fused images obtained by GF, GFF, and the TsF methods are shown in Fig. 6(c-1)–(e-1), respectively. From the image details, one can see that the original GF method focuses more on the spectral information (more similar to the input spectral feature), while the GFF method focuses more on the spatial information representation (more similar to the input spatial feature). However, in both GF and GFF fusion results, details of land-cover objects area are either eliminated or overexaggerated, and thus, they are not properly inherited from the two input features [Fig. 6(c-1) and (d-1)]. In contrast, the proposed approach is able to generate better fusion results by integrating the distinctive information of two input features. This can be more clearly verified from the 3-D visualization of the different features (see Fig. 6 row 2): the proposed TsF resulted in a better fusion output [Fig. 6(e-2)] than two reference methods [Fig. 6(c-2) and (d-2)]. It preserves the original spectral shapes and spatial modeling information, but also enhances the feature representation, thus providing more discriminable features in the fusion result [Fig. 6(e-2)].

Fig. 7 row 1 presents the spectral, spatial features along with three fused features for a subset of the ZH2 scene, while the 3-D visualizations in row 2 correspond to the five features in row 1. It is clear that the TsF method offers the optimal fusion output due to the fact that both the spectral and the spatial information are well preserved [see Fig. 7(e)]. For instance, some bright objects associated with building and trees in Fig. 7(a-2) and (b-2) are well represented in reasonable peaks and plains in Fig. 7(e-2). However, the fused features of the reference GF [see Fig. 7(c-2)] and GFF [see Fig. 7(d-2)] methods only tend to keep partial information that is dominated by the spectral or the spatial features.

D. Quantitative Analysis and Comparison of Classification Results Based on Different Features

To further validate the effectiveness of the proposed TsF approach for VHR image classification, quantitative analysis was carried out by comparing the SVM classification results obtained on the baseline features, i.e., two raw input f_{spe} and f_{spa} features (i.e., the original four spectral bands and eight AP

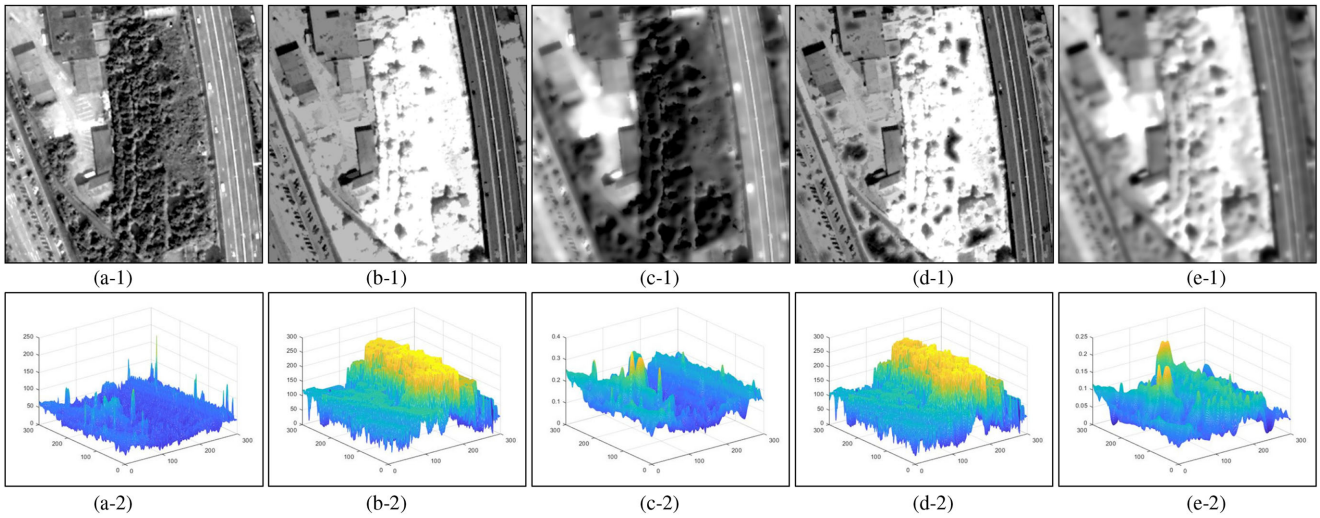

Fig. 7. Visual comparison between different features obtained on the ZH2 scene subset. (a) Spectral feature. (b) Spatial feature. (c) GF feature. (d) GFF feature. (e) TsF feature. Row 2 shows the 3-D visualization corresponds to the features in row 1.

TABLE II
COMPARISON OF THE OA VALUES OBTAINED BY SIX CONSIDERED METHODS WITH DIFFERENT TRAINING SAMPLES (ZH1 DATASET)

Methods	Classification accuracies (%) obtained by different training samples (pixels/class)						
	20	50	100	200	500	1000	2000
f_{spe}	64.84±2.82	67.43±1.52	69.27±1.13	71.73±0.66	72.79±0.54	73.30±0.40	73.73±0.33
f_{spa}	73.14±3.70	79.18±1.12	80.85±0.75	82.59±1.21	84.30±0.46	84.94±0.44	85.49±0.17
$f_{spe+spa}$	75.64±2.83	80.22±1.15	81.81±1.07	84.13±0.92	85.91±0.51	86.81±0.28	87.30±0.12
GF	72.71±2.48	77.84±1.41	80.17±0.91	81.93±0.61	84.27±0.43	85.63±0.33	87.16±0.27
GFF	72.50±2.76	79.07±1.61	83.28±0.84	85.68±0.97	88.23±0.34	89.78±0.30	90.95±0.14
TsF	**81.77±2.29**	**86.00±1.40**	**88.28±0.52**	**89.78±0.72**	**91.60±0.18**	**92.33±0.16**	**92.89±0.20**

TABLE III
COMPARISON OF THE OA VALUES OBTAINED BY SIX CONSIDERED METHODS WITH DIFFERENT TRAINING SAMPLES (ZH2 DATASET)

Methods	Classification accuracies (%) obtained by different training samples (pixels/class)						
	20	50	100	200	500	1000	2000
f_{spe}	75.79±2.20	76.88±1.72	79.10±0.97	79.77±0.62	80.15±0.44	80.51±0.20	80.70±0.27
f_{spa}	78.30±2.24	80.68±1.39	83.61±1.09	85.00±0.66	86.12±0.35	86.64±0.24	87.22±0.35
$f_{spe+spa}$	78.21±2.43	80.60±2.16	83.36±1.36	85.41±0.64	87.21±0.41	87.93±0.28	88.53±0.18
GF	79.21±2.79	81.63±1.32	83.64±0.72	85.49±0.48	87.15±0.41	88.80±0.32	90.09±0.31
GFF	75.94±2.43	80.08±1.26	83.92±1.43	86.24±0.56	89.04±0.52	90.70±0.30	91.81±0.16
TsF	**82.99±1.94**	**84.81±1.87**	**87.20±1.03**	**88.86±0.44**	**90.94±0.31**	**92.07±0.23**	**93.02±0.17**

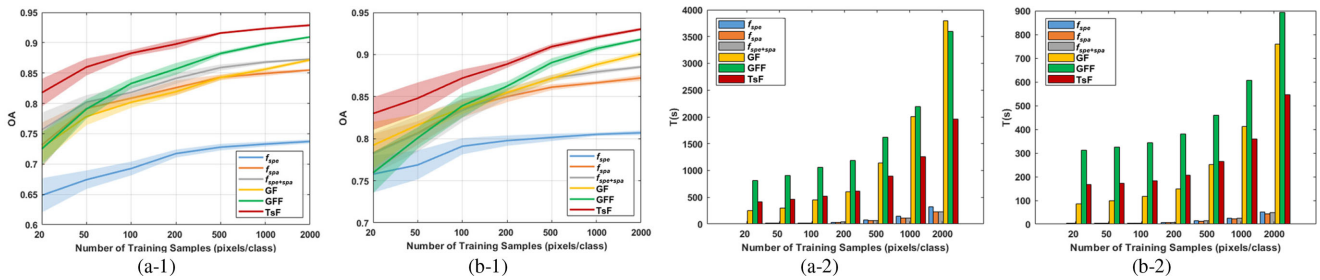

Fig. 8. Comparison of the accuracy and time cost obtained by six considered methods with different training samples on (a) ZH1 and (b) ZH2 scene datasets.

features based on the structure attribute d), and four fusion results obtained by feature stacking ($f_{\mathrm{spe+spa}}$), the GF, the GFF, and the proposed TsF methods. In particular, we randomly generated ten groups of training samples for testing. Numerical experimental results are shown in Tables II and III. The standard variances of OA after ten times of randomizations are illustrated by the shaded areas [see Fig. 8(a-1) and (b-1)], and the average T was shown in Fig. 8(a-2) and (b-2). Classification maps obtained on two scene datasets with 2000 training samples per class are compared in Figs. 9 and 10.

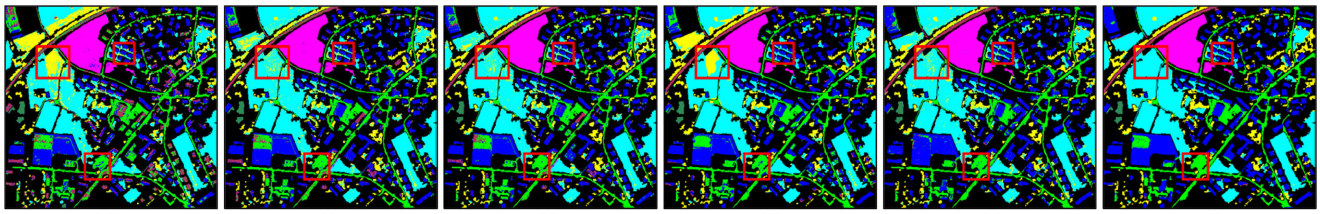

Fig. 9. Classification maps obtained on the ZH1 scene dataset based on 2000 training samples per class and (a) f_{spe}, (b) f_{spa}, (c) $f_{\mathrm{spe+spa}}$, (d) GF, (e) GFF, and (f) TsF six methods.

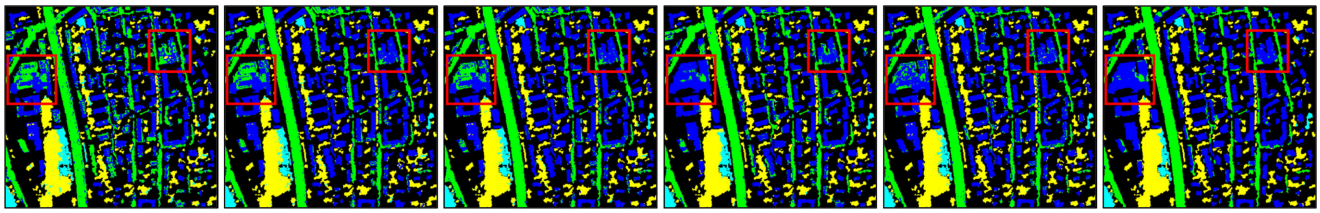

Fig. 10. Classification maps obtained on the ZH2 scene dataset based on 2000 training samples per class and (a) f_{spe}, (b) f_{spa}, (c) $f_{\mathrm{spe+spa}}$, (d) GF, (e) GFF, and (f) TsF six methods.

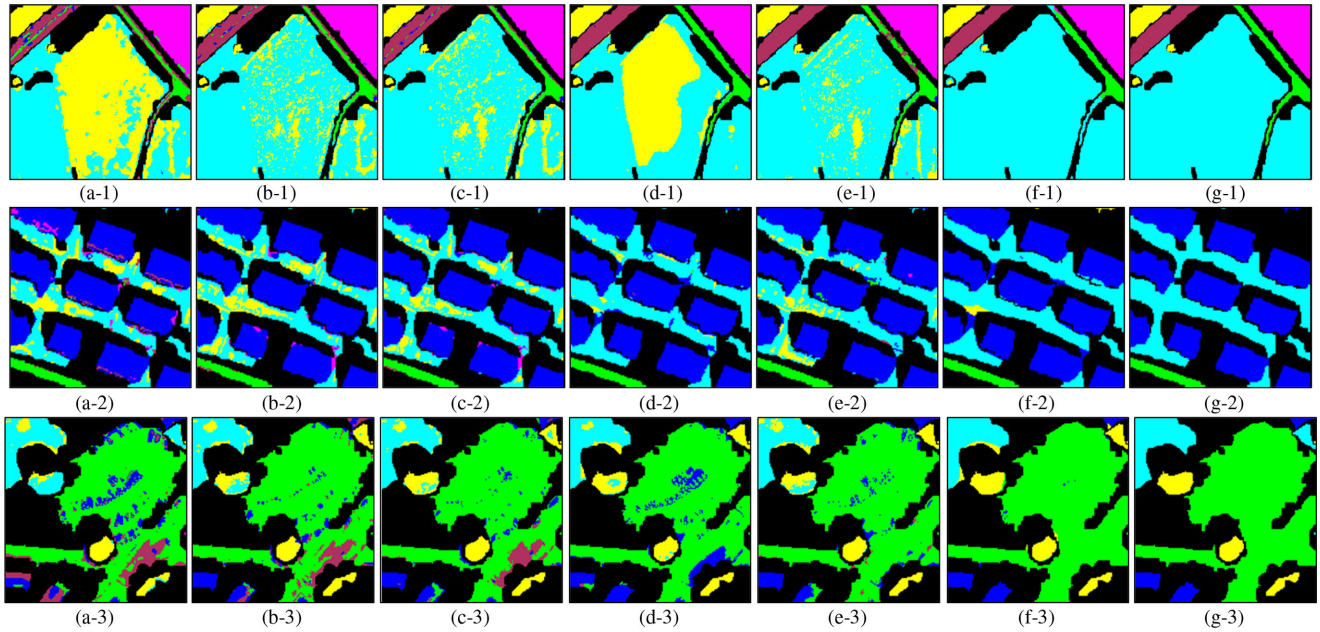

Fig. 11. Comparison of the obtained classification maps at the local scale on the ZH1 scene dataset. (a) f_{spe}. (b) f_{spa}. (c) $f_{\mathrm{spe+spa}}$. (d) GF. (e) GFF. (f) TsF methods. (g) Reference map. Rows 1–3 correspond to the subregion highlighted in red boxes in Fig. 9.

From the experimental results, it is clear that the proposed TsF approach significantly outperformed the state-of-the-art fusion methods and the baseline ones in terms of the highest OA values with different training samples (see both dataset results in Tables I and II). Moreover, as shown in Fig. 8(a-1) and (b-1), the improvement is also significant especially when the number of training samples is small (e.g., 20 pixels/class). With the increase of training samples, we can notice that the change of accuracy tends to be more stable, and the influence of random samples becomes negligible.

From the computational cost point of view [see Fig. 8(a-2) and (b-2)], it is obvious that the f_{spe}, f_{spa}, and $f_{\mathrm{spe+spa}}$ methods have low computational cost as they do not implement complex fusion operations. However, their accuracies are lower than the GF-based fusion methods. Among the GF, GFF, and TsF fusion methods, the computational cost of the proposed TsF is at a relatively low level, whereas its classification accuracy is the highest.

From the visual comparison of the obtained classification maps based on 2000 training samples per class (with the highest OA output) (see Figs. 9 and 10) with respect to the reference maps shown in Figs. 2(b) and 3(b), the raw spectral features f_{spe} [see Figs. 9(a) and 10 (a)] resulted in the worst classification results in the two datasets having numerous commission and omission errors. Considering only the AP spatial features f_{spa} [see Figs. 9(b) and 10(b)], the classification results improved,

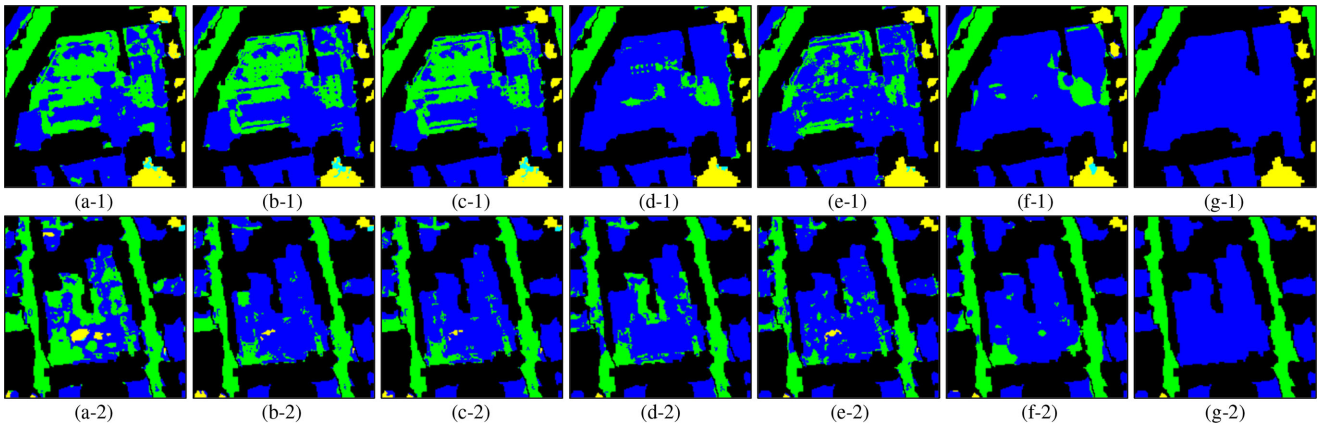

Fig. 12. Comparison of the obtained classification maps at the local scale on the ZH2 scene dataset. (a) f_{spe}. (b) f_{spa}. (c) $f_{\mathrm{spe+spa}}$. (d) GF. (e) GFF. (f) TsF methods. (g) Reference map. Rows 1–2 correspond to the subregion highlighted in red boxes in Fig. 10.

but there are still many misclassified pixels. Among the four fusion methods compared, we can see that the proposed TsF approach achieved the best results [see Figs. 9(f) and 10(f)] outperforming the other three fusion methods with respect to less false alarms or misclassification. It indicates that more discriminant information can be obtained and enhanced after the effective fusion steps of the proposed TsF approach to reduce the classification errors.

In order to further evaluate the classification performance on local subsets, we investigated the classification results at a local scale. Fig. 11 shows the classification and reference data for three subsets of the entire map. Each row corresponds to the highlighted regions in red boxes, as shown in Fig. 9. Compared to the reference map [see Fig. 11 (g)], we can observe that the TsF method produced the most accurate classification results [see Fig. 11(f)]. Moreover, the geometric boundaries of the land-cover objects are well depicted while the inner-class spectral homogeneity is preserved to a great extent. For the five reference methods, commission errors are mainly presented between the following classes: grass (yellow color) and trees (cerulean color), roads (green color) and buildings (blue color). In particular, in this case, f_{spe} [see Fig. 11(a)] and the original GF [see Fig. 11(d)] resulted in a higher misclassification rate compared to the other methods.

Two subsets of the ZH2 scene dataset [see the highlighted red boxes in Fig. 10)] are selected and further compared in Fig. 12. Unlike the previous dataset, this image scene is dominated by roads (green color) and buildings (blue color), so these classes are more likely to be mixed and may lead to classification errors. In the two reference baseline results and the three fusion methods results, it is also clear that many pixels of the buildings are misclassified as roads, as shown in Fig. 12 (a)–(e). The proposed TsF approach presents the best classification performance in such complex local regions [see Fig. 12(f)].

IV. Conclusion

In this article, a novel TsF approach has been proposed to address the multiple feature fusion problem in VHR remote sensing image classification. The main novelty of this work is the design of a sequential fusion process that can not only preserve discriminative information via the intragroup fusion step, but can also capture more significant information via the intergroup fusion step. Moreover, feature redundancy is eliminated and the most significant information in different types of features is preserved without losing their discriminative capability. Experimental results obtained on two VHR scene datasets demonstrate the effectiveness of the proposed TsF approach in terms of higher classification accuracy. Both qualitative and quantitative evaluation of the classification results at global and local scales further confirmed its superiority. Future developments will be focused on the design and improvement of the fusion strategies to take full advantages of different types of features.

Acknowledgment

The authors would like to thank Dr. Michele Volpi from Swiss Federal Institute of Technology Zurich for providing the QB dataset.

References

[1] A. Samat, C. Persello, S. Liu, E. Li, Z. Miao, and J. Abuduwaili, "Classification of VHR multispectral images using extratrees and maximally stable extremal region-guided morphological profile," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 11, no. 9, pp. 3179–3195, Sep. 2018.

[2] Y. Sun *et al.*, "Geo-parcel-based crop classification in very-high-resolution images via hierarchical perception," *Int. J. Remote Sens.*, vol. 41, no. 4, pp. 1603–1624, 2019.

[3] D. Brunner, G. Lemoine, and L. Bruzzone, "Earthquake damage assessment of buildings using VHR optical and SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2403–2420, May 2010.

[4] L. Chen, G. Mei, K. Yan, W. Hao, and X. Yu, "Species discrimination of plantations in subtropical China using 4-band VHR imagery and an operational image analysis framework," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 11, no. 8, pp. 2800–2813, Aug. 2018.

[5] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1923–1938, Apr. 2019.

[6] S. Liu, D. Marinelli, L. Bruzzone, and F. Bovolo, "A review of change detection in multitemporal hyperspectral images: Current techniques, applications, and challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 140–158, Jun. 2019.

[7] S. Liu, Q. Du, X. Tong, A. Samat, L. Bruzzone, and F. Bovolo, "Multiscale morphological compressed change vector analysis for unsupervised multiple change detection," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 10, no. 9, pp. 4124–4137, Sep. 2017.

[8] N. Falco, M. D. Mura, F. Bovolo, J. A. Benediktsson, and L. Bruzzone, "Change detection in VHR images based on morphological attribute profiles," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 3, pp. 636–640, May 2013.
[9] T. Zhang, N. Ai, L. Wang, J. Wang, and J. Peng, "Spectral-spatial hyperspectral image classification based on sparse representation and edge preserving filtering," *Proc. Int. Conf. Front. Adv. Data Sci.*, Xi'an, China, 2017, pp. 165–170.
[10] S. Liu *et al.*, "A multi-scale superpixel-guided filter feature extraction and selection approach for classification of very-high-resolution remotely sensed imagery," *Remote Sens.*, vol. 12, no. 5, 2020, Art. no. 862.
[11] M. Liang, L. Jiao, S. Yang, F. Liu, B. Hou, and H. Chen, "Deep multiscale spectral-spatial feature fusion for hyperspectral images classification," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 11, no. 8, pp. 2911–2924, Aug. 2018.
[12] A. Sellami, Ali Ben Abbes, Vincent Barra, and Imed Riadh Farah, "Fused 3-D spectral-spatial deep neural networks and spectral clustering for hyperspectral image classification," *Pattern Recognit. Lett.*, vol. 138, pp. 594–600, 2020.
[13] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2020.3015157.
[14] D. Hong *et al.*, "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2020.3016820.
[15] X. Wang, P. Du, D. Chen, S. Liu, W. Zhang, and E. Li, "Change detection based on low-level to high-level features integration with limited samples," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 6260–6276, Oct. 2020, doi: 10.1109/JSTARS.2020.3029460.
[16] P. Du *et al.*, "Information fusion techniques for change detection from multi-temporal remote sensing images," *Inf. Fusion*, vol. 14, no. 1, pp. 19–27, 2013.
[17] P. Du, S. Liu, P. Gamba, K. Tan, and J. Xia, "Fusion of difference images for change detection over urban areas," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 5, no. 4, pp. 1076–1086, Aug. 2012.
[18] X. Ma, X. Tong, S. Liu, C. Li, and Z. Ma, "A multisource remotely sensed data oriented method for "ghost city" phenomenon identification," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 11, no. 7, pp. 2310–2319, Jul. 2018.
[19] E. Li, Alim Samat, Wei Liu, Cong Lin, and Xuyu Bai, "High-resolution imagery classification based on different levels of information," *Remote Sens.*, vol. 11, 2019, Art. no. 2916.
[20] L. Shu, K. Mcisaac, and G. R. Osinski, "Learning spatial-spectral features for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5138–5147, Sep. 2018.
[21] S. Liu, Q. Du, X. Tong, A. Samat, and L. Bruzzone, "Unsupervised change detection in multispectral remote sensing images via spectral-spatial band expansion," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 12, no. 9, pp. 3578–3587, Sep. 2019.
[22] L. Zhang, L. Zhang, D. Tao, and X. Huang, "On combining multiple features for hyperspectral remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 879–893, Mar. 2012.
[23] Y. Gu, K. Feng, and H. Wang, "Spatial-spectral multiple kernel learning for hyperspectral image classification," in *Proc. 5th Workshop Hyperspectral Image Signal Process., Evol. Remote Sens.*, 2013, pp. 1–4.
[24] Z. Chunsen, Z. Yiwei, and F. Chenyi, "Spectral–spatial classification of hyperspectral images using probabilistic weighted strategy for multifeature fusion," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 10, pp. 1562–1566, Oct. 2016.
[25] S. Niazmardi, A. Safari, and S. Homayouni, "A novel multiple kernel learning framework for multiple feature classification," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 10, no. 8, pp. 3734–3743, Aug. 2017.
[26] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.
[27] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE Trans. image Process.*, vol. 22, no. 7, pp. 2864–2875, Jul. 2013.
[28] A. Jameel, M. M. Riaz, and A. Ghafoor, "Guided filter and IHS-based pan-sharpening," *IEEE Sensors J.*, vol. 16, no. 1, pp. 192–194, Jan. 2016.
[29] Y. Yang, W. Wan, S. Huang, F. Yuan, S. Yang, and Y. Que, "Remote sensing image fusion based on adaptive IHS and multiscale guided filter," *IEEE Access*, vol. 4, pp. 4573–4582, 2016.
[30] M. D. Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3747–3762, Oct. 2010.
[31] L. Shen, Z. Zhu, S. Jia, J. Zhu, and Y. Sun, "Discriminative Gabor feature selection for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 1, pp. 29–33, Jan. 2013.
[32] J. Feng, L. Jiao, F. Liu, T. Sun, and X. Zhang, "Mutual-information-based semi-supervised hyperspectral band selection with high discrimination, high information, and low redundancy," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2956–2969, May 2015.

Sicong Liu (Member, IEEE) received the B.Sc. degree in geographical information system and the M.E. degree in photogrammetry and remote sensing from the China University of Mining and Technology, Xuzhou, China, in 2009 and 2011, respectively, and the Ph.D. degree in information and communication technology from the University of Trento, Trento, Italy, 2015.

He is currently an Assistant Professor with the College of Surveying and Geo-Informatics, Tongji University, Shanghai, China. His research interests include multitemporal remote sensing data analysis, change detection, multispectral/hyperspectral remote sensing, signal processing, and pattern recognition.

Dr. Liu was the winner (ranked as third place) of Paper Contest of the 2014 IEEE GRSS Data Fusion Contest. He is the Technical Co-Chair of the Tenth International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp 2019). He was the Session Chair for many international conferences such as International Geoscience and Remote Sensing Symposium. He is also a Referee for more than 30 international journals.

Yongjie Zheng received the B.S. degree in remote sensing science and technology from Henan Polytechnic University, Jiaozuo, China, in 2018. She is currently working toward the M.S. degree in photogrammetry and remote sensing from Tongji University, Shanghai, China.

Her current research interests include deep learning, feature extraction, feature fusion, and multispectral/hyperspectral image classification.

Qian Du (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Maryland, at Baltimore County, Baltimore, MD, USA, in 2000.

She is currently the Bobby Shackouls Professor with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS, USA. She is also an Adjunct Professor with the College of Surveying and Geo-informatics, Tongji University, Shanghai, China. Her research interests include hyperspectral remote sensing image analysis and applications, pattern classification, data compression, and neural networks.

Dr. Du is a Fellow of the SPIE-International Society for Optics and Photonics. She was a recipient of the 2010 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society. She was a Co-Chair of the Data Fusion Technical Committee of the IEEE Geoscience and Remote Sensing Society from 2009 to 2013, and the Chair of the Remote Sensing and Mapping Technical Committee of the International Association for Pattern Recognition from 2010 to 2014. She was an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, *Journal of Applied Remote Sensing*, and IEEE SIGNAL PROCESSING LETTERS. Since 2016, she has been the Editor-in-Chief of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING. She was the General Chair of the 4th IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, Shanghai, China, in 2012.

Alim Samat (Member, IEEE) received the B.S. degree in geographic information system from Nanjing University, Nanjing, China, in 2009, the M.S. degree in photogrammetry and remote sensing from the China University of Mining and Technology, Xuzhou, China, in 2012, and the Ph.D. degree in cartography and geographic information system from Nanjing University, in 2015.

He is currently an Associate Researcher with the State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi, China. His current research interests include PolSAR and optical remote sensing for land applications, image processing and pattern recognition, and machine learning.

Dr. Samat is a Reviewer for several international journals including the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, *Pattern Recognition*, etc.

Xiaohua Tong (Senior Member, IEEE) received the Ph.D. degree from Tongji University, Shanghai, China, in 1999.

He was a Postdoctoral Researcher with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, China, between 2001 and 2003. He was a Research Fellow with The Hong Kong Polytechnic University in 2006, and a Visiting Scholar with the University of California, Santa Barbara, CA, USA, between 2008 and 2009. His current research interests include remote sensing, GIS, uncertainty and spatial data quality, image processing for high resolution, and hyperspectral images.

Dr. Tong is the Vice-Chair of the Commission on Spatial Data Quality of the International Cartographical Association, and the Co-Chair of the ISPRS working group (WG II/4) on Spatial Statistics and Uncertainty Modeling.

Michele Dalponte (Senior Member, IEEE) received the M.Sc. degree in telecommunications engineering and the Ph.D. degree in information and communication technologies from the University of Trento, Trento, Italy, in 2006 and 2010, respectively.

He was a Postdoctoral Researcher with the Norwegian University of Life Sciences, Norway, and with the University of Cambridge, U.K. He is currently a Researcher with the Forest Ecology and Biogeochemical Cycles Group, Research and Innovation Center, Edmund Mach Foundation, San Michele all'Adige, Italy. His work has been published in international journals and presented at international conferences. His research interests include the field of remote sensing, in particular the analysis of hyperspectral, multispectral, and LIDAR data for forest monitoring.

Dr. Dalponte is a Reviewer for many remote sensing journals.