Urban Classification by the Fusion of Thermal Infrared Hyperspectral and Visible Data

Article *in* Photogrammetric Engineering and Remote Sensing · November 2015

Impact Factor: 1.61 · DOI: 10.14358/PERS.81.12.901

6 authors, including:

Jiayi Li

Wuhan University

16 PUBLICATIONS **112** CITATIONS

SEE PROFILE

Hongyan Zhang

Wuhan University

40 PUBLICATIONS **442** CITATIONS

SEE PROFILE

Urban Classification by the Fusion of Thermal Infrared Hyperspectral and Visible Data

Jiayi Li, Hongyan Zhang, Min Guo, Liangpei Zhang, Huanfeng Shen, and Qian Du

Abstract

The 2014 Data Fusion Contest, organized by the Image Analysis and Data Fusion (IADF) Technical Committee of the IEEE Geoscience and Remote Sensing Society, involved two datasets acquired at different spectral ranges and spatial resolutions: a coarser-resolution long-wave infrared (LWIR, thermal infrared) hyperspectral data set and fine-resolution data acquired in the visible (VIS) wavelength range. In this article, a novel multi-level fusion approach is proposed to fully utilize the characteristics of these two different datasets to achieve improved urban land-use and land-cover classification. Specifically, road extraction by fusing the classification result of the TI-HSI dataset and the segmentation result of the VIS dataset is first proposed. Thereafter, a novel gap inpainting method for the VIS data with the guidance of the TI-HSI data is presented to deal with the swath width inconsistency, and to facilitate an accurate spatial feature extraction step. The experimental results with the 2014 Data Fusion Contest datasets suggest that the proposed method can alleviate the multi-spectral-spatial resolution and multi-swath width problem to a great extent, and achieve an improved urban classification accuracy.

Introduction

Large remote sensing datasets for the study of urban land-use and land-cover are very important for the study of human activities and urbanization progress monitoring. Remote sensing data processing and analysis for urban areas often benefits from the integration of different information, such as different spectral and spatial resolutions for image pan-sharpening (Guo *et al.*, 2014), different spatial and spectral features for land-cover classification (Huang and Zhang, 2011), and different temporal observations for surface change detection (Huang *et al.*, 2014).

Among the above-mentioned fusion tasks, urban surface land-cover and land-use classification has been the subject of a great deal of interest. First, a lot of meaningful spatial features have been designed to alleviate the discriminative limitation of the spectral interpretability (Dalla Mura *et al.*, 2010). Based on the multiple features, both feature-level fusion (Li *et al.*, 2014) and decision-level fusion (Huang and Zhang, 2011), frameworks can be built for urban surface classification. Second, object-oriented analysis techniques utilizing the fusion of pixel-level labeling and a segmentation map can enhance the performance and stability in a homogenous parcel, as well as speeding up the classification progress (Hay and Blaschke, 2010; Li *et al.*, 2014). Meanwhile, most of the current data fusion methods focus on utilizing passive optical remote sensing data ranging from the visible to the near-infrared (that is, from 400 nm to 1100 nm) (Zhang *et al.*, 2012). The radiant energy collected by a thermal infrared sensor can also contribute meaningful complementary information to urban remote sensing land-cover classification. The superiority of thermal infrared imagery is down to the all-weather and all-time capability. Recently, a new airborne measurement combination, involving thermal infrared hyperspectral imagery (TI-HSI) (ranging from 780nm to 1150 nm) and a simultaneously acquired VIS dataset, was released by the Image Analysis and Data Fusion (IADF) Technical Committee of the IEEE Geoscience and Remote Sensing Society (GRSS) for the 2014 Data Fusion Contest (*http://cucciolo.dibe.unige.it/IPRS/*IEEE_GRSS_IADF*TC_2014_Data_Fusion_Contest.htm*). These datasets are suitable for urban land-use and land-cover classification because the thermal infrared data can describe surface-emitted energy differences caused by human activity (Rodríguez-Galiano *et al.*, 2012); meanwhile, the drawback of the coarse spatial resolution can be overcome by integration with the very high spatial resolution (VHSR) VIS data.

With the aforementioned datasets released by the IADF Technical Committee, this article proposes a new fusion framework for urban surface land-cover and land-use classification. Among the recent state-of-the-art research, the winners of the classification task in the 2014 Data Fusion Contest designed a feature fusion approach by combining several of the top principal components (PCs) of the TI-HSI data and some of the spectral and spatial features of the VIS imagery. The runners-up in the contest utilized various spectral-spatial features as a filter to extract the urban land-use and land-cover classes one by one. Details of the processing strategies of the other top results can be found on the website of the Data Fusion Contest (*http://cucciolo.dibe.unige.it/IPRS/*IEEE_GRSS_IADF*TC_2014_Classification_Contest_Results.htm*). It is notable that most of these methods directly take the spectral information of the TI-HSI data into a supervised machine learning procedure, without considering the specific discriminability of the different urban objects in the thermal infrared spectral domain.

Interpretation of this new dataset combination with different resolutions and different swath widths is quite challenging. In view of this, the following issues must be taken into consideration in the data analysis process. With regard to the TI-HSI data: (a) the low energy and low signal-to-noise ratio (SNR) in each band significantly affect the extraction of discriminative features (Zhang *et al.*, 2014); (b) the high interband correlation reveals significant spectral redundancy (Yan and Niu, 2014); (c) there can be spectral variation in the same

Jiayi Li, Hongyan Zhang, Min Guo, and Liangpei Zhang are with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, and the Collaborative Innovation Center for Geospatial Technology, Wuhan University, P.R. China, 129 Luoyu Road, Wuhan, Hubei, 430079 P.R. China (zhanghongyan@whu.edu.cn).

Huanfeng Shen is with the School of Resource and Environment Science, Wuhan University, P.R. China, 129 Luoyu Road, Wuhan, Hubei, 430079 P.R. China.Qian Du is with the Department of Electrical and Computer Engineering, Mississippi State University, 406 Hardy Road, Mississippi State, MS 39762.

Photogrammetric Engineering & Remote Sensing
Vol. 81, No. 12, December 2015, pp. 9–847.
0099-1112/15/9–847

doi: 10.14358/PERS.81.12.9

land-cover object collected at different times, which is also aggravated by human activity in man-made objects (especially for land-use and land-cover classes in urban areas) (Miliaresis, 2014); and (d) ambiguous boundaries of land objects are caused by heat diffusion (Rodríguez-Galiano *et al.*, 2012), which seriously affects the classification performance at a fine spatial resolution (0.2 m in this study). As for the VIS dataset, interpretation is also difficult due to the fact that VHSR imagery exhibits severe intra-class variations. As for the combination of these two datasets, the swath width of the TI-HSI data is greater than that of the VIS data, which results in large image gaps within the VIS data. In our proposed approach, an effective inpainting method is proposed to undertake the gap filling.

Based on this analysis of the two datasets, ambiguous boundaries are ubiquitous in the TI-HSI data, while the boundaries in the VIS data with a fine spatial resolution are very clear. The significant intra-class variations and relatively low inter-class separability of the VIS data also limits the classification performance. In addition, roads in the TI-HSI data are significantly different from the other classes, and the rest of the classes can be effectively distinguished in light of the fine spatial details of the VIS data. In order to tackle the aforementioned problems, a multi-level fusion framework for land-use and land-cover classification is proposed to integrate the complementary and discriminative features of the two datasets. The major contributions of this article are: (a) a full analysis of the challenges posed by the newly released dataset combination is presented; and (b) a novel framework that fully considers the challenges of the classification task is proposed.

The rest of this paper is organized as follows. The next section introduces the datasets and analyzes the problems encountered when dealing with this classification task, followed by the details for the proposed multi-level fusion framework. The experimental results and analysis are then described leading to our conclusions.

Table 1. Environmental Data During Acquisition of the Two Datasets. The Variations Clearly Influence the Radiant Energy of the TI-HSI dataset, as can be Visually Seen in Figure 1b

UTC date and time	Temperature [°C]	Dew point temp. [°C]	Rel. hum. [%]	Pressure [kPa]
21 May 2013, 22:00:00	14.2	9.6	74	96.41
21 May 2013, 23:00:00	13.1	9.4	78	96.48
22 May 2013, 00:00:00	12.5	9.3	81	96.43

Challenges and Analysis

As this airborne dataset combination is newly released, an introduction to the datasets and an analysis of the problems encountered in their integration are needed.

Introduction to the Datasets

The airborne TI-HSI data consist of 84 spectral bands in the 868 to 1280 cm^{-1} region (7.8 to 11.5μm), at a spectral resolution of 6 cm^{-1} (full-width-half-maximum) and with an approximately 1.0 m spatial resolution. The data have been calibrated to at-sensor spectral radiance in $W/(m^2srcm^{-1})$. The VIS data were acquired almost simultaneously, and have been resampled to a 0.2 m spatial resolution. In view of this, classifying the urban land-use and land-cover types at the 0.2 m spatial resolution level is a multi-resolution fusion task. The average height of the sensor platforms was 807 m.

The grayscale map of the red band of the VIS data and the grayscale map of the first band of the TI-HSI data are visually shown in Figure 1a and 1b, respectively. The two datasets cover an urban area near Thetford Mines in Québec, Canada, on 21 May 2013, between the time 22:27:36 to 23:46:01 UTC. A nearby meteorological station, located at geographic coordinates 46°02'57.002"N and 71°15'58.004"W and, and 430.0 m

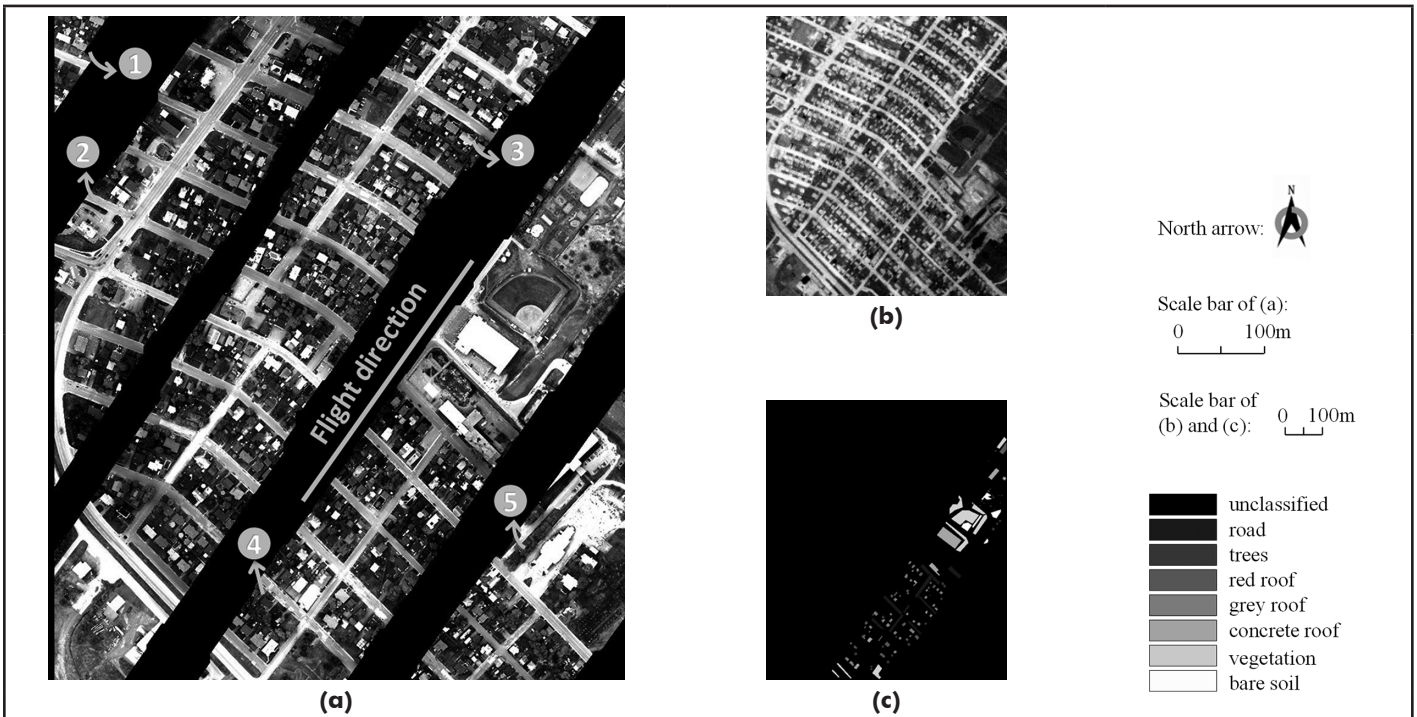


Figure 1. Datasets and labeled sample locations of the study area: (a) The red band of the VIS data; there are five strips in the image - the width of the gaps between two neighboring strips is large; (b) The first spectral band of the TI-HSI data, showing the obvious intensity discrepancy across the flight direction; and (c) Labeled sample locations of the study area, which are all in the fourth strip. The class label grayscale bar applies to all the figures. It should be noted that the spatial ratio between these two datasets is 5:1, but the sizes of the sub-figures here do not follow this ratio, which should not affect the readability.

elevation, recorded the environmental data during image acquisition, as shown in Table 1.

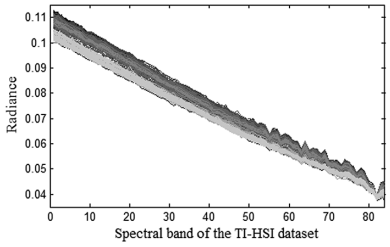
Challenges for Classification

Classification with such a subtle spectral resolution in the thermal infrared range and with such high spatial resolution datasets certainly poses some challenges.

Challenges Posed by the TI-HSI Dataset

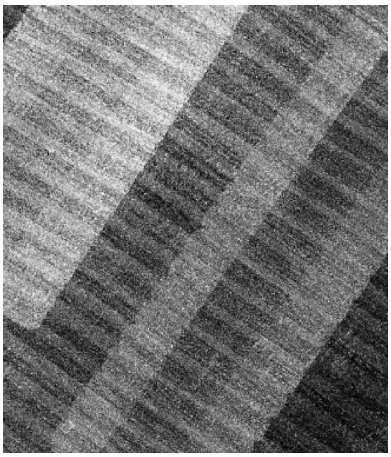
A. Low Energy, Low SNR, and High Inter-band Correlation

Figure 2a shows the representative spectral curves of each class in the original TI-HSI dataset, where the horizontal axis records the number of bands, and the vertical axis represents the radiant energy. It can be seen in Figure 2a that the radiant energy is quite limited, as the maximum vertical value is less than 0.12. Furthermore, there is significant linear correlation between the radiant energies in different bands, which means that high spectral redundancy exists in the dataset. For the noise issue, taking the 80th spectral band shown in Figure 2b as an example, it can be seen that the noise is non-negligible. In addition, Figure 2c displays the uninformative 82nd band of the TI-HSI data, which contains little useful information.



(a)

(b)



(c)

North arrow: Scale bar: 0 100m

Figure 2. The low quality of the TI-HSI dataset: (a) Representative spectral curves of each class in the original TI-HSI dataset; (b) grayscale map of the 80th spectral band of the original TI-HSI dataset; and (c) grayscale map of the 82nd spectral band of the original TI-HSI dataset.

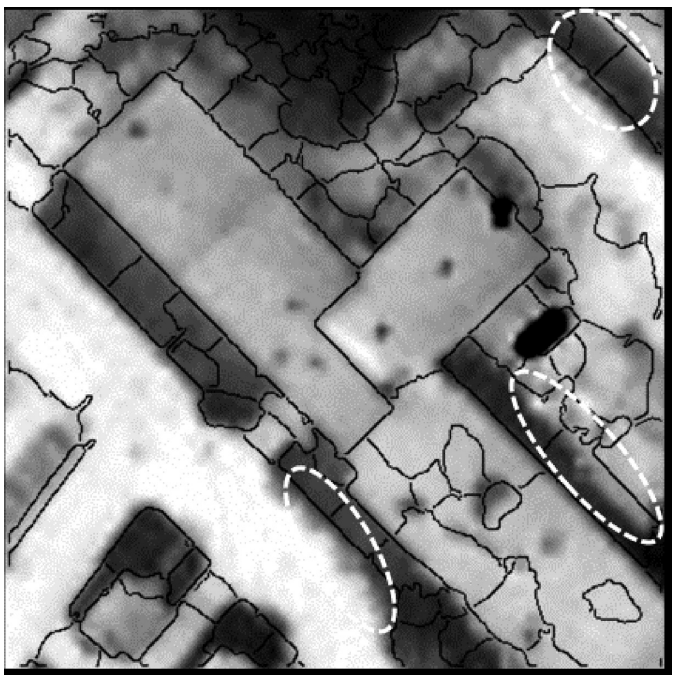
B. Spectral Variation and the Over-fitting Issue

As can be seen in Figure 1, five sequentially acquired strips make up the whole scene. Compared with the VIS data, the thermal radiant energy and spectral discrimination are determined not only by the land-cover material type, but also by the temperature. The radiant energy of the TI-HSI dataset is sensitive to the environmental change occurring during the data acquisition (see Table 1), as can be seen in the obvious intensity discrepancy across the flight direction, as shown in Figure 1b, while the VIS reflectance is relatively stable.

For classification, over-fitting is one of the most important problems, as noted in Chapter 1.4.7 in Murphy (2012). For the study area, the locations of the labeled samples show obvious spatial correlation and redundancy, as shown in Figure 1c. Both the redundant training samples in a local region and spectral variation will aggravate the over-fitting issue.

C. Ambiguous Boundaries of Land Objects

For TI-HSI data, it is well known that the ambiguous boundaries of land objects seriously affect image interpretation accuracy at a fine spatial resolution. To illustrate this problem, some of the blurry boundaries of the TI-HSI data are highlighted in white ellipses in Figure 3, and compared with the segmentation result of the VIS dataset. Since the spatial



Scale bar: 0 100m North arrow:

Figure 3. Illustration of boundaries in the two datasets: a subregion of the TI-HSI data with the VIS segmentation boundaries. As the VIS dataset can offer dedicated edge information, the VIS segmentation boundaries are considered believable.

Table 2. Jeffries-Matusita Distance and Transformed Divergence for the VIS and TI-HSI Datasets

	Road		Trees		Red roof		Gray roof		Concrete roof		Vegetation	
Trees	1.9975 1.9999	1.8556 1.9732							VIS	TI-HSI		
Red roof	1.8861 1.9992	1.8176 1.9102	1.9683 1.9997	**0.8171** **0.1635**				JMD TD				
Gray roof	**0.6343** **0.7374**	1.7233 1.8431	1.9447 1.9999	**1.1751** **1.3970**	1.6400 1.9904	**0.6226** **1.0991**						
Concrete roof	1.7070 1.9259	1.6212 1.9408	2.0000 2.0000	**1.0111** 1.7344	1.9999 2.0000	**0.6383** **0.9885**	1.8703 1.9999	**0.9557** 1.6675				
Vegetation	1.9961 1.9999	1.8424 1.9219	**0.9936** 1.4547	**0.4736** **0.6372**	1.9476 1.9841	**0.9873** 1.6496	1.9566 1.9999	1.2945 1.5322	2.0000 2.0000	**1.0807** 1.8971		
Bare soil	1.9318 1.9600	1.9922 2.0000	1.9984 1.9998	1.9153 2.0000	**1.3953** 1.6779	1.9338 2.0000	1.9515 1.9615	1.9782 2.0000	1.9999 2.0000	1.8897 2.0000	1.9827 1.9964	1.8545 2.0000

Scope for both indices is [0,2] and >1.9→desirable separability, [1.4,1.8]→qualified sample, <1.4→bad classification result

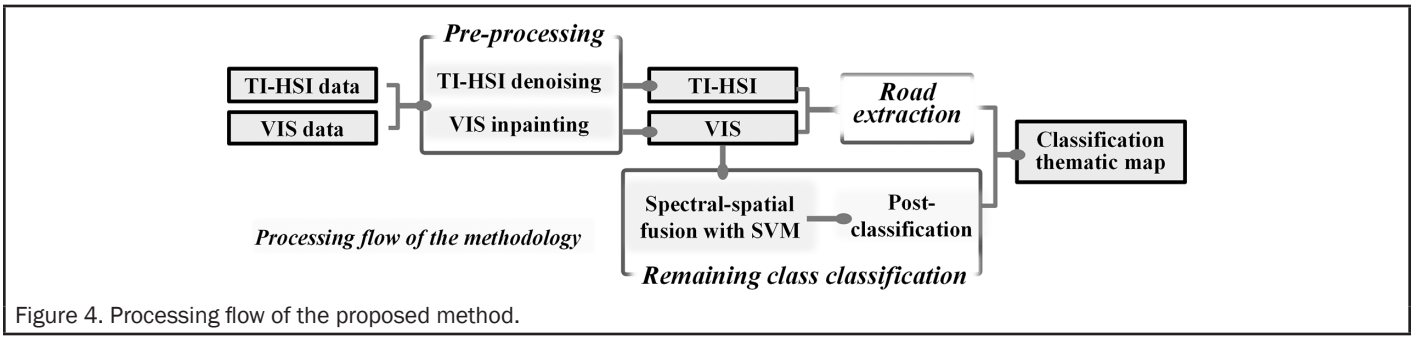


Figure 4. Processing flow of the proposed method.

resolution of the original TI-HSI dataset is 1.0 m, it can be seen that such ambiguity can involve dozens of pixels at the 0.2-m spatial resolution level.

Challenges Posed by the VIS Dataset

In addition to the well-studied problem of dealing with high spatial resolution visible image classification, the large gaps between the available strips aggravates the challenge.

For high spatial resolution imagery, the spatial information can be of great assistance to the spectral information variability (Dalla Mura *et al.*, 2010; Huang and Zhang, 2011), and can be extracted by a set of linear or nonlinear combinations of the surrounding pixels. Although the spatial information can improve the discriminability of the land-cover, it calls for completeness in the spatial domain, and obvious spatial structure corruption will decrease the performance of the spatial descriptor. Unfortunately, due to the inconsistency of the swath widths of the VIS and TI-HSI data, large image gaps are found within the VIS data.

Proposed Method

The first important aspect for the classification is the discriminative feature extraction. The following section analyzes the spectral separability of the training sample set, as it can be utilized as a foundation to improve the classification. The Jeffries-Matusita distance (JMD) and the transformed divergence (TD) are two well-used evaluation indices for measuring the degree of discriminability between two categories. The definitions and physical meanings of these indices can be found in Richards (1993). For the VIS dataset, it can be first observed that the road/concrete class pair show weak separability, as marked in bold in Table 2, and similar observations can be made for the red roof/bare soil pair, due to the discriminative limitation of the visible spectral information. For the TI-HSI dataset, the road pixels can be easily discriminated from the rest of the classes, while the separability for the other class pairs is poor.

In view of the above phenomena, the entire classification framework is composed of the following three parts: (a) data pre-processing; (b) road extraction; and (c) remaining class classification (see Figure 4).

Preprocessing

TI-HSI *Denoising and Dataset Matching*

To alleviate the low-SNR problem, the first step of the pre-processing procedure refers to noise removal with low rank matrix recovery (Zhang et al., 2014). Considering the severe noise, as shown in Figure 2c, the 82nd band is abandoned. With regard to the different spatial resolutions of these two datasets, calibration and resampling techniques are called for in the matching step. The calibration procedure is based on affine transformation. In fact, we just carry out the translation and isometric scaling without rotation, which is enough for the following process. The reason for this is that the two cameras are mounted on the same airborne platform, so the LWIR image and the VIS image are captured at almost the same time and imaging conditions. In this way, upsampling by bicubic interpolation of the denoised TI-HSI data is implemented to match the VIS data at a finer spatial resolution. Bicubic interpolation refers to cubic convolution interpolation, which determines the gray level value from the weighted average of the closest pixels to the specified input coordinates, and assigns that value to the output coordinates, in two dimensions. In this way, the resulting image is slightly smoother than that produced by bilinear interpolation, and it does not have the staircase appearance produced by nearest neighbor interpolation. It is suggested that there is less spectral distortion after

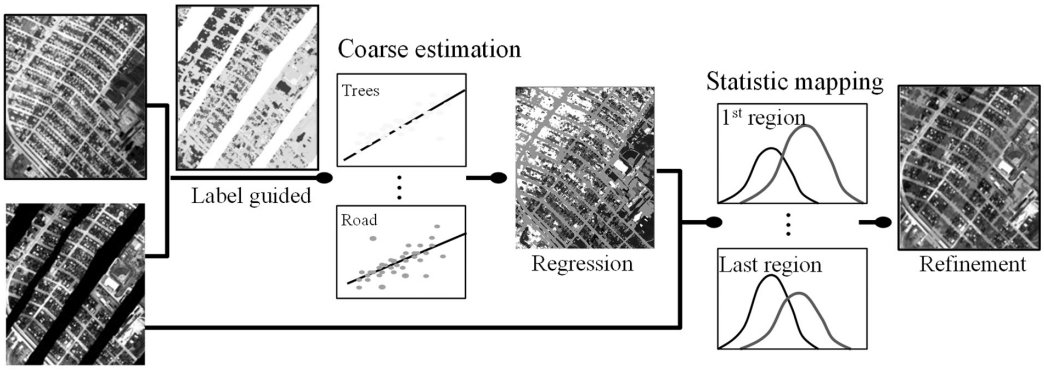


Figure 5. The flowchart of VIS data inpainting.

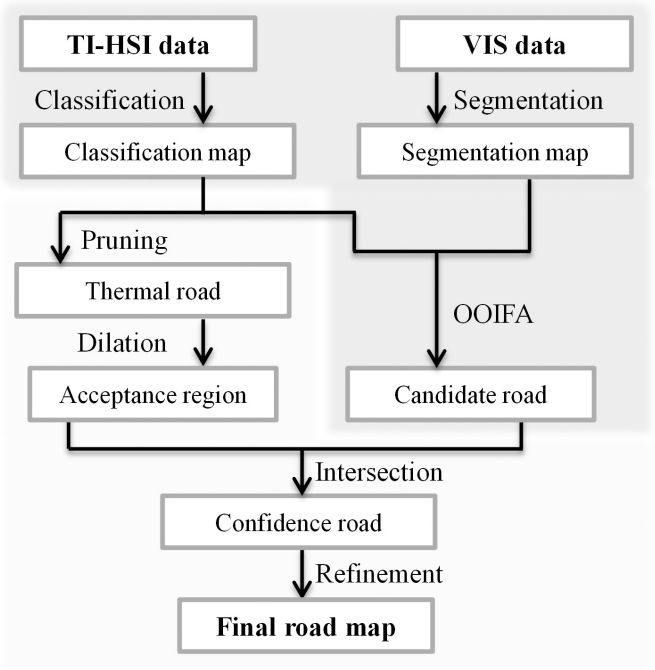


Figure 6. The flowchart for road extraction as the first step in the land-use and land-cover classification framework. The upper part is associated with OOIFA, while the bottom part is the following refinement procedure.

image interpolation, as the spectral information of an image, which is first down-sampled to a certain resolution, and then interpolated to the same image size as before, hardly changes. In the following part after the matching step, "TI-HSI data" refers to the denoised and upsampled result after these pre-processing steps.

VIS Data Inpainting

Due to the inconsistency of the swath width of the VIS and TI-HSI data, there are large image gaps within the VIS data, which significantly and negatively affect the extraction of the spatial features. This is actually an inevitable problem in multi-image interpretation and feature description. Considering the huge data dropout, utilizing the simultaneously acquired TI-HSI data is desirable. As mentioned in the last section, energy in the TI-HSI data changes in the different strips, which aggravates the challenge of the inpainting. Fortunately, the TI-HSI label information can be utilized to build the mapping relationship between the spectral information of the VIS data and that of the thermal infrared measurement in a supervised fashion.

In the proposed approach, the inpainting algorithm includes two steps. (a) Coarse inpainting per class: as all the labeled samples are obtained at the same time, the linear mapping is a function of the two spectral features per class, without considering the change in temperature. That is, a fraction of the labeled samples are randomly selected to linearly predict the mapping coefficients, and several independent trials are taken to correct the unexpected corruption and inappropriate labels of the TI-HSI dataset (see Figure 5). Based on the estimated regression parameter sets, the linear transformation is then performed. (b) Local luminance adjustment: the estimated VIS band may inherit the luminance change, which should be adjusted, from the TI-HSI dataset. In a local spatial parcel, it is assumed that pixels in a local spatial part, belonging to the missing or the known region, should follow the same distribution. Therefore, mapping the distribution of the estimated band and that of the given VIS band in a known region, and matching the means of the associated regions, can alleviate the luminance effects.

Road Extraction

As shown in Figure 6, the flowchart for road extraction includes two steps: (a) object-oriented image fusion analysis (OOIFA); and (b) road refinement in the confidence region.

Object-Oriented Image Fusion Analysis

For VHSR image processing, it is believed that the object-oriented approach is desirable to alleviate the "salt-and-pepper" misclassification phenomenon caused by the intra-class spectral variation. Meanwhile, it was also demonstrated in the previous sub-section that the TI-HSI data contain a superior identification ability for road pixels, and the VIS data have a fine object boundary extraction capability. In view of this, it is intuitive to combine the superiorities of these two datasets in an object-oriented fusion manner. Here, we first classify the TI-HSI dataset by the use of a linear support vector machine (SVM) classifier (Boser *et al.*, 1992; Vapnik, 1999), and segment the VIS dataset with the mean-shift algorithm (Comaniciu and Meer, 2002; Li *et al.*, 2014). Guided by the clusters of the segmentation map, a majority voting approach (Huang and Zhang, 2011) is utilized to achieve the object-oriented fusion result. In this way, the classification map of the TI-HSI dataset and the segmentation result of the VIS dataset are integrated to produce a candidate road map, as illustrated in Figure 6.

Road Refinement

The aforementioned object-oriented fusion technique can be affected by the segmentation scale, and it is difficult to adaptively match all the land objects in an entire complex scene at one scale. As shown in Figure 7, although most of the concrete roof pixels in the lower-left part of the sub-region with its surrounding road pixels are correctly classified in

Figure 7b, the aforementioned OOIFA strategy misclassifies the concrete roof class as roads, which is caused by the under-segmented blocks in Figure 7c. Due to the scene complexity, it is difficult to determine a suitable segmentation scale, and even a multi-scale segmentation method may not fully address this problem. In this paper, the fine separability and the homogeneity of the road areas in the TI-HSI data are again utilized to deal with the problem from another perspective. It is believed that heat diffusion only affects adjacent pixels, and the morphological dilation operation is conducted with road pixels of the TI-HSI classification map. Pixels combined with their dilation-related ones comprise the so-called "confidence region," and are labeled as potential roads, as shown in Figure 7e. In view of this, only pixels located in the confidence region accept the labeling correction made by the OOIFA, as shown in Figure 7f. Finally, the concept that the road regions are elongated areas with connected components is utilized for pruning, to generate the final road map (Maurya *et al.*, 2011).

Classification of Other Classes

For the remaining classes, we focus on the fine spatial resolution VIS data. However, the gaps in the VIS data may lead to inaccurate spatial feature extraction. Thus, this procedure consists of two components: (a) classification by stacked multi-feature vectors; and (b) post-classification.

For VHSR imagery, the spatial features are valuable information for image interpretation (Li *et al.*, 2014). In this work, four meaningful morphological attribute profiles (Dalla Mura *et al.*, 2010), i.e., area, standard deviation (SD), diagonal box (DB), and moment of inertia (MI), along with the gray level co-occurrence matrix (GLCM) feature (Huang and Zhang, 2013) and the spectral feature, are merged in a vector stacking (VS) manner (Li *et al.*, 2014) for the classification. The selection of the parameter set for the above spatial features depends on the type of images and the application and task being dealt with. For the study in this paper, we first extract the profiles with the parameters in a large range, then compute the derivative of the profile, and pick out those containing meaningful spatial structure. Recently, some automatic methods for the selection of these spatial criteria have been proposed (Ghamisi *et al.*, 2015). Among the various classifiers, linear SVM (Fan *et al.*, 2008) is efficient and effective enough to deal with such a task with a large training sample set. To analyze the study area, the parameters of linear SVM were automatically determined by five-fold cross-validation (Li *et al.*, 2014) from a reasonable range. After the classification, the segmentation map of the VIS data is utilized as an object-based decision fusion step (Sugg *et al.*, 2014) to conduct the post-classification process. The selection of the parameter set in the mean-shift segmentation procedure depends on the spectral and spatial resolutions of the image and the terrain complexity within the scene. For the study area, we set the spectral scale as 17, the spatial scale as 15, and the smallest block size as 200 pixels.

Results and Discussion

Experimental Setting

Labeled samples were provided by the IADF 2014 committee, as visually shown in Figure 1c, and were used for the

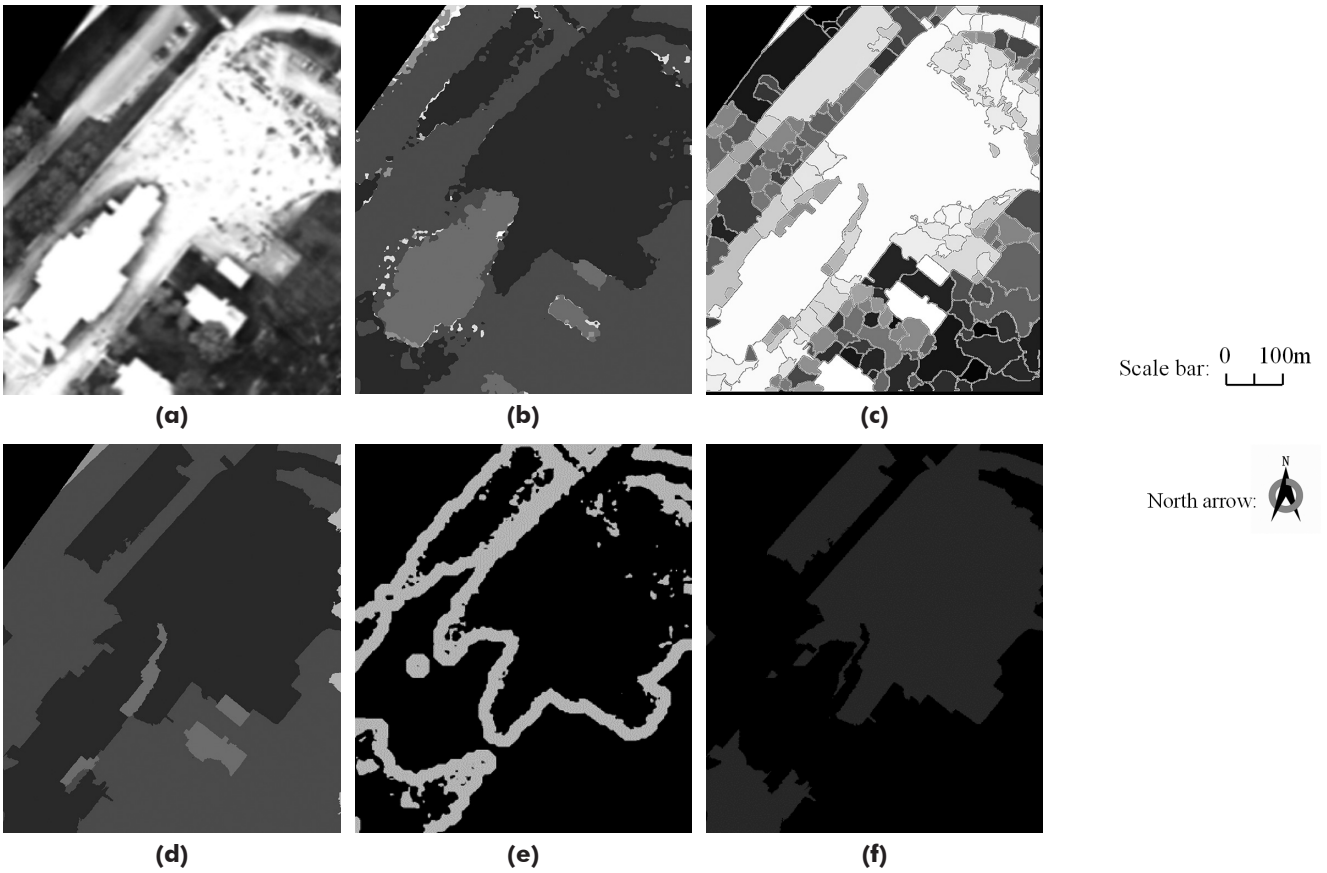


Figure 7. Illustration of the road extraction procedure within a sub-region of the scene: (a) The red band of the VIS dataset; (b) the TI-HSI classification map; (c) the segmentation result and its boundaries in the VIS domain, with each grayscale value associated with the segmentation parcel index; (d) the OOIFA result with the procedure described in the second Section; (e) the highlighted confidence regions, along with pixels labeled as roads in (b); and (f) the final road map after refinement.

Table 3. The Seven Ground-Reference Classes in the Study Area, and the Training and Test Sample Sets for Each Class

No.	Class name	Provided by the contest committee		Edge sample set
		Training samples	Global test set	
1	Road	112457	809098	-
2	Trees	27700	100749	21116
3	Red roof	46578	136697	55702
4	Gray roof	53520	142868	118723
5	Concrete roof	97826	109539	37718
6	Vegetation	185329	103583	105473
7	Bare soil	44738	49212	-
	Total	568148	1451746	338732

Table 4. Comparison of the Classification Approaches

Acronym	Approach
TI-HSI	Original spectral feature of the TI-HSI data
VIS	Original spectral feature of the VIS data
VIS-SF	Single optimal spatial feature of the VIS data
VIS-VS	Vector stacking (VS) of the five spatial features of the VIS data
InVIS-VS	Inpainting of the strips of the VIS data, and then VS of the five spatial features of the VIS data
R-InVIS-VS	Road extraction with the OOIFA approach, then inpainting of the strips, and VS for the remaining class classification
Proposed	R-InVIS-VS, with post-classification

Table 5. Classification Accuracy (%) for the Study Area

Class Name	TI-HSI	VIS	VIS-SF	VIS-VS	InVIS-VS	R-InVIS-VS	Proposed
Road	96.30	88.29	89.86	91.49	94.52	**98.17**	**98.17**
Trees	0.90	12.15	90.16	90.11	88.56	85.94	**91.07**
Red roof	27.71	93.66	96.93	97.85	**98.40**	94.72	95.14
Gray roof	50.10	53.20	67.42	90.24	88.74	95.71	**98.99**
Concrete roof	71.53	**93.77**	92.15	91.49	91.82	90.38	92.29
Vegetation	26.51	**99.40**	91.46	93.46	93.58	94.74	99.14
Bare soil	45.54	88.89	65.55	88.12	88.38	87.69	**89.70**
OA (%)	70.10	81.29	87.80	91.90	93.42	95.57	**96.81**
AA (%)	45.51	75.63	84.79	91.82	92.00	92.48	**94.93**
κ	0.5470	0.7187	0.8160	0.8795	0.9009	0.9324	**0.9514**

Table 6. Summary of the Classification Comparisons Undertaken in the Study Area. A Resampling Method was Used to Conduct the McNemar's Test, to Compare the Proportions of the Correctly Allocated Pixels. All the tests Shown were One-sided, and a 5 Percent Level of Significance was Selected

Classifier 1	Classifier 2	Comparison of the proportions and disagreement		
		ΔOA(%)	$\lvert z \rvert$	Significant?
TI-HSI	VIS	−10.19	1.8204	No
VIS-SF	VIS	6.51	1.6713	No
VIS-VS	VIS	10.61	3.6742	Yes
Proposed	VIS	15.52	4.5962	Yes

training. For testing, there were two test sample sets, the first set being provided by the contest committee for the whole image, while the second set was collected by the authors on the edges near the missing strips, to evaluate the performance of the inpainting step. In the second test set, the road class was omitted, and the bare soil class was mostly absent in the edge parts. Details of the training and test samples for each class of the reference data are shown in Table 3.

General Classification Performance

The different classification approaches described in Table 4 and the top-two results in the contest were selected for comparison. Each thematic map is visually shown in Plate 1. The quantitative evaluation consisted of the classification accuracy for each class, the overall accuracy (OA), the average accuracy (AA), and the kappa coefficient (κ) (Congalton, 1991), as shown in Table 5. For the spectral-spatial feature fusion related approaches, the related parameters were selected with consideration of the spatial distribution and resolution, as suggested in Dalla Mura *et al.* (2010). In addition, the κ values for the winners of the classification contest are given in the bottom row of Table 5. In this table, the best results for each quality index are highlighted in bold, and the second-best results are underlined.

It can be observed that most of the accuracy statistics in Table 5 suggest that the proposed method can obtain the best or at least a desirable performance, except for the bare soil pixels, which are often mixed with grass and tree pixels in the study area. In Plate 1d through 1i), the classification accuracies of the thematic results improve one by one. In Plate 1c, it is suggested that although the road pixels still show a satisfactory discrimination, the imaging environmental changes (i.e., the temperature change, pressure variation, etc., shown in Table 1) reduce the discriminative ability for the rest of the classes, especially the pixels in the second strip, in which most of the pixels (except for the road pixels) are classified as concrete roof. In both Plate 1a and 1b, it can be seen that many corridors existing between the roofs in the community are misclassified as concrete roof pixels, as the material of these corridors is also concrete. Since there is no corridor class type in the reference data, this kind of misclassification is ignored in the quantitative statistics. The proposed method interpreted these pixels as road, which is more suitable in the semantic-level exploitation. Compared with the other thematic maps, Plate 1h and 1i show the best road detection performance, while the latter figure also shows the best bare soil land-cover recognition.

Table 6 shows a statistical comparison between the different approaches. The ΔOA values are the difference in the quantity disagreement between practical Classifier 1 and Classifier 2, and they confirm the inferiority of the former classifier when ΔOA <0. McNemar's test (Foody 2004), which is a non-parametric statistical significance test of the difference

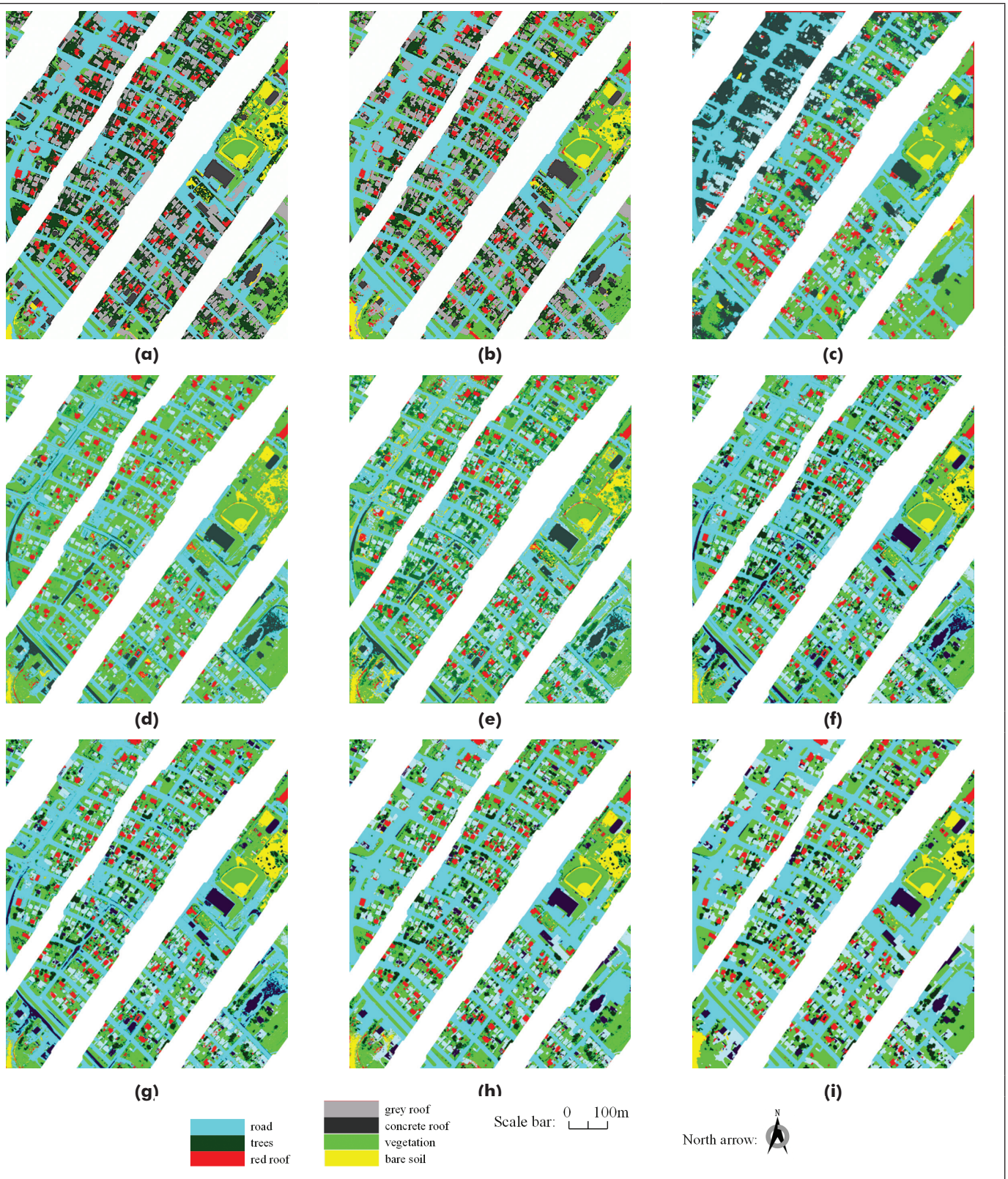


Plate 1. Thematic maps of the study area: (a) runner-up in the 2014 Data Fusion Contest; (b) 2014 Data Fusion Contest winner; (c) TI-HSI; (d) VIS; (e) VIS-SF; (f) VIS-VS; (g) In-VIS-VS; (h) R-InVIS-VS; and (i) the proposed method. These thematic maps are utilized to allow a visual comparison between the proposed method, the winning methods in the 2014 Data Fusion Contest, and the other methods listed in Table 4.

between two classification results, is also utilized in this paper. Although the single spatial feature can improve the discrimination, it is observed that the multiple-feature fusion can achieve a better performance. Furthermore, the proposed approach combining gap inpainting, road extraction, and VS fusion can obtain the best result. However, some issues with the current fusion approach need to be mentioned. First, the computational complexity is high because of the multiple

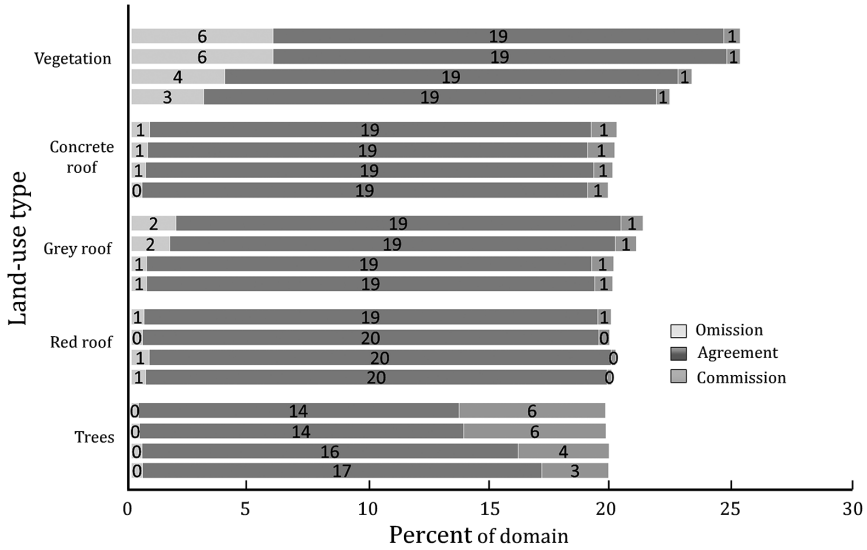


Figure 8. Classification omission, agreement, and commission for each class for the related classifiers with the sub-test set, to display the performance for inpainting. For each group: the upper bar is for VIS-SF (i.e., the GLCM feature of the VIS), the second bar is associated with InVIS-SF (i.e., the GLCM feature of the inpainted VIS), the third bar represents VIS-VS, and the bottom bar refers to the InVIS-VS approach. With regard to these evaluation indices, it can be seen that the classification performance of each category has been improved by the image gap inpainting.

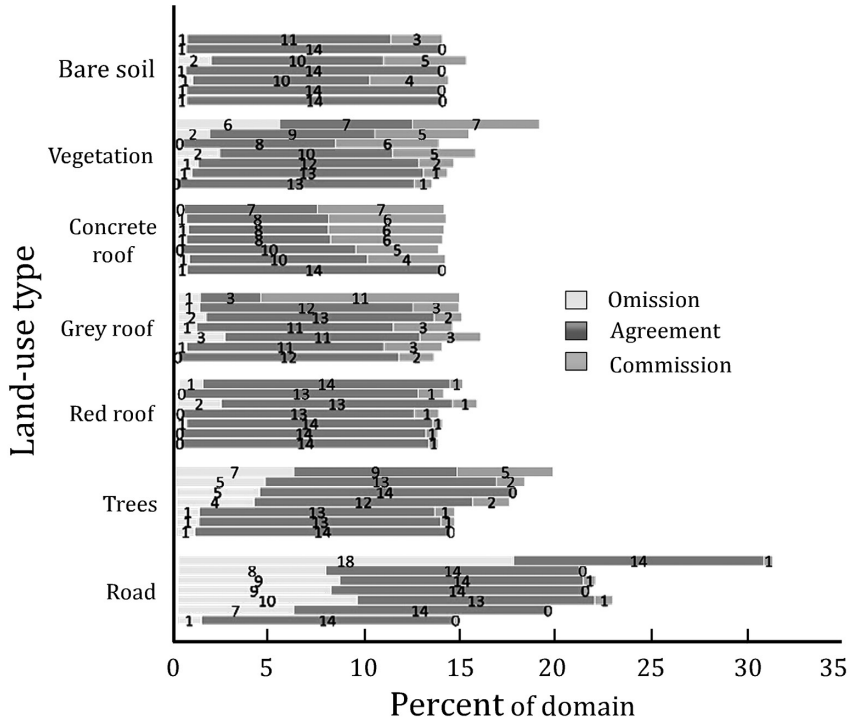


Figure 9. Classification omission, agreement, and commission for each class for the related classifiers with the global test set, to display the discriminative ability for each class. The results are shown for each spatial feature related approach for the remaining class classification of the study area. For each group, results of the SD, area, moment, DB, GLCM, VS, and the proposed approach are shown in order from the top to bottom. With regard to these evaluation indices, it can be seen that the VS approach utilizes the complementary spatial descriptions to effectively improve the classification accuracy.

features utilized in the VS image interpretation. Second, due to the large-scale data classification task (with as many as 568,148 training samples), it is likely that a complicated classifier with greater computational complexity could not be utilized, which will be further studied in the future.

Performances when Dealing with the Challenges

Improved TI-HSI SNR

To evaluate the effectiveness, the no-reference metric Q (Xiang and Milanfar, 2010) was utilized, which indicates a better image quality with a larger value. Due to the space limitations, we omit the visual comparison between the original

and the denoised image, which obtained metric Q values of 0.0020 and 0.0023, respectively.

Enhanced Discriminative Feature Description near Missing Regions

To validate the function of the VIS image gap inpainting for the classification, five single spatial features and the multiple-feature VS approach were utilized, respectively. Table 7 records the classification accuracy results, and the quantitative assessments for each class are shown in Figure 8.

TABLE 7. CLASSIFICATION ACCURACY ON THE EDGE SUB-PART OF THE SECOND TEST SET FOR EACH SPATIAL FEATURE RELATED APPROACH FOR THE REMAINING CLASS CLASSIFICATION OF THE STUDY AREA, BEFORE AND AFTER IMAGE INPAINTING

	Index	Area	SD	DB	MI	GLCM	VS
With inpainting	OA	0.9096	0.9092	0.9103	0.8960	0.9127	**0.9350**
	AA	0.8819	0.8628	0.8856	0.8430	0.9322	**0.9448**
	κ	0.8790	0.8778	0.8800	0.8604	0.8843	**0.9133**
Without inpainting	OA	0.8935	0.8388	0.8950	0.8951	0.9080	0.9274
	AA	0.8721	0.8138	0.8807	0.8321	0.9293	0.9365
	κ	0.8572	0.7881	0.8596	0.8590	0.8781	0.9031

In Table 7, it can be seen that in all the comparison pairs, especially the SD pair, the image gap inpainting can improve the classification accuracy. In Figure 8, the classification omission, agreement, and commission (Li *et al.*, 2014) are shown as the sub-bars for each category in a group, and the detailed number in each sub-bar denotes the associated proportion. Overall, the classification performance of each category has been improved by the image gap inpainting, as shown in Figure 8. Specifically, comparing VIS-VS with In-VIS-VS, it can be seen that the omission decreases for the vegetation, concrete roof, and red roof classes, and the agreement for tree pixels increases. For the commission, similar observations can be made, which further confirm the effectiveness of the simple inpainting step.

Discriminative Ability Analysis for each Class

To illustrate the discriminability between the land-use and land-cover types, the classification omission, agreement, and commission for the global test set are shown as the sub-bars for each class in a group in Figure 9. Here, it can be first seen that the road discriminability of the TI-HSI data is superior to that of the VIS data, although the spatial features are represented in the VIS data. It should be mentioned that the proposed OOIFA approach can achieve the goal of road extraction, and can also alleviate the omission error of the other classes. The remaining classes consist of bare soil, vegetation, trees, and buildings (i.e., red roof, concrete roof, and gray roof). For these classes, it is believed that their spatial description is more useful. For the first three class types in Figure 9 (i.e., bare soil, vegetation, and concrete roof), it is suggested that the contextual feature is superior, as each of these classes has a specific contextual pattern, which can be seen in Figure 1a. It can also be observed that the VS approach can maintain the superiority of the contextual description, to conduct the classification task. For the three building (roof) classes in Figure 9, it can be seen that a single feature cannot describe the discriminative characteristic satisfactorily, and the VS approach utilizes the complementary spatial descriptions to improve the classification accuracy.

Conclusions

The study of the newly released multi-sensor, multi-spectral-spatial resolution, and multi-swath width TI-HSI and VIS datasets is a challenging topic, and this paper presents a multi-level fusion approach for discriminative information mining to achieve urban land-use and land-cover classification. The specific superiorities of the TI-HSI and VIS datasets are integrated to improve the classification accuracy, which was confirmed by a quantitative assessment. In particular, a novel image gap inpainting method for the VIS data with the guidance of the TI-HSI data is applied to deal with the swath width inconsistency and facilitate accurate spatial feature extraction, thereby improving the overall classification accuracy. In summary, it is suggested that utilizing the TI-HSI data together with the VIS data for urban surface exploitation is both promising and meaningful.

Acknowledgments

The authors would like to thank Telops, Inc. (Canada) for providing the data used in this study, and the IEEE GRSS Image Analysis and Data Fusion Technical Committee for organizing 2014 Data Fusion Contest. Thanks are also due to the handling editor and the anonymous reviewers for their insightful and constructive comments. The authors would also like to thank the supporting from the National Basic Research Program of China (973 Program) under Grant 2011CB707105, the National Natural Science Foundation of China under Grants 61201342 and 41431175, and the Program for Changjiang Scholars and Innovative Research Team in University IRT1278.

References

Boser, B.E., I.M. Guyon, and V.N. Vapnik, 1992. A training algorithm for optimal margin classifiers, *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, 1992, pp. 144–152.

Comaniciu, D., and P. Meer, 2002. Mean shift: A robust approach toward feature space analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619.

Congalton, R., 1991. A review of assessing the accuracy of classifications of remotely sensed data, *Remote Sensing of Environment*, 37(1):35-46.

Dalla Mura, M., J.A. Benediktsson, B. Waske, and L. Bruzzone, 2010. Morphological attribute profiles for the analysis of very high resolution images, *IEEE Transactions on Geoscience and Remote Sensing*, 48(10):3747–3762.

Fan, R.-E., K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, 2008. LIBLINEAR: A library for large linear classification, *The Journal of Machine Learning Research*, 9:1871–1874.

Foody, G.M., 2004. Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy, *Photogrammetric Engineering & Remote Sensing*, 70(5):627–634.

Ghamisi, P., M. Dalla Mura, and J.A. Benediktsson, 2015. A survey on spectral-spatial classification techniques based on attribute profiles, *IEEE Transactions on Geoscience and Remote Sensing*, 53(5):2335–2353.

Guo, M., H. Zhang, J. Li, L. Zhang, and H. Shen, 2014. An online coupled dictionary learning approach for remote sensing image fusion, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(4):1284–1294.

Hay, G.J., and T. Blaschke, 2010. Special issue: Geographic object-based image analysis (GEOBIA), *Photogrammetric Engineering & Remote Sensing*, 76(2):121–122.

Huang, X., and L. Zhang, 2011. A multilevel decision fusion approach for urban mapping using very high-resolution multi/hyperspectral imagery, *International Journal of Remote Sensing*, 33(11):3354–3372.

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

Huang, X., L. Zhang, and T. Zhu, 2014. Building change detection from multitemporal high-resolution remotely sensed images based on a morphological building index, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(1):105–115.

IEEE Geoscience and Remote Sensing Society Image Analysis and Data Fusion Technical Committee 2014 Data Fusion Contest: Details of the Data Fusion Contest, URL: *http://cucciolo.dibe.unige.it/IPRS/IEEE_GRSS_IADFTC_2014_Data_Fusion_Contest.htm* (last date accessed: 19 October 2015).

IEEE Geoscience and Remote Sensing Society Image Analysis and Data Fusion (IADF) Technical Committee 2014 Data Fusion Contest: Results, URL: *http://cucciolo.dibe.unige.it/IPRS/IEEE_GRSS_IADFTC_2014_Classification_Contest_Results.htm* (last date accessed: 19 October 2015).

Li, J., H. Zhang, and L. Zhang, 2014. Column-generation kernel nonlocal joint collaborative representation for hyperspectral image classification, *ISPRS Journal of Photogrammetry and Remote Sensing*, 94(0):25–36.

Li, J., H. Zhang, and L. Zhang, 2014. Supervised segmentation of very high resolution images by the use of extended morphological attribute profiles and a sparse transform, *IEEE Geoscience and Remote Sensing Letters*, 11(8):1409–1413.

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

Maurya, R., Gupta, P.R. and Shukla A.S., 2011, Road extraction using K-Means clustering and morphological operations, *Proceedings of the International Conference on Image Information Processing (ICIIP 2011)*, DOI:10.1109/ICIIP.2011.6108839.

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

Miliaresis, G.C., 2014. Daily temperature oscillation enhancement of multitemporal LST imagery, *Photogrammetric Engineering & Remote Sensing*, 80(5):423–428.

Murphy, K.P., 2012. *Machine Learning: A Probabilistic Perspective*, The MIT Press.

Richards, J.A., 1993. *Remote Sensing Digital Image Analysis*, Springer Berlin Heidelberg.

Rodríguez-Galiano, V.F., B. Ghimire, E. Pardo-Igúzquiza, M. Chica-Olmo, and R.G. Congalton, 2012. Incorporating the downscaled Landsat TM thermal band in land-cover classification using random forest, *Photogrammetric Engineering & Remote Sensing*, 78(2):129–137.

Sugg, Z.P., T. Finke, D.C. Goodrich, M.S. Moran, and S.R. Yool, 2014. Mapping impervious surfaces using object-oriented classification in a semiarid urban region, *Photogrammetric Engineering & Remote Sensing*, 80(4):343–352.

Vapnik, V., 1999. *The Nature of Statistical Learning Theory*, Springer.

Yan, L., and X. Niu, 2014. Spectral-angle-based Laplacian eigenmaps for nonlinear dimensionality reduction of hyperspectral imagery, *Photogrammetric Engineering & Remote Sensing*, 80(9):849-861.

Zhang, H., W. He, L. Zhang, H. Shen, and Q. Yuan, 2014. Hyperspectral image restoration using low-rank matrix recovery, *IEEE Transactions on Geoscience and Remote Sensing*, 52(8): 4729-4743.

Zhang, H., L. Zhang, and H. Shen, 2012. A super-resolution reconstruction algorithm for hyperspectral images, *Signal Processing*, 92(9):2082–2096.

Zhu, X., and P. Milanfar, 2010. Automatic parameter selection for denoising algorithms using a no-reference measure of image content, *IEEE Transactions on Image Processing*, 19(12):3116–3132.

(Received 14 September 2014; accepted 02 April 2015; final version 11 June 2015)