ISPRS Journal of Photogrammetry and Remote Sensing 90 (2014) 36–48

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

A multi-index learning approach for classification of high-resolution remotely sensed images over urban areas

Xin Huang *, Qikai Lu, Liangpei Zhang

The State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, PR China

ARTICLE INFO

Article history:
Received 4 June 2013
Received in revised form 10 January 2014
Accepted 19 January 2014

Keywords:
High spatial resolution
Classification
SVM
Morphological
Texture
Feature extraction

ABSTRACT

In recent years, it has been widely agreed that spatial features derived from textural, structural, and object-based methods are important information sources to complement spectral properties for accurate urban classification of high-resolution imagery. However, the spatial features always refer to a series of parameters, such as scales, directions, and statistical measures, leading to high-dimensional feature space. The high-dimensional space is almost impractical to deal with considering the huge storage and computational cost while processing high-resolution images. To this aim, we propose a novel multi-index learning (MIL) method, where a set of low-dimensional information indices is used to represent the complex geospatial scenes in high-resolution images. Specifically, two categories of indices are proposed in the study: (1) *Primitive indices* (*PI*): High-resolution urban scenes are represented using a group of primitives (e.g., building/shadow/vegetation) that are calculated automatically and rapidly; (2) *Variation indices* (*VI*): A couple of spectral and spatial variation indices are proposed based on the 3D wavelet transformation in order to describe the local variation in the joint spectral-spatial domains. In this way, urban landscapes can be decomposed into a set of low-dimensional and semantic indices replacing the high-dimensional but low-level features (e.g., textures). The information indices are then learned via the multi-kernel support vector machines. The proposed MIL method is evaluated using various high-resolution images including GeoEye-1, QuickBird, WorldView-2, and ZY-3, as well as an elaborate comparison to the state-of-the-art image classification algorithms such as object-based analysis, and spectral-spatial approaches based on textural and morphological features. It is revealed that the MIL method is able to achieve promising results with a low-dimensional feature space, and, provide a practical strategy for processing large-scale high-resolution images.

1. Introduction

High-resolution remotely sensed data can provide a large amount of detailed ground information, and open new avenues for remote sensing applications such as precise land use/land cover mapping, landscape analysis, urban facility retrieval, regional environment monitoring. However, increase of the spatial resolution does not signify increase of the image processing accuracy. Specifically, increase of the intra-class variation and decrease of the inter-class variation lead to reduction of the separability of the spatial patterns in the spectral domain (Bruzzone and Carlin, 2006). As a result, the traditional pixel-based and spectral-based image classification techniques are inadequate for high-resolution data (Huang and Zhang, 2013). In this context, researchers proposed to exploit spatial information for complementing the spectral feature space and enhancing separability of the spectrally similar classes. The so-called spectral-spatial classification methods can be divided into the following two categories.

1.1. Exploration of spatial features

In this case, the structural and textural features are used as additional bands to enhance the spectral information and raise accuracies of high-resolution image classification. The commonly used spatial bands include wavelet textures (Ouma and Tateishi, 2008), gray-level co-occurrence matrix (GLCM) (Aguera et al., 2008), pixel shape index (Huang et al., 2007), geometric image features (Inglada 2007), Gabor textural features (Reis and Tasdemir, 2011), morphological filters (Pingel et al., 2013), and morphological profiles (Fauvel et al., 2012).

* Corresponding author. Tel.: +86 27 68778525.
E-mail address: huang_whu@163.com (X. Huang).

http://dx.doi.org/10.1016/j.isprsjprs.2014.01.008
0924-2716/

On the other hand, in recent years, the integrative algorithms combining multiple features have received increasing interest since a single feature descriptor is not adequate for representing complex high-resolution urban scenes. Huang and Zhang (2013) proposed to integrate the spectral, structural, and semantic features for classification of multi/hyperspectral imagery with high spatial resolution. Dell'Acqua et al. (2009) exploited boundary, textural, and morphological features for rapid mapping of high resolution SAR scenes. Fusion of LiDAR and optical data also received much attention for urban scene classification (Guo et al., 2011) and residential building detection (Awrangjeb et al., 2010), since the height information extracted from LiDAR is a useful feature source for urban images.

1.2. Object-based image analysis (OBIA)

Its basic idea is to segment the spatially adjacent pixels into spectrally similar homogeneous objects and then conduct image analysis on objects as the minimum processing unit (Blaschke, 2010). Compared to the pixel-based approach, advantages of the OBIA mainly lie in that it is able to reduce the local spectral variation and intra-class variance, and, hence, avoid the salt-pepper effect. OBIA is an active research area for remote sensing image interpretation. Some studies focus on segmentation, which is the core step of OBIA, such as adaptive mean-shift procedure (Huang and Zhang, 2008), anisotropic morphological leveling (Tzotsos et al., 2011), boundary-constrained multi-scale segmentation (Zhang et al., 2013). Other relevant literature refers to scale parameter selection (Liu et al., 2012) and optimization (Johnson and Xie, 2011) in order to adapt the segmentation. In recent years, OBIA has been successfully applied to mapping private gardens (Mathieu et al., 2007), urban objects extraction (Sebari and He, 2013), change detection (Hussain et al., 2013), tree crown delineation (Jing et al., 2012), spatial pattern analysis of vegetation cover (Yang et al., 2013), etc.

In spite of promising progress achieved for high-resolution image processing, few studies have paid attention to the practical techniques for information extraction from large-scale high-resolution remote sensing data. The spectral-spatial approach, which is the existing mainstream strategy for high resolution image classification, is always subject to the problem of high-dimensional feature space, constructed by a series of parameters such as scales, directions, statistical measures, and basis images. The high-dimensional space poses a big challenge to storage and computation cost for applications of high-resolution data.

In order to address this problem, we propose an innovative multi-index learning (MIL) method for urban scene classification based on high-resolution imagery. The notable characteristic of the MIL is to describe the urban landscape using a set of low-dimensional semantic indices that replace the high-dimensional and low-level features. The first step of the MIL framework is to decompose urban scenes into a series of information indices. In this study, two categories of indices are proposed.

1) Primitive indices (PI): Urban landscapes can be represented by a group of basic elements such as buildings, shadow, and vegetation. These basic elements are calculated automatically without training samples via the morphological building index (MBI) (Huang and Zhang, 2011), morphological shadow index (MSI) (Huang and Zhang, 2012), and normalized difference vegetation index (NDVI). It should be underlined that these indices actually provide more information than the three classes. They also contain information for non-buildings, non-shadow, and non-vegetation, respectively. Therefore, these primitive indices are potential for description of the semantic feature space in urban image scenes.
2) Variation indices (VI): Spectral and spatial variation indices are extracted by 3D wavelet transformation (3D-WT). The notable property of 3D-WT is that it processes an image as a cube, and, therefore, simultaneously describes variation information in the joint spectral-spatial feature space (Yoo et al., 2009). Accordingly, based on the 3D-WT, we propose a couple of variation indices as a representation of spectral and spatial information in a local image scene.

PI and VI are used to describe stationary (basic elements) and dynamic (variation information) features, respectively. Subsequently, we propose to use the multi-kernel learning (Tuia et al., 2010) for interpretation of the multi-index features. The proposed multi-index learning (MIL) method is tested through a series of sophisticated experiments, conducted on the GeoEye-1, QuickBird, WorldView-2, and ZY-3 images. The first three datasets are used for evaluation of MIL method by a detailed comparison to the existing state-of-the-art high-resolution image classification methods, e.g., object-based image analysis (Bruzzone and Carlin, 2006; Blaschke, 2010), spectral-spatial classification using multiscale and multidirectional gray level co-occurrence matrix (GLCM) (Pesaresi et al., 2008), and differential morphological profiles (DMP) (Pesaresi and Benediktsson, 2001). Furthermore, performance of the MIL method is then assessed using a large-scale high-resolution data, ZY-3 satellite image over the urban area of Wuhan city (260 km^2) in central China. In this case, the high-dimensional textural and structural features (e.g., GLCM, DMP) are impractical due to the limitation of memory and processor of a personal computer. In addition, ZY-3 satellite, launched on 9th January 2012, is the China's first civilian high-resolution satellite. To our best knowledge, this study is the first assessment of ZY-3 satellite imagery for information extraction and urban mapping.

The remainder of this paper is organized as follows. The multi-index feature extraction is described in Section 2. Then, the multi-index learning is introduced in Section 3. Experimental results and the comparative study are presented in Section 4, followed by the conclusions and remarks in Section 5.

2. Multi-index feature extraction

This section describes the primitive and variation indices, as well as a demonstration of the multi-index urban scene description.

2.1. Primitive indices (PI)

PI includes a group of basic urban classes: buildings, shadow, and vegetation, which are automatically calculated using MBI, MSI, and NDVI, respectively. Information for other primitive classes (e.g., roads, water) is implicitly contained in the three indices. For instance, MBI is able to discriminate between buildings and roads, while water and shadow can be distinguished based on MSI and NDVI.

2.1.1. Morphological building index (MBI)

The basic idea of MBI is to build the relationship between the spectral-spatial characteristics of buildings (e.g., local contrast, size, isotropy, and brightness) and the morphological operators (Huang and Zhang, 2011). It is constructed based on the fact that the relatively high reflectance of roofs and the spatially adjacent shadows lead to high local contrast of buildings (Pesaresi et al., 2008; Huang and Zhang, 2011). The calculation of MBI is briefly described as follows.