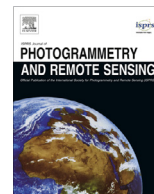Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Review Article

A survey on object detection in optical remote sensing images

Gong Cheng, Junwei Han *

School of Automation, Northwestern Polytechnical University, Xi'an 710072, China

A B S T R A C T

Object detection in optical remote sensing images, being a fundamental but challenging problem in the field of aerial and satellite image analysis, plays an important role for a wide range of applications and is receiving significant attention in recent years. While enormous methods exist, a deep review of the literature concerning generic object detection is still lacking. This paper aims to provide a review of the recent progress in this field. Different from several previously published surveys that focus on a specific object class such as building and road, we concentrate on more generic object categories including, but are not limited to, road, building, tree, vehicle, ship, airport, urban-area. Covering about 270 publications we survey (1) template matching-based object detection methods, (2) knowledge-based object detection methods, (3) object-based image analysis (OBIA)-based object detection methods, (4) machine learning-based object detection methods, and (5) five publicly available datasets and three standard evaluation metrics. We also discuss the challenges of current studies and propose two promising research directions, namely deep learning-based feature representation and weakly super-vised learning-based geospatial object detection. It is our hope that this survey will be beneficial for the researchers to have better understanding of this research field.

© 2016 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.

1. Introduction

Object detection in optical remote sensing images (RSIs) is to determine if a given aerial or satellite image contains one or more objects belonging to the class of interest and locate the position of each predicted object in the image. The term 'object' used in this survey refers to its generalized form, including man-made objects (e.g. vehicles, ships, buildings, etc.) that have sharp boundaries and are independent of background environment, as well as landscape objects, such as land-use/land-cover (LULC) parcels that have vague boundaries and are parts of background environment. As a fundamental problem in the field of aerial and satellite image analysis, object detection in optical RSIs plays an important role for a wide range of applications, such as environmental monitoring, geological hazard detection, LULC mapping, geographic information system (GIS) update, precision agriculture, and urban planning.

Object detection in optical RSIs often suffers from several increasing challenges including the large variations in the visual appearance of objects caused by viewpoint variation, occlusion, background clutter, illumination, shadow, etc., the explosive growth of RSIs in quantity and quality, and the various require-ments of new application areas. To address these challenges, the topic of geospatial object detection has been extensively studied since the 1980s. The low spatial resolution of earlier satellite images (such as Landsat) would not allow the detection of separate man-made or natural objects. Therefore, researchers mostly focused on extracting the region properties from these images. With the advances of remote sensing technology, the very high res-olution (VHR) satellite (e.g. IKONOS, SPOT-5, and Quickbird) and aerial images have been providing us more detailed spatial and textural information. Aside from region properties, a greater range of man-made objects become recognizable and even can be sepa-rately identified than ever before because of the increased sub-meter resolution. This opens new prospects in the field of auto-matic detection of geospatial objects.

During the last decades, considerable efforts have been made to develop various methods for the detection of different types of objects in satellite and aerial images, such as roads (Barsi and Heipke, 2003; Barzohar and Coope, 1996; Chaudhuri et al., 2012; Das et al., 2011; Hu et al., 2007; Huang and Zhang, 2009; Kim et al., 2004; Laptev et al., 2000; Leninisha and Vani, 2015; Li et al., 2010; Maillard and Cavayas, 1989; Mayer et al., 2006; McKeown and Denlinger, 1988; Mena, 2003; Mokhtarzade and Zoej, 2007; Movaghati et al., 2010; Song and Civco, 2004; Trinder and Wang, 1998; Ünsalan and Sirmacek, 2012; Wang et al.,

* Corresponding author.
E-mail address: Junweihan2010@gmail.com (J. Han).

2015; Wang and Zhang, 2011; Zhang et al., 2011a; Zhang and Couloigner, 2006; Zhou et al., 2006; Zhu et al., 2005), buildings (Ahmadi et al., 2010; Akçay and Aksoy, 2010; Aytekın et al., 2012; Benedek et al., 2012; Durieux et al., 2008; Hofmann et al., 2002; Karantzalos and Paragios, 2009; Lefèvre et al., 2007; Lhomme et al., 2009; Mayer, 1999; Ok, 2013; Ok et al., 2013; Peng and Liu, 2005; Peng et al., 2005; Senaras et al., 2013; Shufelt, 1999; Sirmaçek and Ünsalan, 2009; Sirmacek and Ünsalan, 2011; Stankov and He, 2013, 2014; Wegner et al., 2011a, 2011b), trees (Haala and Brenner, 1999; Hung et al., 2012; Malek et al., 2014; Moustakidis et al., 2012; Yang et al., 2013), and vehicles (Eikvil et al., 2009; Grabner et al., 2008; Jin and Davis, 2007; Kembhavi et al., 2011; Leitloff et al., 2010; Moon et al., 2002; Niu, 2006; Tuermer et al., 2013; Wen et al., 2015; Yang et al., 2013, 2015; Zhao and Nevatia, 2003; Zheng et al., 2013). While enormous methods exist, a deep review of the literature concerning generic object detection is still lacking. This paper aims to provide a review of the recent progress in this field. Different from several previously published surveys that focus on a specific object class such as building (Mayer, 1999; Shufelt, 1999) and road (Mayer et al., 2006; Mena, 2003), we concentrate on more generic object categories including, but are not limited to, road, building, tree, vehicle, ship, airport, urban-area. Covering about 270 publications we survey (1) template matching-based object detection methods, (2) knowledge-based object detection methods, (3) object-based image analysis (OBIA)-based object detection methods, (4) machine learning-based object detection methods, and (5) five publicly available datasets and three standard evaluation metrics for object detection. We also discuss open problems and challenges of current studies, and propose two promising research directions in future for constructing more effective object detection framework. This survey will be especially beneficial for the researchers to have better understanding of this research field. Furthermore, to the best of our knowledge, this is the first survey paper in the literature that focuses on generic object detection in optical RSIs.

The rest of the paper is organized as follows. Section 2 briefly introduces the taxonomy of methods for object detection. In Sections 3–6, we exhaustively review template matching-based object detection methods, knowledge-based object detection methods, OBIA-based object detection methods, and machine learning-based object detection methods, respectively. In Section 7, we review five publicly available datasets and three standard evaluation metrics. Section 8 discusses the challenges of current studies and proposes two promising research directions to advance the field. Finally, conclusions are drawn in Section 9.

2. Taxonomy of methods for object detection

In the last decades, a large number of methods have been developed for object detection from aerial and satellite images. We can generally divide them into four main categories: template matching-based methods, knowledge-based methods, OBIA-based methods, and machine learning-based methods. These four categories are not necessarily independent and sometimes the same method exists with different categories. Fig. 1 shows a taxonomy of geospatial object detection studies, in which rounded rectangles with solid borders illustrate our scope in this paper.

According to the template type selected by a user, template matching-based methods are further divided into two classes as rigid template matching and deformable template matching. As for knowledge-based object detection methods, we mainly review two kinds of most widely used prior knowledge, namely geometric information and context information. Generally, OBIA-based object detection methods involve two steps: image segmentation and object classification. With regards to machine learning-based methods, we mainly focus on reviewing three crucial steps that play important roles in the performance of object detection. They are feature extraction, optional feature fusion and dimension reduction, and classifier training. In the step of feature extraction, we introduce five types of typical features including Histogram
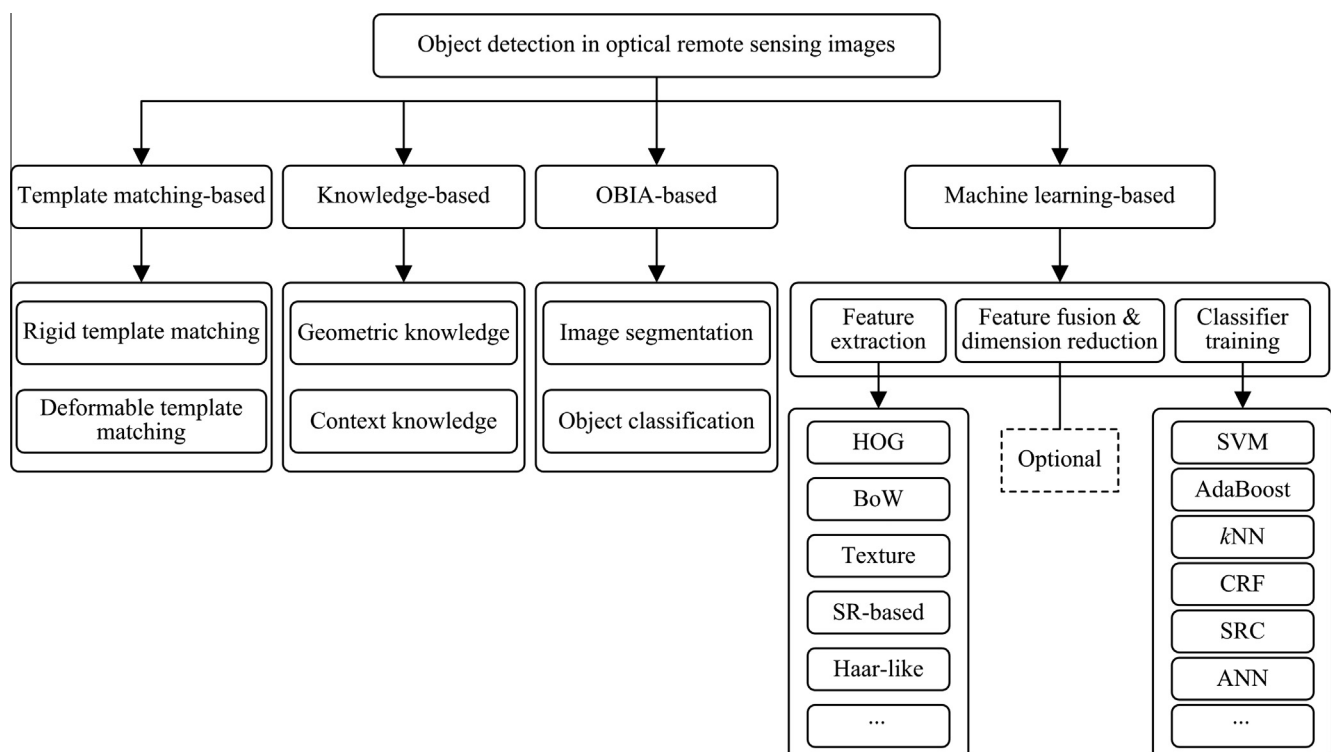

Fig. 1. Taxonomy of methods for object detection in optical RSIs. Rounded rectangles with solid borders illustrate our scope in this paper.

of oriented gradients (HOG) feature, bag-of-words (BoW) feature, texture features, sparse representation (SR)-based features, and Haar-like features. In the step of classifier training, we exhaustively review six kinds of machine learning algorithms including support vector machine (SVM), AdaBoost, k-nearest-neighbor (kNN), conditional random field (CRF), sparse representation-based classification (SRC), and artificial neural network (ANN).

3. Template matching-based object detection

Template matching-based methods are one category of the simplest and earliest approaches for object detection. Fig. 2 gives the flowchart of template matching-based object detection. As shown in Fig. 2, there are two main steps in template matching-based object detection framework. (1) Template generation: a template T for each to-be-detected object class should be firstly generated by hand-crafting or learning from the training set. (2) Similarity measure: given a source image, the stored template T is used to match the image at each possible position to find the best matches, according to the minimum distortion or maximum correlation measures, while taking into account all allowable translation, rotation, and scale changes. The most popular similarity measures are the sum of absolute differences (SAD), the sum of squared differences (SSD), the normalized cross correlation (NCC), and the Euclidean distance (ED). According to the template type selected by a user, the template matching-based object detection approaches are generally categorized into two groups: rigid template matching and deformable template matching.

3.1. Rigid template matching

Early research in this area mainly concentrated on rigid template matching. Various rigid templates have been designed for detecting specific objects with simple appearance and small variations such as road (Chaudhuri et al., 2012; Kim et al., 2004; McKeown and Denlinger, 1988; Zhang et al., 2011a; Zhou et al., 2006). For instance, McKeown and Denlinger (1988) introduced a road tracking method based on road profile correlation, in which the distortion of a reference profile and the target profile is measured by identifying two geometric parameters (shift and width) and two radiometric parameters (brightness and contrast). Zhou et al. (2006) introduced a road tracking system by using two profiles: one orthogonal to the road direction and the other parallel to the road direction. Kim et al. (2004) utilized a rectangular template instead of the profiles to track ribbon roads in urban areas via least squares correlation template matching method. Zhang et al. (2011a) proposed a semi-automatic template matching method to track roads, in which a rectangular reference template is generated by user inputting a seed point on a selected road and using a spoke wheel algorithm to obtain road direction, road width, and starting point. The matching is then performed by calculating the SSD and Euclidean distance transformation of the reference template and a target template.

The morphological hit-or-miss transform (HMT) is a powerful tool dedicated to template matching. Since its first definition for binary images (Lefèvre et al., 2007), its recent extensions to gray-scale images (Stankov and He, 2013, 2014; Zheng et al., 2013) and multispectral images (Weber and Lefèvre, 2008, 2012) have proven its ability to solve various template matching problems. Lefèvre et al. (2007) presented a method for building extraction in Quickbird images based on an adaptive binary HMT with varying sizes and shapes of the structuring element. In this method, a binary image was firstly generated from a gray-level panchromatic input image before building detection. To make use of the spectral information, the authors in (Stankov and He, 2013, 2014) generated grayscale images from the spectral bands and then applied the grayscale HMT to building detection, where (Stankov and He, 2013) is a supervised method that requires a set of reference windows for each roof color present in the image while the method in (Stankov and He, 2014) is unsupervised. Furthermore, Weber and Lefèvre (2008, 2012) introduced a new definition of the HMT for multivariate image analysis and illustrated its potential as a template matching operator for coastline extraction and petroleum tank detection.

However, although the rigid template matching is effective in some applications, it has a number of disadvantages, resulting in their utility is limited. A shortcoming of the rigid template matching is that it requires the template to be very precise, so it is sensitive to shape and density variation. In most applications, an exact geometric template of the object is not available because of the viewpoint change, or large intra-class variations among the objects. For instance, most of the road trackers mentioned above do not work well when they encounter irregular geometric deformation in tracking process, such as appearances of road junctions, material changes, occlusions of cars, and shadows and lane markings.

3.2. Deformable template matching

The concept of deformable template was first introduced to computer vision community by Fischler and Elschlager (1973) with the spring-loaded templates. Deformable template matching is more powerful and flexible than rigid shape matching in dealing with shape deformations and intra-class variations because of its capability to both impose geometrical constraints on the shape and to integrate local image evidence. There has been a substantial amount of studies on deformable template matching in recent years. These studies can be roughly divided into two classes: free-form deformable templates and parametric deformable templates (Jain et al., 1998).

3.2.1. Free-form deformable templates

The free-form deformable templates represent an arbitrary object shape by constraining some general regularization (e.g. continuity, smoothness, etc.) and the most popular methods are the active contour models, also known as snake models. In these approaches, an energy-minimizing contour, called a "snake", is
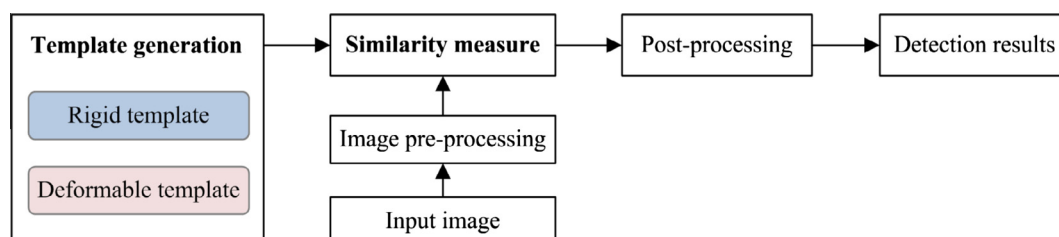

Fig. 2. The flowchart of template matching-based object detection.

controlled by a combination of three forces or energies. Briefly, a snake is modeled as being able to deform elastically, but any deformation increases its internal energy causing a "restitution force", which tries to bring it back to its original shape (Jain et al., 1998). For different applications, various snake-based methods have been developed (Ahmadi et al., 2010; Jing et al., 2011; Laptev et al., 2000; Leninisha and Vani, 2015; Liu et al., 2013b; Niu, 2006; Peng et al., 2005; Wang and Zhang, 2011; Xu and Duan, 2010). Peng et al. (2005) proposed an improved snake model for building detection from gray-level aerial images by modifying the traditional snake model in two aspects: the criteria for the selection of initial seeds and the external energy function. Niu (2006) introduced a semi-automatic framework based on geometric active contour model for highway extraction and vehicle detection from aerial photographs. In (Liu et al., 2013b), the authors proposed an algorithm to extract geospatial objects with regular shape using shape-based global minimization active contour model. Xu and Duan (2010) described a shape-matching approach for aircraft recognition at low altitude, in which an artificial bee colony (ABC) algorithm with edge potential function (EPF) was proposed.

3.2.2. Parametric deformable templates

The parametric deformable templates parameterize a specific object class and its variations by a parametric formula (Barzohar and Coope, 1996; Hung et al., 2012; Lhomme et al., 2009; Liu et al., 2013a; Movaghati et al., 2010) or using a prototype and deformation modes (Karantzalos and Paragios, 2009; Lin et al., 2015; Sirmaçek and Ünsalan, 2009; Sun et al., 2010; Tao et al., 2011), which are commonly used when some prior information of the geometrical shape is available. For instance, Lhomme et al. (2009) proposed a "Discrimination by Ratio of Variance" (DRV) parameter to quantify the spatial distribution of grey-level variation of a building and its close neighborhood for the detection of buildings. Liu et al. (2013a) proposed an aircraft recognition method in high-resolution satellite images using coarse-to-fine

shape prior, in which a parametric shape model for aircraft is derived by applying principal component analysis (PCA) and kernel density function. In (Sirmaçek and Ünsalan, 2009), the authors firstly defined two template building images, one for a bright building and the other one for a dark building, and represented them by scale invariant feature transform (SIFT) features (Lowe, 2004). Then, the urban area was extracted using a multiple subgraph matching method and separate buildings in the urban area were extracted using a graph cut method.

A brief summary of template matching-based object detection methods is given in Table 1.

4. Knowledge-based object detection

Knowledge-based object detection methods are another type of popular approaches for object detection in optical RSIs. An extensive collection of papers on knowledge-based object detection have been published for buildings in (Akçay and Aksoy, 2010; Haala and Brenner, 1999; Hofmann et al., 2002; Huertas and Nevatia, 1988; Irvin and McKeown, 1989; Lin and Nevatia, 1998; Liow and Pavlidis, 1990; McGlone and Shufelt, 1994; Ok, 2013; Ok et al., 2013; Peng and Liu, 2005; Shufelt, 1996; Stilla et al., 1997; Weidner and Förstner, 1995), roads in (Hu et al., 2007; Leninisha and Vani, 2015; Maillard and Cavayas, 1989; Trinder and Wang, 1998; Wang and Newkirk, 1988; Zhu et al., 2005), and other more general object extraction applications like landslide, bridges, vehicles, urban land changes, crops, drainage channels, forests in (Chaudhuri and Samal, 2008; Janssen and Middelkoop, 1992; Martha et al., 2011; Moon et al., 2002; Solberg, 1999; Tchoku et al., 1996; Wang, 1993). A review of knowledge-based object extraction in RSIs is also given in (Baltsavias, 2004). This type of approaches generally translates object detection problem into hypotheses testing problem by establishing various knowledge and rules. Fig. 3 gives the flowchart of knowledge-based object detection. As shown in Fig. 3, the establishment of knowledge and rules is the most important step.

Table 1
A brief summary of template matching-based object detection methods.

| Templates	Objects and publications	Strengths	Limitations
Rigid template	Building (Lefèvre et al., 2007; Stankov and He, 2013, 2014); coastline and storage tank (Weber and Lefèvre, 2008, 2012); road (Chaudhuri et al., 2012; Kim et al., 2004; McKeown and Denlinger, 1988; Zhang et al., 2011a; Zhou et al., 2006)	Simple and easy to implement	Scale and rotation dependent; Sensitive to shape and viewpoint change
Deformable template	Airplane, ship, and storage tank (Lin et al., 2015; Liu et al., 2013a, 2013b; Sun et al., 2010; Xu and Duan, 2010); airport (Tao et al., 2011); building (Ahmadi et al., 2010; Karantzalos and Paragios, 2009; Sirmaçek and Ünsalan, 2009); edges of oil slick (Jing et al., 2011); road (Barzohar and Coope, 1996; Laptev et al., 2000; Leninisha and Vani, 2015; Lhomme et al., 2009; Movaghati et al., 2010; Niu, 2006; Peng et al., 2005; Wang and Zhang, 2011); tree crown (Hung et al., 2012)	More powerful and flexible than rigid shape matching in dealing with shape deformations and intra-class variations	Need more prior information and parameters of the geometrical shape for template designing; Computationally expensive
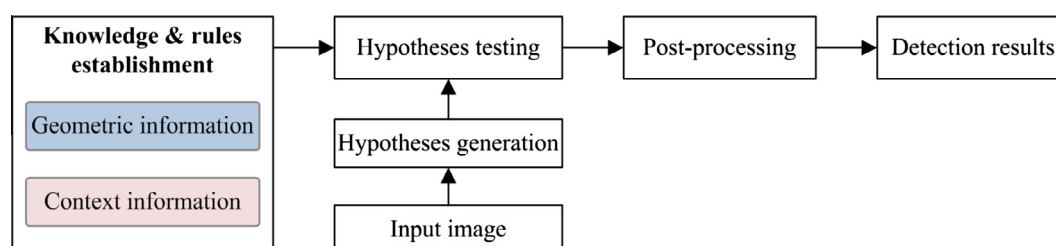
Fig. 3. The flowchart of knowledge-based object detection.

Two kinds of widely used knowledge on target objects are geometric knowledge (Huertas and Nevatia, 1988; McGlone and Shufelt, 1994; Shufelt, 1996; Trinder and Wang, 1998; Weidner and Förstner, 1995) and context knowledge (Akçay and Aksoy, 2010; Huertas and Nevatia, 1988; Irvin and McKeown, 1989; Lin and Nevatia, 1998; Ok, 2013; Ok et al., 2013; Peng and Liu, 2005; Stilla et al., 1997).

4.1. Geometric knowledge

The object geometric information is the most important and widely used knowledge for object detection, which encodes prior knowledge by taking parametric specific or generic shape models (Huertas and Nevatia, 1988; Leninisha and Vani, 2015; McGlone and Shufelt, 1994; Trinder and Wang, 1998; Weidner and Förstner, 1995). For example, Trinder and Wang (1998) proposed a road model including geometric and radiometric properties, in which the hypotheses of roads are generated using hand-crafted rules, and a top-down process is applied to verify the road hypotheses. Huertas and Nevatia (1988) assumed that the buildings are rectangular or composed of rectangular components (e.g. "box," "T," "L," and "E" shapes) and used a generic model of the shapes to detect buildings. Weidner and Förstner (1995) developed an approach for extracting 3D shape of buildings from high-resolution digital elevation models (DEMs) by establishing and using explicit geometric constraints knowledge in the form of parametric and prismatic building models. McGlone and Shufelt (1994) proposed to include the target geometric and metric knowledge into the building extraction system for the generation of building hypotheses, and the generated hypotheses were finally verified with the shadow information.

4.2. Context knowledge

The context knowledge is another crucial cue for knowledge-based object detection and the most widely used context knowledge is the spatial constraints or relationships between objects and background, or the information regarding how the object interacts with its neighboring regions (Akçay and Aksoy, 2010; Chaudhuri and Samal, 2008; Liow and Pavlidis, 1990; Ok et al., 2013; Peng and Liu, 2005). As a representative example, shadow evidence has been considered to be one of the most important clues for building detection (Irvin and McKeown, 1989; Lin and Nevatia, 1998; Liow and Pavlidis, 1990; Ok, 2013; Ok et al., 2013). For instance, Irvin and McKeown (1989) exploited the relationship between man-made structures and their cast shadows to predict the locations and shape of buildings, and Liow and Pavlidis (1990) used shadow information to complete the boundary grouping process. Ok et al. (2013) employed the shadow evidence to automatically detect buildings with arbitrary shapes from monocular VHR images. In their work, the authors modeled the directional spatial relationship between buildings and their shadows by proposing a new fuzzy landscape generation approach. Besides, the shadow areas derived from monocular images were extensively used during the verification of the building hypotheses. For example, Lin and Nevatia (1998) proposed an approach to detect buildings from oblique aerial images, in which the hypothesized rectangular buildings are verified with both shadow and wall evidences by following the basic assumption that the building shapes are rectilinear. Peng and Liu (2005) developed a shadow-context model to extract buildings in dense urban aerial images via combining shadow information with context to verify building regions.

It is worth noting that the core of knowledge-based object detection methods is how to effectively transform the implicit knowledge understanding on target objects into the explicit detection rules. If the defined rules are too strict, some target objects will be missed; conversely, too loose rules will cause false positives. A brief summary of knowledge-based object detection methods is given in Table 2.

5. OBIA-based object detection

Recently, with the increasing availability and wide utilization of sub-meter imagery, object-based image analysis (OBIA or GEOBIA for geospatial object based image analysis) (Blaschke, 2010; Blaschke et al., 2008) has become a new methodology or paradigm (Blaschke et al., 2014) to classify or map VHR imagery into meaningful objects (or rather, grouping of relatively local homogeneous pixels). Fig. 4 gives the flowchart of OBIA-based object detection. As shown in Fig. 4, OBIA involves two steps: image segmentation and object classification. Firstly, imagery is first segmented into homogeneous regions (segments also called objects) representing a relatively homogeneous group of pixels by selecting desired scale, shape, and compactness criteria. And in a second step, a classification process is applied to these objects. Since OBIA offers the potential to exploit geographical information system (GIS) functionality, such as the incorporation of the spatial context or object shape in the classification, it provides a framework for overcoming the limitations of conventional pixel-based image classification methods and has been successfully applied to landslide mapping (Feizizadeh et al., 2014; Li et al., 2015b; Martha et al., 2010, 2011, 2012; Stumpf and Kerle, 2011), land cover and land use mapping (Baker et al., 2013; Benz et al., 2004; Blaschke, 2003; Blaschke et al., 2008, 2011; Contreras et al., 2015; D'Oleire-Oltmanns et al., 2014; De Pinho et al., 2012; Doleire-Oltmanns et al., 2013; Drăguţ and Blaschke, 2006; Drăguţ and Eisank, 2012; Duro et al., 2012; Eisank et al., 2011; Goodin et al., 2015; Hay et al., 2003; Hofmann et al., 2011; Kim et al., 2011; Leon and Woodroffe, 2011; Li et al., 2014; Li and Shao, 2013;

Table 2
A brief summary of knowledge-based object detection methods.

| Objects and publications	Strengths	Limitations
Bridge (Chaudhuri and Samal, 2008); building (Akçay and Aksoy, 2010; Haala and Brenner, 1999; Hofmann et al., 2002; Huertas and Nevatia, 1988; Irvin and McKeown, 1989; Lin and Nevatia, 1998; Liow and Pavlidis, 1990; McGlone and Shufelt, 1994; Ok, 2013; Ok et al., 2013; Peng and Liu, 2005; Shufelt, 1996; Stilla et al., 1997; Weidner and Förstner, 1995); channel (Tchoku et al., 1996); crops (Janssen and Middelkoop, 1992); drainage forest (Solberg, 1999); landslide (Martha et al., 2011); road (Hu et al., 2007; Leninisha and Vani, 2015; Maillard and Cavayas, 1989; Trinder and Wang, 1998; Wang and Newkirk, 1988; Zhu et al., 2005); urban land change (Wang, 1993); vehicle (Moon et al., 2002)	Detection can be performed through coarse-to-fine hierarchical structure	How to define prior knowledge and detection rules is subjective; Too loose rules will cause false positives and vice versa
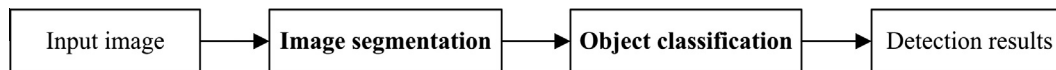
Fig. 4. The flowchart of OBIA-based object detection.

Lisita et al., 2013; Macfaden et al., 2012; Mallinis et al., 2008; Mishra and Crews, 2014; Moskal et al., 2011; Myint et al., 2011; Phinn et al., 2012; Tzotsos et al., 2011; Walker and Blaschke, 2008; Walker and Briggs, 2007; Weng, 2009, 2011; Xie et al., 2008; Yu et al., 2006; Zhou, 2013; Zhou et al., 2009; Zhou and Troy, 2008), and change detection (Bontemps et al., 2008; Chen and Hay, 2012; Contreras et al., 2016; Dissanska et al., 2009; Doxani et al., 2008, 2012; Hussain et al., 2013; Im et al., 2008; Nebiker et al., 2014; Walter, 2004). Two reviews of the literature on OBIA/GEOBIA for remote sensing are also given in (Blaschke, 2010; Blaschke et al., 2014).

Image segmentation is the first step and a necessary prerequisite for generating the building blocks of OBIA based on a type of segmentation algorithm, the color, texture and shape of groups of pixels, and the required level or scale of spatial detail. The delineation quality of the target objects has a direct influence on the accuracy of the subsequent image classification. During the last decades, numerous image segmentation techniques have been developed and applied in remote sensing image analysis (Ardila et al., 2012; Baatz and Schäpe, 2000; Benz et al., 2004; Blaschke et al., 2004; Drăguţ et al., 2014; Drăguţ et al., 2010; Esch et al., 2008; Gao et al., 2011; Hay et al., 2005; Jungho et al., 2014; Kim et al., 2008, 2011; Ming et al., 2015). However, as Hay et al. (2005) pointed out, the real challenge is to define appropriate segmentation parameters (typically based on spectral homogeneity, size, or both) for the varying sized, shaped, and spatially distributed image-objects composing a scene, so that segments can be generated to satisfy user requirements.

Multi-resolution segmentation (MRS) algorithm (Baatz and Schäpe, 2000) is probably the most popular one for the purpose of the delineation of relatively homogeneous and meaningful objects. Implemented in the eCognition® software (Trimble Geospatial Imaging), this algorithm has quickly became one of the most important segmentation algorithms within the OBIA domain. MRS uses three parameters to partition an image into objects: shape, compactness, and scale. The shape parameter defines to which percentage the homogeneity of shape is weighted against the homogeneity of spectral values. The compactness parameter is a sub-parameter of shape and is used to optimize image objects with regard to compactness or smoothness. The scale parameter is used for controlling the internal (spectral) heterogeneity of the resulting objects and is therefore correlated with their average size, i.e., a larger value of the scale allows a higher internal heterogeneity, which increases the number of pixels per object and vice versa (Baatz and Schäpe, 2000; Benz et al., 2004; Flanders et al., 2003).

Since the scale is a crucial control parameter in MRS and heavily impacts on the classification accuracy, how to select a more objective value of scale (at least traceable or reproducible) is a hot topic in OBIA (Blaschke, 2010; Blaschke et al., 2014). The traditional procedure for the selection of segmentation scale parameters often depends on subjective trial-and-error methods, which are mainly based on a visual assessment of segmentation suitability. While allowing flexibility in incorporating expert knowledge in OBIA, this procedure is hardly reproducible and raises some limitations with respect to the robustness of the approach. Based on the previous work of Woodcock and Strahler (1987), the concept of local variance (LV) graphs that reveals the spatial structure of images was introduced in the context of OBIA by Kim et al. (2008) to determine the optimal segmentation scale for alliance-level forest classification of multispectral IKONOS images. Drăguţ et al. (2010) carried

the concept of LV further and presented a tool called Estimation of Scale Parameters (ESP) to objectively identify the most suitable range of scale parameters. In parallel, aiming at the automation of the segmentation process in OBIA, a suite of techniques (Ardila et al., 2012; Drăguţ et al., 2014; Drăguţ and Eisank, 2012; Esch et al., 2008; Hay et al., 2005) have also been developed.

Once segments are generated, one can extract object features, such as spectral information as well as size, shape, texture, geometry, and contextual semantic features. In addition, GIS-like functionality and expert knowledge can also be flexibly incorporated in classification procedures to make OBIA context-aware and multi-source capable (Blaschke et al., 2014). These features are then selected (optional, e.g. the work of Stumpf and Kerle (2011)) and fed to a classifier (e.g. membership function classifier, nearest neighbor classifier, decision tree, neural network, SVM, etc.) for classification.

For OBIA, accuracy assessment that measures how well objects extracted from high-resolution images match existing geographic objects is very important (Blaschke et al., 2008, 2014; Congalton and Green, 2009; Drăguţ et al., 2010, 2014). It has been suggested by Congalton and Green (2009) that objects, instead of pixels, should be used as sampling units for the thematic accuracy assessment of object-based classification results. Thus, object-based accuracy assessment should account for the thematic accuracy of the class labels as well as the spatial characteristics of represented objects. Recently, a number of researchers have proposed several object-based metrics for assessing both geometric and thematic accuracies of object-based image classification (Albrecht, 2010; Clinton et al., 2010; Lizarazo, 2014; MacLean and Congalton, 2012; Radoux et al., 2011; Zhan et al., 2005; Zhen et al., 2013). However, accuracy assessments of OBIA are still a complex subject and some problems remain to be solved, such as (1) the impossibility of specifying a single, all-purpose measure of classification accuracy and (2) the selection of an appropriate sample size as well as the specification and the use of a measure of accuracy appropriate to the application in-hand.

A brief summary of OBIA-based object detection methods is given in Table 3.

6. Machine learning-based object detection

With the advance of machine learning techniques, especially the powerful feature representations and classifiers, many recent approaches regarded object detection as a classification problem and have achieved significant improvements. Fig. 5 gives the flowchart of machine learning-based object detection, in which object detection can be performed by learning a classifier that captures the variation in object appearances and views from a set of training data in a supervised or semi-supervised or weakly supervised framework. The input of the classifier is a set of regions (sliding windows or object proposals) with their corresponding feature representations and the output is their corresponding predicted labels, i.e., object or not. As can been seen from Fig. 5, feature extraction, feature fusion and dimension reduction (optional), and classifier training play the most important roles in the performance of object detection and hence we mainly focus on reviewing these three crucial steps.
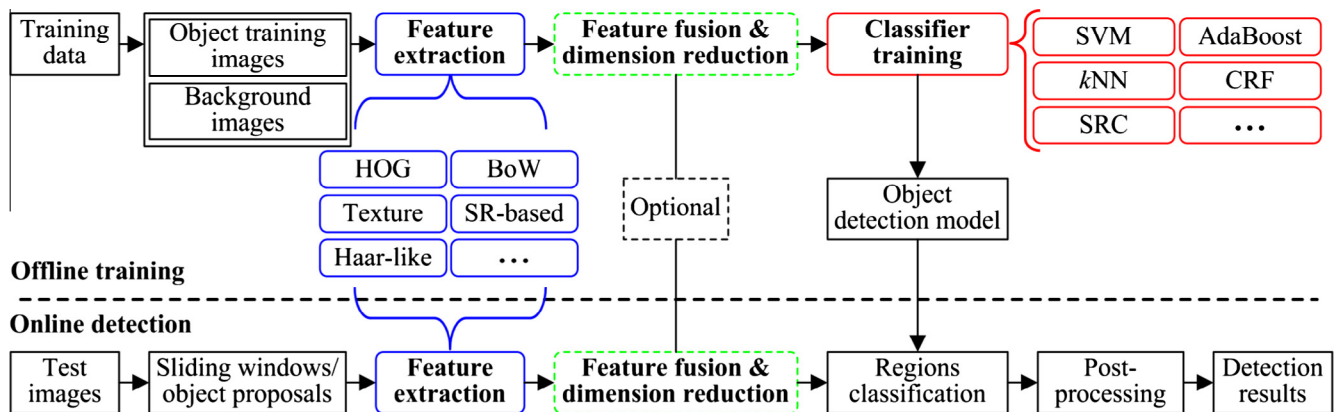
6.1. Feature extraction

Feature extraction is the mapping from raw image pixels to a discriminative high-dimensional data space. Since high-performance object detection is usually carried out in feature

Table 3
A brief summary of OBIA-based object detection methods.

Objects and publications	Strengths	Limitations
Change detection (Bontemps et al., 2008; Chen and Hay, 2012; Contreras et al., 2016; Dissanska et al., 2009; Doxani et al., 2012, 2008; Hussain et al., 2013; Im et al., 2008; Nebiker et al., 2014; Walter, 2004); land cover and land use mapping including vegetation, tree, water, residential, etc. (Baker et al., 2013; Benz et al., 2004; Blaschke, 2003; Blaschke et al., 2011, 2008; Contreras et al., 2015; D'Oleire-Oltmanns et al., 2014; De Pinho et al., 2012; Doleire-Oltmanns et al., 2013; Drăguţ and Blaschke, 2006; Drăguţ and Eisank, 2012; Duro et al., 2012; Eisank et al., 2011; Goodin et al., 2015; Hay et al., 2003; Hofmann et al., 2011; Kim et al., 2011; Leon and Woodroffe, 2011; Li et al., 2014; Li and Shao, 2013; Lisita et al., 2013; Macfaden et al., 2012; Mallinis et al., 2008; Mishra and Crews, 2014; Moskal et al., 2011; Myint et al., 2011; Phinn et al., 2012; Tzotsos et al., 2011; Walker and Blaschke, 2008; Walker and Briggs, 2007; Weng, 2009, 2011; Xie et al., 2008; Yu et al., 2006; Zhou, 2013; Zhou et al., 2009; Zhou and Troy, 2008); landslide mapping (Feizizadeh et al., 2014; Li et al., 2015b; Martha et al., 2010, 2011, 2012; Stumpf and Kerle, 2011)	The flexible incorporation of shape, texture, geometry, and contextual semantic features, as well as GIS-like functionality and expert knowledge, makes OBIA context-aware and multi-source capable	Generic solutions to the full automation of segmentation process are still missing; The expert knowledge for how to define the classification rules are still subjective

Fig. 5. The flowchart of machine learning-based object detection.

space, the feature extraction is quite important to construct high-performance object detection systems. Here we mainly focus on reviewing some widely used features for object detection in optical RSIs. Some reviews on feature extraction and feature coding in computer vision are available in (Andreopoulos and Tsotsos, 2013; Huang et al., 2014; Li et al., 2015c).

6.1.1. Histogram of oriented gradients (HOG) feature and its extensions

HOG feature was first proposed by Dalal and Triggs (2005) to represent objects by the distribution of gradient intensities and orientations in spatially distributed regions, which has been widely acknowledged as one of the best features to capture the edge or local shape information of the objects. Since its introduction, it has shown great success in many geospatial object detection algorithms (Cheng et al., 2013b; Cheng et al., 2014a, 2014b; Grabner et al., 2008; Kembhavi et al., 2011; Tuermer et al., 2013). Besides, some part model-based methods (Cheng et al., 2013b, 2014a, 2014b; Zhang et al., 2014b, 2015b) and the sparelets work (Cheng et al., 2015a, 2015b, 2015c) that built on the HOG feature have also shown impressive performance.

The implementation of HOG feature extraction can be briefly summarized as follows. (1) Calculate gradients at orthogonal directions with centered derivative mask $[-1, 0, +1]$. (2) Compute gradient magnitudes and orientations for each pixel point. (3) Divide each image into fine and non-overlapping spatial regions called cells. (4) Accumulate a local one-dimensional histogram of gradient orientations over the pixels within each cell. The gradient at each pixel is quantized into one of nine orientation bins. Each pixel

votes into the corresponding orientation bin with a voting weight based on the gradient magnitude. (5) Spatially connected cells are grouped into coarser blocks. (6) Run a normalization process on each block, by the division of the gradient energy inside the block, to provide strong illumination invariance and to reduce the sensitivity to gradient strength. Suppose $\boldsymbol{v}$ is the descriptor vector from a block, and the ℓ_p norm of $\boldsymbol{v}$ is used for normalization. Typical choices are

$$\ell_2\text{-norm}: \ \boldsymbol{v} \leftarrow \boldsymbol{v}\Big/\sqrt{\|\boldsymbol{v}\|_2^2 + \xi^2}, \quad \ell_1\text{-norm}: \ \boldsymbol{v} \leftarrow \boldsymbol{v}/(\|\boldsymbol{v}\|_1 + \xi)$$

(1)

where ξ is a small constant and the results are insensitive to the value of ξ over a large range. The HOG feature is the concatenation of normalized vectors in distributed blocks. Since histograms are computed over regions, they are fairly robust to some variability in the location of the parts of the object. The HOG descriptor is also invariant to rotations smaller than the size of the histogram orientation bin.

To further enhance the description capability of HOG for optical RSIs, several extensions are developed (Shi et al., 2014; Zhang et al., 2014b, 2015b). In (Zhang et al., 2014b), rotation invariant HOG feature was achieved by rotating the coordinates of the descriptor and the gradient orientations according to the dominant orientation as SIFT descriptor (Lowe, 2004). Shi et al. (2014) developed a circle frequency-HOG feature for ship detection by combining circle frequency feature and HOG feature. Zhang et al. (2015b) introduced a strategy to generate a variant of HOG, namely HOG normalized by polar angle, to steer the rotation problem, in which the gradient

direction of each pixel is normalized by its polar angle and then the HOG is generated according to the new directions.

6.1.2. Bag-of-words (BoW) feature

Among various features developed for visual object recognition, the BoW model (Li and Perona, 2005) is probably one of the most popular during the last decade. The main advantages of the BoW model is its simplicity, efficiency and invariance under viewpoint changes and background clutter, which was widely adopted by the community and results in good performance for geographic image classification (Xu et al., 2010; Yang and Newsam, 2010, 2011, 2013) and geospatial object detection (Bai et al., 2014; Cheng et al., 2013a; Sun et al., 2012; Zhang et al., 2015a).

There are generally five steps to construct BoW model and we will briefly state these steps. The first step is to detect keypoints from images. Many algorithms have been proposed for keypoint detection such as Harris-Laplacian detector (Mikolajczyk and Schmid, 2001) and difference of Gaussian (DoG) detector (Lowe, 2004). The second step is to compute local descriptors for the detected keypoints. There is a rich literature on local descriptors, among which SIFT descriptor (Lowe, 2004) is the most popular. The third step is to construct the visual vocabulary using clustering technique such as k-means over the computed descriptors. Given the dictionary, the fourth step is to vector quantize each keypoint into a visual word in the dictionary, a step that is often referred to as the encoding module. The last step of the BoW model is the pooling step that pools encoded local descriptors into a global histogram representation. Various pooling strategies have been proposed for the BoW model such as mean and max-pooling. Studies in (Lazebnik et al., 2006) have shown that it is important to take into account the spatial layout of keypoints in the pooling step. A well known example of this approach is spatial pyramid matching (SPM) (Lazebnik et al., 2006) which partitions the image into increasingly finer spatial sub-regions and constructs a histogram for each sub-region separately.

6.1.3. Texture features

Texture features aim to describe the local density variability and patterns inside the surface of an object, which are very important for identifying textural objects such as airport (Aytekin et al., 2013; Tao et al., 2011), buildings (Senaras et al., 2013), urban area (Zhong and Wang, 2007), and vehicles (Eikvil et al., 2009; Grabner et al., 2008). Here we briefly review two commonly used texture features, namely Gabor feature and local binary patterns (LBP) feature.

Gabor feature. The Gabor feature (Jain et al., 1997) is a classical kind of texture features. It is computed by applying a bank of scale and orientation selective Gabor filters to an image. Gabor filters often appear in pairs as even and odd components with the following mathematical form

$$G_{\text{even}}(x,y) = \cos(k_x x + k_y y)\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right), \quad G_{\text{odd}}(x,y)$$
$$= \sin(k_x x + k_y y)\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \qquad (2)$$

where k_x and k_y are spatial frequencies at that the filters respond most strongly. $\arctan(k_x/k_y)$ determines the orientation and σ denotes the scale of the filters. A bank of Gabor filters is generated by selecting different combinations of k_x, k_y, and σ. Gabor filtering responds higher when the local texture changes at the same orientation with the same frequency. The Gabor feature dimension is determined by the image size and the filter bank size.

LBP feature. The LBP feature (Ojala et al., 2002) is a theoretically very simple yet efficient approach for texture description by computing the frequencies of local patterns in sub-regions. The local

pattern is defined with a center pixel p_c, pixel number N and radius r. A set of interpolated surrounding pixels $P = \{p_1, \ldots, p_N\}$ is generated with predefined N and r. The values of surrounding pixels P are then compared with the value of center pixel p_c and a N-bit vector is generated. The LBP code is defined by the decimal value of the N-bit vector

$$LBP_{N,r}(p_c) = \sum_{i=1}^{N} 2^{i-1}\phi(I(p_c), I(p_i)) \qquad (3)$$

where $I(p_c)$ and $I(p_i)$ are the values of pixels p_c and p_i, respectively. $\phi(I(p_c), I(p_i)) = 1$ only if $I(p_c) < I(p_i)$, otherwise it is 0. By computing the decimal LBP code for each point, the frequencies of LBP codes appearing in spatially distributed cells can be recorded as the bin value of the LBP histogram. The bin number is 2^N. Besides, by circularly right shifting the bitwise vector and calculating the minimal value for the histogram binning we can obtain rotation-invariant LBP.

6.1.4. Sparse representation (SR)-based features

Lately, with the development of compressed sensing theory, SR-based features have been extensively applied to hyperspectral image denoising (Zhao and Yang, 2015), hyperspectral image classification (Chen et al., 2011a, 2013; Qian et al., 2013; Tao et al., 2012), and object detection in RSIs (Chen et al., 2011b, 2011c; Cheng et al., 2014c; Du and Zhang, 2014; Han et al., 2014; Liu and Shi, 2014; Yokoya and Iwasaki, 2015; Zhang et al., 2014a, 2015c, 2015d). The core idea of SR is to sparsely encode high-dimensional original signals by a few structural primitives in a low-dimensional manifold. The procedure of seeking the sparsest representation for test sample in terms of an over-complete dictionary endows itself with a discriminative nature to perform classification. The SR-based feature can be generally calculated via resolving a least-square-based optimization problem with constraints on the sparse coefficients

$$\boldsymbol{\alpha} = \arg\min_{\boldsymbol{\alpha}}\{\|\boldsymbol{x} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda\varphi(\boldsymbol{\alpha})\} \qquad (4)$$

where $\|\cdot\|_2$ denotes the ℓ_2 norm, λ is a scalar constant, $\boldsymbol{x} \in R^d$ is an input signal, $\mathbf{D} = [\boldsymbol{d}_1, \ldots, \boldsymbol{d}_M] \in R^{d \times M}$ is an over-complete dictionary with each column $d_i(i = 1, \ldots, M)$ representing an atom, and $\boldsymbol{\alpha} \in R^M$ is a sparse coefficient vector used as the feature of input signal $\boldsymbol{x}$. In general, the number of atoms M is large, whereas the signal dimension d is relatively small. The least-square term $\|\boldsymbol{x} - \mathbf{D}\boldsymbol{\alpha}\|_2^2$ pursues accurate reconstruction, i.e., a signal can be described with a small error, and the constraint term $\varphi(\boldsymbol{\alpha})$ pursues discriminative description, i.e., similar/different features obtain similar/different representations. The main difference among various SR-based features lies in the constraint term. A widely used constraint function is ℓ_1 norm constraint $\varphi(\boldsymbol{\alpha}) = \sum_{i=1}^{M}|\boldsymbol{\alpha}(i)|$. With the ℓ_1 norm constraint, SR achieves the effect that similar signals share a part of dictionary.

In addition to ℓ_1 norm constraint, there are many other SR-based feature extraction methods in the recent literature, such as discriminative sparse coding (Han et al., 2014), joint sparse coding (Chen et al., 2011b), SR-based Hough voting (Yokoya and Iwasaki, 2015), nonlinear SR (Zhang et al., 2015d), sparse transfer manifold embedding (Zhang et al., 2014a), Laplacian SR (Chen et al., 2011c), and class-specific and discriminative SR (Du and Zhang, 2014). All of them extend SR by substituting or modifying the constraint term. Due to limited space, we do not introduce them one by one here.

6.1.5. Haar-like features

Haar-like features were first introduced by Viola and Jones (2001) for face detection and now are widely used for object detec-

tion in RSIs such as (Grabner et al., 2008; Leitloff et al., 2010). Haar-like features can be defined by a set of rectangular regions which are either positive or negative weighted. The feature value is calculated as the sum of pixel values in rectangular regions with their corresponding weights. Haar-like feature can be mathematically formed as $\{(Rec_i, w_i)|i = 1, \ldots, N\}$, where N is the number of the rectangles, Rec_i denotes the i-th rectangular region, and w_i denotes the assigned weight. The feature value is computed as $Haar(Rec_i, w_i) = \sum_{i=1}^{N} w_i \cdot Sum(Rec_i)$, where $Sum(Rec_i)$ refers to the summed pixel values inside rectangle Rec_i. Besides, a normalization is required to reduce the effects of illumination changes. Suppose μ and σ are the mean value and standard derivation, respectively, of pixel values inside local regions, then the normalized feature value is denoted as $(Haar(Rec_i, w_i) - \mu)/\sigma$.

6.2. Feature fusion and dimension reduction

After feature extraction, some optional techniques can be applied to further improve feature representation performance. We now briefly discuss two kinds of commonly used techniques, namely feature fusion and dimension reduction.

6.2.1. Feature fusion

Although many visual descriptors have been proposed to describe geospatial objects, as reviewed in Section 6.1, different features generally describe different aspects of the visual characteristics. Due to the diversity of object appearances, the fusion of heterogeneous visual features is somewhat necessary for high-performance object detection. Feature fusion motivates the integration of the information of multiple visual cues in order to generate more comprehensive feature representation. The fusion of complementary features can lead to significant improvement in object detection.

A simple and widely used feature fusion technique is known as linear vector concatenation. Briefly, suppose we have N different kinds of feature vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_N\}$, where each feature vector $\boldsymbol{v}_i$ represents a specific visual property, vector concatenation-based feature fusion is to connect all feature vectors into a new long-length feature vector $\boldsymbol{v} = [\boldsymbol{v}_1; \boldsymbol{v}_2; \ldots; \boldsymbol{v}_N]$. In some cases, normalization is required to avoid the bias caused by scale differences. For instance, in (Grabner et al., 2008), an on-line boosting algorithm is used to integrate and select features from Haar-like features, orientation histograms, and LBP features for comprehensive object representation. In (Zhong and Wang, 2007), the authors presented a multiple CRFs ensemble model to incorporate five types of texture features, namely gray-level occurrence feature, GLCM feature, Gabor feature, gradient orientation feature, and straight-line length feature, for urban area detection. Furthermore, some non-linear feature fusion methods were also proposed such as heterogeneous feature machines (HFM) (Cao et al., 2009) and sparse multimodal learning (SMML) approach (Wang et al., 2013).

6.2.2. Dimension reduction

Dimension reduction aims to generate compact low-dimensional features from high-dimensional ones, in order to reduce computational complexity. Compactness and distinctiveness are two important properties of dimension reduction. Dimension reduction techniques can be broadly classified into unsupervised and supervised methods. PCA is a classical linear unsupervised method, which aims to find and preserve projections with maximal variance for energy preservation. However, the generalization capability of unsupervised dimension reduction is limited because they are dataset-dependent. For the purpose of classification, supervised dimension reduction methods hold an advantage over unsupervised methods due to their use of the class

information within the training data set, and hence have attracted more attention such as partial least squares (Kembhavi et al., 2011), Fisher discriminant analysis (Sugiyama, 2007), and linear discriminant analysis (LDA) (Hariharan et al., 2012). In (Sugiyama, 2007), interclass separation is emphasized by replacing the total covariance matrix in PCA by Fisher criterion. In (Hariharan et al., 2012), LDA is used to remove the correlations of HOG and obtain the Whitened Histograms of Orientation (WHO) features.

6.3. Classifier training

After feature extraction, feature fusion and dimension reduction, a classifier can be trained using a number of possible approaches with the objective of minimizing the misclassification error on the training dataset. In practice, many different learning approaches can be selected including, but are not limited to, support vector machine (SVM), AdaBoost, k-nearest-neighbor (kNN), conditional random field (CRF), sparse representation-based classification (SRC), and artificial neural network (ANN).

6.3.1. Support vector machine (SVM)

SVM is one of the most popular and effective machine learning algorithms for solving classification problems, which was proposed by Vapnik and Vapnik (1998) and now has been widely used for various object detection applications, such as man-made objects recognition (Inglada, 2007), road extraction (Das et al., 2011; Huang and Zhang, 2009; Song and Civco, 2004), change detection (Bovolo et al., 2008; De Morsier et al., 2013), multi-class object detection (Bai et al., 2014; Cheng et al., 2013b; Cheng et al., 2014a, 2014b; Li and Itti, 2011), ship detection (Bi et al., 2012; Zhu et al., 2010), airport detection (Tao et al., 2011), airplane detection (Sun et al., 2012; Zhang et al., 2014b, 2015b), weakly supervised object detection (Han et al., 2015; Zhang et al., 2015a). A recent review about the applications of SVMs in remote sensing can also be found in (Mountrakis et al., 2011).

In its simplest form, SVM is a linear binary classifier. For multiple classification problem, such as multi-class object detection tasks, adjustments are only made to the simple SVM binary classifier to operate as a multi-class classifier by using methods such as one-against-all and one-against-others. Furthermore, SVM can also be used as a non-linear classifier (called kernel SVM) to further improve the separation between classes by projecting the samples onto a higher dimensional feature space.

Taking a binary classification problem as an example, given a training set of N samples $(X, Y) = \{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_N, y_N)|\boldsymbol{x}_i \in R^d, i = 1, \ldots, N\}$, where $\boldsymbol{x}_i$ is the feature vector corresponding to the i-th sample labeled by $y_i \in \{-1, +1\}$. For the linear and separable case, the classifier is trained by solving the following optimization problem

$$\min \frac{1}{2}\|\boldsymbol{w}\|^2 \quad \text{s.t. } y_i(\langle \boldsymbol{x}_i, \boldsymbol{w} \rangle + b) \geqslant 1, \quad i = 1, \ldots, N \qquad (5)$$

where $\boldsymbol{w} \in R^d$ and $b \in R$ are to-be-learned classifier parameters, and $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ denotes the inner product of $\boldsymbol{x}$ and $\boldsymbol{y}$. In order to deal with non-separable data, the soft-margin hyperplane is used. A set of slack variables ξ_i is introduced to allow errors or data points inside the margin and a hyperparameter C is used to tune the trade-off between the amount of accepted errors and the maximization of the margin. The optimization problem now becomes

$$\min \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N}\xi_i \quad \text{s.t. } y_i(\langle \boldsymbol{x}_i, \boldsymbol{w} \rangle + b) \geqslant 1 - \xi_i, \quad i = 1, \ldots, N, \quad \xi_i \geqslant 0$$

$$(6)$$

Given a test sample $\boldsymbol{x}' \in R^d$, the SVM then classifies the test sample with the following decision function, which determines on which side of the separating hyperplane the sample $\boldsymbol{x}'$ lies.

$$f(\mathbf{x}') = \text{sign}(\langle \mathbf{x}', \mathbf{w} \rangle + b) \tag{7}$$

For non-linear classification problems, the data are first mapped into a higher dimensional feature space S by $\mathbf{x} \in R^d \mapsto \Phi(\mathbf{x}) \in S$ in which a separating hyperplane is built. This leads to the decision function becoming $f(\mathbf{x}') = \text{sign}(\langle \Phi(\mathbf{x}'), \mathbf{w} \rangle + b)$, where $\mathbf{w}$ is now a new vector of S. However, non-linear SVM is sensitive to the selection of kernel. In the case when the appropriate kernel is selected with optimal parameters, SVM can give very good classification results.

6.3.2. AdaBoost

The AdaBoost algorithm (Freund, 1995; Freund and Schapire, 1996) is a widely used machine learning algorithm that combines many weak classifiers to form a strong classifier by adjusting the weights of training samples. In many object detection applications such as vehicle detection (Grabner et al., 2008; Leitloff et al., 2010; Tuermer et al., 2013), ship detection (Shi et al., 2014), and airport detection (Aytekin et al., 2013), the AdaBoost classifier has shown good performance.

Various variants of AdaBoosting have been developed, such as Discrete AdaBoost (Freund and Schapire, 1997), Real AdaBoost (Schapire and Singer, 1999), and Gentle AdaBoost (Lienhart et al., 2003). It has been shown in (Lienhart et al., 2003) that Gentle Ada-Boost outperforms both Discrete AdaBoost and Real AdaBoost due to its less sensitivity to errors in the training set. This property is very useful when working with RSIs since the acquisition of accurate annotation of training data in complicated scenes cannot always be guaranteed. Consequently, we here focus on a brief review of the Gentle AdaBoost algorithm. The basic algorithm works as follows: Given a training set $(X, Y) = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N) | \mathbf{x}_i \in R^d, y_i \in \{-1, +1\}, i = 1, \ldots, N\}$, where $\mathbf{x}_i$ is the feature vector of i-th example and y_i is its class label. The weights for all examples are initialized uniformly by $w_i = 1/N$. Based on the training set, a weak classifier $f_m(\mathbf{x})$ is trained by weighted least-squares fitting of the labels Y to the feature vectors X with weights $\mathbf{w}$. Then, the weights are updated by $w_i \leftarrow w_i \cdot \exp(-y_i f_m(\mathbf{x}_i))/Z$ with the normalization factor Z defined by $Z = \sum_{i=1}^{N} w_i$. At each boosting iteration a new weak classifier is added and the process is repeated until a certain stopping condition is met (e.g. a predefined number of M weak classifiers are trained). Finally, a strong classifier $F(\mathbf{x})$ is computed as a linear combination of all weak classifiers

$$F(\mathbf{x}) = \sum_{m=1}^{M} f_m(\mathbf{x}) \tag{8}$$

In object detection, the sign of $F(\mathbf{x})$ decides the predicted object label, while its absolute value is a confidence measure for this prediction. The higher the absolute value is, the more confident is the result.

6.3.3. k-nearest-neighbor (kNN)

The kNN classifier is firstly proposed by Cover and Hart (1967). It is one of the most simple and traditional classification tools and has been extensively used for geospatial object detection and image classification (Blanzieri and Melgani, 2008; Cheng et al., 2013a; Haapanen et al., 2004; Ma et al., 2010; Yang et al., 2010; Zhu and Basir, 2005). It performs the classification by firstly consulting a reference set of labeled training samples. Then, given a test sample, we identify a subset of k training samples that are the closest to it. Finally, various decision strategies are adopted to classify the test sample and the most widely used strategy is assigning it to the class that appears most frequently within this subset.

Algorithmically, let $(X, Y) = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N) | \mathbf{x}_i \in R^d, y_i \in \{1, 2, \ldots, C\}, i = 1, \ldots, N\}$ be the training set with N training samples, where $\mathbf{x}_i$ is the feature vector of the i-th sample and y_i

is its class label (assuming we have C classes). Given a test sample $\mathbf{x}' \in R^d$ and an ordering function $\mathbf{f}_{\mathbf{x}'}$ (e.g. an Euclidean metrics $\mathbf{f}_{\mathbf{x}'}(\mathbf{x}) = \|\mathbf{x} - \mathbf{x}'\|$), we can order the set of training samples X to obtain its k nearest neighbors $X' = \{\mathbf{x}'_1, \mathbf{x}'_2, \ldots, \mathbf{x}'_k\}$ with their corresponding labels denoted as $Y' = \{y'_1, y'_2, \ldots, y'_k\}$. The kNN then classifies the test sample $\mathbf{x}'$ with the following majority voting rule

$$j^* = \arg\max_{j=1,\ldots,C} \sum_{i=1}^{k} \delta(y'_i, j) \tag{9}$$

where δ is the Kronecker delta defined as

$$\delta(i, j) = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases} \tag{10}$$

As can be seen from the above description, on the one hand, there are some attractive properties of the kNN classifier, such as the simplicity of the learning rule and the tuning of only one free parameter. On the other hand, the selection of the neighbor size k is a challenging problem because different k values will result in different performances. If k is considerably huge, the neighbors searching will take large time whereas a small value of k may decrease the prediction accuracy. To deal with this issue, different variants have been proposed in the literature (Blanzieri and Melgani, 2008; Dudani, 1976; Zhu and Basir, 2005).

6.3.4. Conditional random field (CRF)

CRF is introduced in (Lafferty et al., 2001) for labeling 1-D data. Kumar and Hebert (2003) further extended CRF to classify 2-D images. CRF has the advantage of incorporating the spatial contextual information to assign probabilities to the final label instead of only providing crisp decisions and hence has been applied successfully to various visual recognition tasks such as building detection (Li et al., 2015a; Wegner et al., 2011a, 2011b), urban area detection (Zhong and Wang, 2007), and airport detection (Yao et al., 2015).

Here we briefly introduce CRF models concerning with binary classification task of distinguishing objects from background, such as the works in (Li et al., 2015a; Wegner et al., 2011a, 2011b; Yao et al., 2015; Zhong and Wang, 2007). Given an input image, let $G = (S, E)$ be a graph denoting a CRF of $(\mathbf{x}, \mathbf{y})$, where $\mathbf{x} = \{\mathbf{x}_i\}_{i \in S}$ is the observed data from the input image with $\mathbf{x}_i$ denoting the data from the i-th region, $\mathbf{y} = \{y_i | y_i \in \{-1, +1\}\}_{i \in S}$ is the corresponding labels at the image regions, S is a set of nodes and each node corresponds to an image region data $\mathbf{x}_i$ and E is a set of edges connecting certain pairs of neighboring nodes. In CRF framework, the random variables $\mathbf{x}$ and $\mathbf{y}$ are jointly distributed in a discriminative manner and the posterior distribution $P(\mathbf{y}|\mathbf{x})$ of the labels $\mathbf{y}$ given the observations $\mathbf{x}$ can be written as

$$P(\mathbf{y}|\mathbf{x}) = \frac{1}{Z} \exp\left(\sum_{i \in S} A_i(\mathbf{x}, y_i) + \sum_{i \in S} \sum_{j \in N_i} I_{ij}(\mathbf{x}, y_i, y_j) \right) \tag{11}$$

where Z is a normalizing constant known as the partition function, $A_i(\mathbf{x}, y_i)$ is the association potential, $I_{ij}(\mathbf{x}, y_i, y_j)$ is the interaction potential, and N_i is the neighbors of the node i in G.

The association potential $A_i(\mathbf{x}, y_i)$ measures how likely a node i is labeled y_i with given the data $\mathbf{x}$, as in (Wegner et al., 2011a, 2011b) which is generally defined as

$$A_i(\mathbf{x}, y_i) = \exp\left(y_i \mathbf{w}^T \mathbf{h}_i(\mathbf{x}) \right) \tag{12}$$

where $\mathbf{h}_i(\mathbf{x})$ denotes the feature representation vector of node i and $\mathbf{w}^T$ is a weight vector containing the weights of the features in $\mathbf{h}_i(\mathbf{x})$ that are learned during the training process.

The interaction potential $I_{ij}(\mathbf{x}, y_i, y_j)$ describes how two nodes i and j interact, as in (Wegner et al., 2011a, 2011b) which is defined as

$$I_{ij}(\boldsymbol{x}, y_i, y_j) = \exp\left(y_i y_j \boldsymbol{v}^\mathrm{T} \boldsymbol{\mu}_{ij}(\boldsymbol{x})\right) \tag{13}$$

where $\boldsymbol{\mu}_{ij}(\boldsymbol{x})$ denotes a feature vector which is simply calculated by either subtracting or concatenating $\boldsymbol{h}_i(\boldsymbol{x})$ and $\boldsymbol{h}_j(\boldsymbol{x})$, and vector $\boldsymbol{v}^\mathrm{T}$ contains the weights of the features $\boldsymbol{\mu}_{ij}(\boldsymbol{x})$ that are adjusted during the training process. It should be noted that the interaction potential $I_{ij}(\boldsymbol{x}, y_i, y_j)$ is not only a function of adjacent labels y_i and y_j in the local neighborhood, but of all data, too. This means that the interaction potential can effectively incorporate both local and global context information which is actually a major advantage concerning automatic analysis of remote sensing data.

6.3.5. Sparse representation-based classification (SRC)

SRC was first proposed by Wright to tackle face recognition problem (Wright et al., 2009), which aims at looking for the sparsest representation of a test example in a dictionary composed of all training data across all classes. In the last few years, SRC and its variants have been widely used in RSI analysis field, including geospatial object detection (Chen et al., 2011b, 2011c, 2014c; Han et al., 2014; Liu and Shi, 2014; Zhang et al., 2015c, 2015d) and hyperspectral image classification (Chen et al., 2011a; Qian et al., 2013; Tao et al., 2012), and have shown excellent performances.

A brief review of SRC is as follows. Assume we have C classes and we are given m_i training data $\left\{\boldsymbol{a}_{ij} \in R^d\right\}_{j=1}^{m_i}$ for each class $i (i = 1, \ldots, C)$. We denote by $\mathbf{A}_i = [\boldsymbol{a}_{i1}, \ldots, \boldsymbol{a}_{im_i}] \in R^{d \times m_i}$ the collection of training data of the i-th class and denote by $\mathbf{A} = [\mathbf{A}_1, \ldots, \mathbf{A}_C]$ the collection of all training data across all classes. Given a test example $\boldsymbol{x} \in R^d$, which belongs to one of the C classes, our goal is to find the class to which the test example belongs. The SRC method is based on the idea that a test example has a sparse representation in the dictionary of all the training data across different classes. Therefore, in principle, we can sparsely code $\boldsymbol{x}$ on $\mathbf{A}$ via solving the following optimization problem

$$\boldsymbol{\alpha} = \arg\min_{\boldsymbol{\alpha}}\{\|\boldsymbol{x} - \mathbf{A}\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1\} \tag{14}$$

We can then determine the class of a given test example as the class that best represents the test example using its training data. More precisely, for a given test example $\boldsymbol{x}$, we define the i-th residual, i.e., error between the test example and the reconstruction from training samples in the i-th class, to be

$$r_i(\boldsymbol{x}) = \|\boldsymbol{x} - \mathbf{A}_i\hat{\boldsymbol{\alpha}}_i\|_2, \quad i = 1, \ldots, C \tag{15}$$

where $\hat{\boldsymbol{\alpha}} = [\hat{\boldsymbol{\alpha}}_1; \ldots; \hat{\boldsymbol{\alpha}}_C]$ is the optimal solution of (14) and $\hat{\boldsymbol{\alpha}}_i$ is the coefficient vector associated with the i-th class. The class of $\boldsymbol{x}$ is determined as the one with the minimal residual

$$\mathrm{class}(\boldsymbol{x}) = \arg\min_{i=1,\ldots,C} r_i(\boldsymbol{x}) \tag{16}$$

6.3.6. Artificial neural network (ANN)

ANN, known as a kind of pattern classifiers, was proposed in the early 1980s. It has the capability to learn patterns whose complexity makes them difficult to analyze using other conventional approaches. Some typical neural network architectures, such as multilayer perceptron (MLP) (Li et al., 2001; Mokhtarzade and Zoej, 2007; Pacifici et al., 2009), hopfield neural networks (Ghosh et al., 2007), extreme learning machine (ELM) (Malek et al., 2014; Tang et al., 2015), and convolutional neural network (CNN) (Jin and Davis, 2007; Wang et al., 2015), have been successfully used in many remote sensing applications including ship detection (Tang et al., 2015), vehicle detection (Jin and Davis, 2007), road detection (Mokhtarzade and Zoej, 2007; Wang et al., 2015), tree detection (Malek et al., 2014), fire smoke detection (Li et al.,

2001), change detection (Ghosh et al., 2007), and land-use classification (Pacifici et al., 2009). Here, we briefly review the structure of neural networks. For more details and the latest advances, readers can refer to (Bishop, 1995; LeCun et al., 2015).

Neural networks are made up of a large number of simple processing units called nodes or neurons. The main task of a neuron is to receive input from its neighbors, to compute an output and to send the output to its neighbors. Neurons are usually organized into layers with full or random connections between successive layers. An ANN generally consists of three types of layers, namely input layer, hidden layer, and output layer, that receive, process and present the final results, respectively. There are two main stages in the operation of an ANN classifier, i.e., learning (training) and recalling. Learning is the process of adapting or modifying the connection weights so that the network can fulfill a specific task. This process is mainly performed with a supervised learning algorithm using a training set, in which random weights are first given at the beginning of training, and then the algorithm performs weights tuning by minimizing the error of misclassification. Recalling asks how the network can operate based on what it has learned in the training stage. It actually uses the trained network for interpolation and extrapolation, such as classification and regression.

6.3.7. Other classifiers

Besides the aforementioned classifiers, some other machine learning algorithms for object detection in optical RSIs are also used, such as Gaussian mixture model (GMM) (Ari and Aksoy, 2014; Bhagavathy and Manjunath, 2006), Markov random fields (MRF) (Benedek et al., 2015; Kasetkasem and Varshney, 2002; Le Hégarat-Mascle and André, 2009), random forest (Dong et al., 2015; Yin et al., 2015), texton forests (Lei et al., 2012, 2014), support tensor machine (STM) (Zhang et al., 2011b), Bayesian minimum risk classifier (Zhao and Nevatia, 2003), partial least squares classifier (Kembhavi et al., 2011), logistic regression classifier (Corbane et al., 2010; Zhang and Couloigner, 2006), LDA (Eikvil et al., 2009), quadratic discriminant analysis (QDA) (Eikvil et al., 2009), and decision fusion (Fauvel et al., 2006; Senaras et al., 2013). Readers can refer to the relevant publications for details.

A brief summary of machine learning-based object detection methods is given in Table 4.

7. Datasets and evaluation metrics

During the last decades, many efforts have been made to develop various methods for object detection from aerial and satellite images. In order to make the results of such algorithms more comparable, it is of paramount importance to introduce some publicly available benchmarking datasets and some standard evaluation metrics to help researchers conduct further studies and compare different algorithms. Here we summarize five publicly available datasets and three standard evaluation metrics.

7.1. Datasets

NWPU VHR-10 dataset (Cheng et al., 2014a).[1] This is a challenging 10-class geospatial object detection dataset, which can be used for both single class and multi-class objects detection. These ten classes of objects are airplane, ship, storage tank, baseball diamond, tennis court, basketball court, ground track field, harbor, bridge, and vehicle. This dataset contains totally 800 VHR optical RSIs, from which 757 airplanes, 302 ships, 655 storage tanks, 390 baseball diamonds, 524 tennis courts, 159 basketball courts, 163 ground track fields, 224 harbors, 124 bridges, and 477 vehicles were manually

---

[1] http://pan.baidu.com/s/1hqwzXeG.

Table 4
A brief summary of machine learning-based object detection methods.

Objects and publications	Strengths	Limitations
Airplane (Lei et al., 2012; Liu and Shi, 2014; Sun et al., 2012; Zhang et al., 2014b, 2015b); airport (Aytekin et al., 2013; Cheng et al., 2014c; Tao et al., 2011; Yao et al., 2015); building (Ari and Aksoy, 2014; Lei et al., 2012; Li et al., 2015a; Senaras et al., 2013; Wegner et al., 2011a, 2011b); change detection (Benedek et al., 2015; Bovolo et al., 2008; De Morsier et al., 2013; Ghosh et al., 2007; Kasetkasem and Varshney, 2002; Lei et al., 2014); forest (Haapanen et al., 2004); landslide (Cheng et al., 2013a); multi-class objects (e.g. airport, airplane, ship, storage tank, etc.) (Bai et al., 2014; Cheng et al., 2013b, 2014a, 2014b; Han et al., 2014, 2015; Li and Itti, 2011; Yokoya and Iwasaki, 2015; Zhang et al., 2015a; Zhou et al., 2015a); road (Das et al., 2011; Huang and Zhang, 2009; Inglada, 2007; Mokhtarzade and Zoej, 2007; Song and Civco, 2004; Wang et al., 2015; Zhang and Couloigner, 2006); ship (Bi et al., 2012; Corbane et al., 2010; Shi et al., 2014; Tang et al., 2015; Zhu et al., 2010); tree (Malek et al., 2014); land-use/land-cover (Bhagavathy and Manjunath, 2006; Fauvel et al., 2006; Pacifici et al., 2009; Zhong and Wang, 2007); vehicle (Eikvil et al., 2009; Grabner et al., 2008; Jin and Davis, 2007; Kembhavi et al., 2011; Leitloff et al., 2010; Tuermer et al., 2013; Zhao and Nevatia, 2003)	Object model (detector) can be automatically established via machine learning technique; Detection system is scalable and compatible; Have high detection accuracy	Need a lot of training samples of objects and non-objects to learn classifiers; Detection accuracy depends on the training samples

annotated with axis-aligned bounding boxes used for ground truth. 715 color images were acquired from Google Earth with the spatial resolution ranging from 0.5-m to 2-m, and 85 pansharpened color infrared (CIR) images were acquired from Vaihingen data with a spatial resolution of 0.08-m. The Vaihingen data was provided by the German Society for Photogrammetry, Remote Sensing and Geoinformation (DGPF) (Cramer, 2010): http://www.ifp.uni-stuttgart.de/dgpf/DKEPAllg.html.

SZTAKI-INRIA building detection dataset (Benedek et al., 2012).[2] This dataset is an excellent resource for benchmarking building extraction algorithms. It contains the rectangular footprints of 665 buildings in nine aerial or satellite images taken from Budapest and Szada (both in Hungary), Manchester (U.K.), Bodensee (Germany), and Normandy and Cot d'Azur (both in France), as well as manually annotated ground truth data. Two images (Budapest and Szada) are aerial images, and the remaining images are satellite images acquired from Google Earth, IKONOS, and QuickBird. All of the images contain only RGB information.

TAS aerial car detection dataset (Heitz and Koller, 2008).[3] This dataset consists of 30 images extracted from Google Earth and a total of 1319 manually labeled cars with axis-aligned bounding boxes used for ground truth. The images are color and of size 792×636. The average car window is approximately 45×45 pixels.

Overhead imagery research dataset (OIRDS) (Tanner et al., 2009).[4] The OIRDS is designed for vehicle detection algorithms. It is a large collection of about 900 overhead images, captured using aircraft-mounted cameras. The total number of vehicles annotated in the data set is around 1800. Ground sample distance (GSD) values for the OIRDS images range from a low of 0.0838-m to a high of 0.3048-m.

IITM road extraction dataset (Das et al., 2011).[5] This dataset consists of 200 multispectral satellite images of size 512×512 each, where 100 images are of developed countries and 100 images are of emerging countries. They were created from satellite images with 1-m spatial resolution from Wikimapia (http://www.wikimapia.org/). With the advice (used visual observation and geolocation) of a GIS

expert, this dataset was divided into four categories with 50 images per category. For each image in the dataset, a ground truth road map was also obtained using a human operator.

7.2. Evaluation metrics

Here, we explain three universally-agreed, standard, and easy-to-understand measures for evaluating the object detection methods, namely precision–recall curve, F-measure, and average precision. The first evaluation metric is based on the overlapping area between detections and ground truth. From this metric, we also report the F-measure, which jointly considers recall and precision, and average precision, which is the area under the precision–recall curve.

Precision–recall curve (PRC). The *Precision* measures the fraction of detections that are true positives and the *Recall* measures the fraction of positives that are correctly identified. Let *TP*, *FP*, and *FN* denote the number of true positives, the number of false positives, and the number of false negatives. The *Precision* and *Recall* can be formulated as

$$Precision = TP/(TP + FP) \quad (17)$$

$$Recall = TP/(TP + FN) \quad (18)$$

The PRC measure can evaluate the performance of an object detection algorithm at both pixel-level and object-level. For the pixel-level evaluation, true positives refer to the pixels assigned as target objects in both ground truth and detection result, false positives refer to the pixels assigned as target objects in detection result but not in ground truth, and false negatives refer to the pixels assigned as a target object in ground truth but not in detection result. In an object-level evaluation, a detection is labeled as true positive if the area overlap ratio a_o between the detection and ground truth object exceed a predefined threshold λ by the formula

$$a_o = \frac{area(detection \cap ground\_truth)}{area(detection \cup ground\_truth)} > \lambda \quad (19)$$

where $detection \cap ground\_truth$ denotes the intersection of the detection and ground truth object and $detection \cup ground\_truth$ denotes their union. Otherwise the detection is considered as a false

positive. In addition, if several detections overlap with a same ground truth object, only one is considered as true positive and the others are considered as false positives.

F-measure. Often, neither *Precision* nor *Recall* can comprehensively evaluate the quality of an object detection method. To this end, the F-measure is proposed to combine the *Precision* and *Recall* metrics to a single measure by weighted harmonic mean of them with a non-negative weight β^2

$$F_\beta = \frac{(1 + \beta^2)Precision \times Recall}{\beta^2 Precision + Recall} \tag{20}$$

Setting $\beta^2 < 1$ can raise more importance to the *Precision* value and vice versa. As suggested by many object detection works (Benedek et al., 2012; Han et al., 2014, 2015; Li et al., 2015a; Ok, 2013; Ok et al., 2013), β is often set to 1 to give each an equal importance, which results in $F_\beta = F_1$.

Average precision (AP). The AP computes the average value of *Precision* over the interval from *Recall* = 0 to *Recall* = 1, i.e., the area under the PRC, so the higher the AP value is, the better the performance and vice versa.

8. Promising research directions

Here we propose two promising research directions, namely deep learning-based feature representation and weakly supervised learning-based geospatial object detection, to further advance the development of object detection task.

8.1. Deep learning-based feature representation

Currently, most of geospatial object detection methods, except for few works such as (Han et al., 2015; Tang et al., 2015; Wang et al., 2015; Zhou et al., 2015b), are still dominated by handcrafted features and shallow learning-based features (e.g. SR-based features (Chen et al., 2011b, 2011c; Cheng et al., 2014c; Du and Zhang, 2014; Han et al., 2014; Liu and Shi, 2014; Yokoya and Iwasaki, 2015; Zhang et al., 2014a, 2015c, 2015d)). Although these features have shown impressive success for some specific object detection tasks, the involvement of human ingenuity in feature design or shallow structure significantly influence the representational power as well as the effectiveness for object detection. Especially as the visual recognition task becomes more challenging, the description capability of those features may become limited or even impoverished.

In 2006, a breakthrough in feature learning and deep learning was made by Hinton and Salakhutdinov (2006). Since then, deep learning has shown its stronger feature representation power in a wide range of computer vision applications. On the one hand, compared with handcrafted features which involves in abundant human ingenuity for features design, deep learning features are directly extracted from data in an automated fashion via neural networks of deep architecture. Thus, the burden for human engineering-based feature design has been transferred to the network construction. On the other hand, compared with traditional shallow learning techniques for feature representations (e.g. sparse coding), the deep architecture of neural networks can extract much more powerful feature representation. Besides, the feature maps from higher layers of the deep neural network display semantic abstracting properties. The representation power of deep learning feature can be significantly strengthened with stacked layers of neural networks.

Despite the progress made so far, there are also two notable challenges for deep learning-based feature representation. First, the deep learning features are much dependent on big image data because the deep neural networks need to be trained on a big enough training data set with abundant and diverse images to avoid overfitting. How to reduce the dependency to large-scale training data is a challenging issue. Second, the time cost for feature extraction with deep neural networks for online object detection is still very large. Many efforts are required to further improve the computational efficiency.

8.2. Weakly supervised learning-based geospatial object detection

With the explosive growth of satellite and aerial images in both quantity and quality, supervised learning-based object detection methods encounter a serious challenge caused by the fact that these approaches often require a large number of training samples with accurate manual annotation to train effective detectors. However, since VHR images often contain a greater range of object categories, big to an airport, small to a vehicle, manual annotation of various objects in large image sets is hardly achievable due to its expensive, time consuming and ineffective nature.

A few efforts (Bai et al., 2014; Capobianco et al., 2009; Cheng et al., 2014a; De Morsier et al., 2013; Liu et al., 2008) have been made to alleviate the work of human annotation. One common idea is to adopt the semi-supervised learning methods. Such methods apply a self-learning or an active learning technique where machine learning algorithms can automatically explore the most informative unlabeled examples based on a limited set of available labeled examples. Then, these selected unlabeled examples are combined with the initial labeled examples for the training of object detector or classifier. Although semi-supervised learning methods can considerably reduce the labor of human annotation, they still inevitably require a number of manually labeled training examples.

Weakly supervised learning (WSL) is more desirable for further reducing the human annotation significantly, in which the training set needs only binary labels indicating whether an image contains the target object or not. Although a few WSL approaches have been developed for natural scene image analysis (Deselaers et al., 2012; Zhu and Shao, 2014), those methods cannot be directly used for geospatial object detection because they are insufficient to address the challenges of RSIs, such as variations in the visual appearance of objects and complex background clutter. As an initial effort, our recent work in (Han et al., 2015; Zhang et al., 2015a; Zhou et al., 2015a, 2015b) have shown the feasibility of using WSL for geospatial object detection. However, two challenging problems still exist. The first problem is the existing WSL-based geospatial object detection methods only work for a specific object class. Generally, a large-scale RSI often contains multiple classes of interesting objects instead of only a single one. Therefore, it is of great significance to develop a WSL-based object detection framework that can simultaneously identify multiple classes of objects. The other problem is the performance of existing WSL-based methods is still far from satisfactory. It still needs many efforts to develop more effective methods to improve the detection accuracy.

9. Conclusion

Object detection in optical RSIs has always been a fundamental but challenging issue in the field of aerial and satellite image analysis. During the last decades, considerable efforts have been made to develop various methods for the detection of different types of objects. In this paper, a review of the recent progress in this field was presented. We broadly categorized the methods into four groups, namely template matching-based methods, knowledge-based methods, OBIA-based methods, and machine learning-based methods, and exhaustively reviewed them respectively. Five publicly available datasets and three standard evaluation metrics were also summarized. Besides, we also discussed the challenges of current studies and proposed two promising research directions

in future. This survey will be significantly beneficial for the researchers to better understand this research field.

Acknowledgements

References

Ahmadi, S., Zoej, M.V., Ebadi, H., Moghaddam, H.A., Mohammadzadeh, A., 2010. Automatic urban building boundary extraction from high resolution aerial images using an innovative model of active contours. Int. J. Appl. Earth Observ. Geoinform. 12, 150–157.

Akçay, H.G., Aksoy, S., 2010. Building detection using directional spatial constraints. In: Proc. IEEE Int. Geosci. Remote Sens. Sympos., pp. 1932–1935.

Albrecht, F., 2010. Uncertainty in image interpretation as reference for accuracy assessment in object-based image analysis. In: Proceedings of the Ninth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, pp. 13–16.

Andreopoulos, A., Tsotsos, J.K., 2013. 50 years of object recognition: directions forward. Comput. Vis. Image Understand. 117, 827–891.

Ardila, J.P., Bijker, W., Tolpekin, V.A., Stein, A., 2012. Context-sensitive extraction of tree crown objects in urban areas using VHR satellite images. Int. J. Appl. Earth Obs. Geoinf. 15, 57–69.

Ari, C., Aksoy, S., 2014. Detection of compound structures using a Gaussian mixture model with spectral and spatial constraints. IEEE Trans. Geosci. Remote Sens. 52, 6627–6638.

Aytekin, Ö., Erener, A., Ulusoy, İ., Düzgün, Ş., 2012. Unsupervised building detection in complex urban environments from multispectral satellite imagery. Int. J. Remote Sens. 33, 2152–2177.

Aytekin, Ö., Zöngür, U., Halici, U., 2013. Texture-based airport runway detection. IEEE Geosci. Remote Sens. Lett. 10, 471–475.

Baatz, M., Schäpe, A., 2000. Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. In: Strobl, J., Blaschke, T., Griesebner, G. (Eds.), Angewandte Geographische Informations-Verarbeitung XII. Wichmann Verlag, Heidelberg, pp. 12–23.

Bai, X., Zhang, H., Zhou, J., 2014. VHR object detection based on structural feature extraction and query expansion. IEEE Trans. Geosci. Remote Sens. 52, 6508–6520.

Baker, B.A., Warner, T.A., Conley, J.F., McNeil, B.E., 2013. Does spatial resolution matter? A multi-scale comparison of object-based and pixel-based methods for detecting change associated with gas well drilling operations. Int. J. Remote Sens. 34, 1633–1651.

Baltsavias, E., 2004. Object extraction and revision by image analysis using existing geodata and knowledge: current status and steps towards operational systems. ISPRS J. Photogramm. Remote Sens. 58, 129–151.

Barsi, A., Heipke, C., 2003. Artificial neural networks for the detection of road junctions in aerial images. Int. Arch. Photogramm. Remote Sens. Spatial Inform. Sci. 34, 113–118.

Barzohar, M., Coope, D.B., 1996. Automatic finding of main roads in aerial images by using geometric-stochastic models and estimation. IEEE Trans. Pattern Anal. Mach. Intell. 18, 707–721.

Benedek, C., Descombes, X., Zerubia, J., 2012. Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics. IEEE Trans. Pattern Anal. Mach. Intell. 34, 33–50.

Benedek, C., Shadaydeh, M., Kato, Z., Szirányi, T., Zerubia, J., 2015. Multilayer Markov Random Field models for change detection in optical remote sensing images. ISPRS J. Photogramm. Remote Sens. 107, 22–37.

Benz, U.C., Hofmann, P., Willhauck, G., Lingenfelder, I., Heynen, M., 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. ISPRS J. Photogramm. Remote Sens. 58, 239–258.

Bhagavathy, S., Manjunath, B.S., 2006. Modeling and detection of geospatial objects using texture motifs. IEEE Trans. Geosci. Remote Sens. 44, 3706–3715.

Bi, F., Zhu, B., Gao, L., Bian, M., 2012. A visual search inspired computational model for ship detection in optical satellite images. IEEE Geosci. Remote Sens. Lett. 9, 749–753.

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Oxford University Press.

Blanzieri, E., Melgani, F., 2008. Nearest neighbor classification of remote sensing images with the maximal margin principle. IEEE Trans. Geosci. Remote Sens. 46, 1804–1811.

Blaschke, T., 2003. Object-based contextual image classification built on image segmentation. In: Proc. IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data, pp. 113–119.

Blaschke, T., 2010. Object based image analysis for remote sensing. ISPRS J. Photogramm. Remote Sens. 65, 2–16.

Blaschke, T., Burnett, C., Pekkarinen, A., 2004. Image segmentation methods for object-based analysis and classification. In: Jong, S.M.D., Meer, F.D.V.d. (Eds.), Remote Sensing Image Analysis: Including the Spatial Domain. Springer Netherlands, pp. 211–236.

Blaschke, T., Hay, G.J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Feitosa, R.Q., van der Meer, F., van der Werff, H., van Coillie, F., 2014. Geographic object-based image analysis–towards a new paradigm. ISPRS J. Photogramm. Remote Sens. 87, 180–191.

Blaschke, T., Hay, G.J., Weng, Q., Resch, B., 2011. Collective sensing: integrating geospatial technologies to understand urban systems—an overview. Remote Sens. 3, 1743–1776.

Blaschke, T., Lang, S., Hay, G.J., 2008. Object-based image analysis: spatial concepts for knowledge-driven remote sensing applications. Springer, Heidelberg, Berlin, New York.

Bontemps, S., Bogaert, P., Titeux, N., Defourny, P., 2008. An object-based change detection method accounting for temporal dependences in time series with medium to coarse spatial resolution. Remote Sens. Environ. 112, 3181–3191.

Bovolo, F., Bruzzone, L., Marconcini, M., 2008. A novel approach to unsupervised change detection based on a semisupervised SVM and a similarity measure. IEEE Trans. Geosci. Remote Sens. 46, 2070–2082.

Cao, L., Luo, J., Liang, F., Huang, T.S., 2009. Heterogeneous feature machines for visual recognition. In: Proc. IEEE Int. Conf. Comput. Vis., pp. 1095–1102.

Capobianco, L., Garzelli, A., Camps-Valls, G., 2009. Target detection with semisupervised kernel orthogonal subspace projection. IEEE Trans. Geosci. Remote Sens. 47, 3822–3833.

Chaudhuri, D., Kushwaha, N., Samal, A., 2012. Semi-automated road detection from high resolution satellite images by directional morphological enhancement and segmentation techniques. IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens. 5, 1538–1544.

Chaudhuri, D., Samal, A., 2008. An automatic bridge detection technique for multispectral images. IEEE Trans. Geosci. Remote Sens. 46, 2720–2727.

Chen, G., Hay, G.J., 2012. Object-based change detection. Int. J. Remote Sens. 33, 4434–4457.

Chen, Y., Nasrabadi, N.M., Tran, T.D., 2011a. Hyperspectral image classification using dictionary-based sparse representation. IEEE Trans. Geosci. Remote Sens. 49, 3973–3985.

Chen, Y., Nasrabadi, N.M., Tran, T.D., 2011b. Simultaneous joint sparsity model for target detection in hyperspectral imagery. IEEE Geosci. Remote Sens. Lett. 8, 676–680.

Chen, Y., Nasrabadi, N.M., Tran, T.D., 2011c. Sparse representation for target detection in hyperspectral imagery. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 5, 629–640.

Chen, Y., Nasrabadi, N.M., Tran, T.D., 2013. Hyperspectral image classification via kernel sparse representation. IEEE Trans. Geosci. Remote Sens. 51, 217–231.

Cheng, G., Guo, L., Zhao, T., Han, J., Li, H., Fang, J., 2013a. Automatic landslide detection from remote-sensing imagery using a scene classification method based on BoVW and pLSA. Int. J. Remote Sens. 34, 45–59.

Cheng, G., Han, J., Guo, L., Liu, T., 2015a. Learning coarse-to-fine sparselets for efficient object detection and scene classification. In: Proc. IEEE Int. Conf. Comput. Vision Pattern Recognit., pp. 1173–1181.

Cheng, G., Han, J., Guo, L., Liu, Z., Bu, S., Ren, J., 2015b. Effective and efficient midlevel visual elements-oriented land-use classification using VHR remote sensing images. IEEE Trans. Geosci. Remote Sens. 53, 4238–4249.

Cheng, G., Han, J., Guo, L., Qian, X., Zhou, P., Yao, X., Hu, X., 2013b. Object detection in remote sensing imagery using a discriminatively trained mixture model. ISPRS J. Photogramm. Remote Sens. 85, 32–43.

Cheng, G., Han, J., Zhou, P., Guo, L., 2014a. Multi-class geospatial object detection and geographic image classification based on collection of part detectors. ISPRS J. Photogramm. Remote Sens. 98, 119–132.

Cheng, G., Han, J., Zhou, P., Guo, L., 2014b. Scalable multi-class geospatial object detection in high-spatial-resolution remote sensing images. In: Proc. IEEE Int. Geosci. Remote Sens. Sympos., pp. 2479–2482.

Cheng, G., Han, J., Zhou, P., Yao, X., Zhang, D., Guo, L., 2014c. Sparse coding based airport detection from medium resolution Landsat-7 satellite remote sensing images. In: Proc. Int. Workshop Earth Observ. Remote Sens. Appl., pp. 226–230.

Cheng, G., Zhou, P., Han, J., Guo, L., Han, J., 2015c. Auto-encoder-based shared mid-level visual dictionary learning for scene classification using very high resolution remote sensing images. IET Comput. Vision 9, 639–647.

Clinton, N., Holt, A., Scarborough, J., Yan, L., Gong, P., 2010. Accuracy assessment measures for object-based image segmentation goodness. Photogramm. Eng. Remote Sens. 76, 289–299.

Congalton, R.G., Green, K., 2009. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. Taylor and Francis, London.

Contreras, D., Blaschke, T., Tiede, D., Jilge, M., 2015. Monitoring recovery after earthquakes through the integration of remote sensing, GIS, and ground observations: the case of L'Aquila (Italy). Cartogr. Geogr. Inform. Sci. 43, 115–133.

Contreras, D., Blaschke, T., Tiede, D., Jilge, M., 2016. Monitoring recovery after earthquakes through the integration of remote sensing, GIS, and ground observations: the case of L'Aquila (Italy). Cartogr. Geogr. Inform. Sci. 43, 115–133.

Corbane, C., Najman, L., Pecoul, E., Demagistri, L., Petit, M., 2010. A complete processing chain for ship detection using optical satellite imagery. Int. J. Remote Sens. 31, 5837–5854.

Cover, T.M., Hart, P.E., 1967. Nearest neighbor pattern classification. IEEE Trans. Inf. Theory 13, 21–27.

Cramer, M., 2010. The DGPF test on digital airborne camera evaluation–overview and test design. Photogramm. Eng. Remote Sens. 2010, 73–82.

D'Oleire-Oltmanns, S., Marzolff, I., Tiede, D., Blaschke, T., 2014. Detection of gully-affected areas by applying object-based image analysis (OBIA) in the region of Taroudannt, Morocco. Remote Sens. 6, 8287–8309.

Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection. In: Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., pp. 886–893.

Das, S., Mirnalinee, T., Varghese, K., 2011. Use of salient features for the design of a multistage framework to extract roads from high-resolution multispectral satellite images. IEEE Trans. Geosci. Remote Sens. 49, 3906–3931.

De Morsier, F., Tuia, D., Borgeaud, M., Gass, V., Thiran, J.-P., 2013. Semi-supervised novelty detection using svm entire solution path. IEEE Trans. Geosci. Remote Sens. 51, 1939–1950.

De Pinho, C.M.D., Fonseca, L.M.G., Korting, T.S., De Almeida, C.M., Kux, H.J.H., 2012. Land-cover classification of an intra-urban environment using high-resolution images and object-based image analysis. Int. J. Remote Sens. 33, 5973–5995.

Deselaers, T., Alexe, B., Ferrari, V., 2012. Weakly supervised localization and learning with generic knowledge. Int. J. Comput. Vis. 100, 275–293.

Dissanska, M., Bernier, M., Payette, S., 2009. Object-based classification of very high resolution panchromatic images for evaluating recent change in the structure of patterned peatland. Can. J. Remote Sens. 35, 189–215.

Doleire-Oltmanns, S., Eisank, C., Dragut, L., Blaschke, T., 2013. An object-based workflow to extract landforms at multiple scales from two distinct data types. IEEE Geosci. Remote Sens. Lett. 10, 947–951.

Dong, Y., Du, B., Zhang, L., 2015. Target detection based on random forest metric learning. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 8, 1830–1838.

Doxani, G., Karantzalos, K., Tsakiri-Strati, M., 2012. Monitoring urban changes based on scale-space filtering and object-oriented classification. Int. J. Appl. Earth Obs. Geoinf. 15, 38–48.

Doxani, G., Siachalou, S., Tsakiri-Strati, M., 2008. An object-oriented approach to urban land cover change detection. Int. Arch. Photogramm. Remote Sens. Spat. Inform. Sci. 37, 1655–1660.

Drăguț, L., Blaschke, T., 2006. Automated classification of landform elements using object-based image analysis. Geomorphology 81, 330–344.

Drăguț, L., Csillik, O., Eisank, C., Tiede, D., 2014. Automated parameterisation for multi-scale image segmentation on multiple layers. ISPRS J. Photogramm. Remote Sens. 88, 119–127.

Drăguț, L., Eisank, C., 2012. Automated object-based classification of topography from SRTM data. Geomorphology 141, 21–33.

Drăguț, L., Tiede, D., Levick, S.R., 2010. ESP: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. Int. J. Geogr. Inform. Sci. 24, 859–871.

Du, B., Zhang, L., 2014. A discriminative metric learning based anomaly detection method. IEEE Trans. Geosci. Remote Sens. 52, 6844–6857.

Dudani, S.A., 1976. The distance-weighted k-nearest-neighbor rule. IEEE Trans. Syst., Man, Cybern., 325–327

Durieux, L., Lagabrielle, E., Nelson, A., 2008. A method for monitoring building construction in urban sprawl areas using object-based analysis of Spot 5 images and existing GIS data. ISPRS J. Photogramm. Remote Sens. 63, 399–408.

Duro, D.C., Franklin, S.E., Dubé, M.G., 2012. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. Remote Sens. Environ. 118, 259–272.

Eikvil, L., Aurdal, L., Koren, H., 2009. Classification-based vehicle detection in high-resolution satellite images. ISPRS J. Photogramm. Remote Sens. 64, 65–72.

Eisank, C., Drăguț, L., Blaschke, T., 2011. A generic procedure for semantics-oriented landform classification using object-based image analysis. Geomorphometry, 125–128.

Esch, T., Thiel, M., Bock, M., Roth, A., Dech, S., 2008. Improvement of image segmentation accuracy based on multiscale optimization procedure. IEEE Geosci. Remote Sens. Lett. 5, 463–467.

Fauvel, M., Chanussot, J., Benediktsson, J.A., 2006. Decision fusion for the classification of urban remote sensing images. IEEE Trans. Geosci. Remote Sens. 44, 2828–2838.

Feizizadeh, B., Tiede, D., Rezaei Moghaddam, M.H., Blaschke, T., 2014. Systematic evaluation of fuzzy operators for object-based landslide mapping. South-Eastern Eur. J. Earth Observation Geomatics 3, 219–222.

Fischler, M.A., Elschlager, R.A., 1973. The representation and matching of pictorial structures. EEE Trans. Comput., 67–92

Flanders, D., Hall-Beyer, M., Pereverzoff, J., 2003. Preliminary evaluation of eCognition object-based software for cut block delineation and feature extraction. Can. J. Remote Sens. 29, 441–452.

Freund, Y., 1995. Boosting a weak learning algorithm by majority. Inf. Comput. 121, 256–285.

Freund, Y., Schapire, R.E., 1996. Experiments with a new boosting algorithm. In: Proc. Int. Conf. Mach. Learn., pp. 148–156.

Freund, Y., Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. 55, 119–139.

Gao, Y., Mas, J.F., Kerle, N., Navarrete Pacheco, J.A., 2011. Optimal region growing segmentation and its effect on classification accuracy. Int. J. Remote Sens. 32, 3747–3763.

Ghosh, S., Bruzzone, L., Patra, S., Bovolo, F., Ghosh, A., 2007. A context-sensitive technique for unsupervised change detection based on Hopfield-type neural networks. IEEE Trans. Geosci. Remote Sens. 45, 778–789.

Goodin, D.G., Anibas, K.L., Bezymennyi, M., 2015. Mapping land cover and land use from object-based classification: an example from a complex agricultural landscape. Int. J. Remote Sens. 36, 4702–4723.

Grabner, H., Nguyen, T.T., Gruber, B., Bischof, H., 2008. On-line boosting-based car detection from aerial images. ISPRS J. Photogramm. Remote Sens. 63, 382–396.

Haala, N., Brenner, C., 1999. Extraction of buildings and trees in urban environments. ISPRS J. Photogramm. Remote Sens. 54, 130–137.

Haapanen, R., Ek, A.R., Bauer, M.E., Finley, A.O., 2004. Delineation of forest/nonforest land use classes using nearest neighbor methods. Remote Sens. Environ. 89, 265–271.

Han, J., Zhang, D., Cheng, G., Guo, L., Ren, J., 2015. Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning. IEEE Trans. Geosci. Remote Sens. 53, 3325–3337.

Han, J., Zhou, P., Zhang, D., Cheng, G., Guo, L., Liu, Z., Bu, S., Wu, J., 2014. Efficient, simultaneous detection of multi-class geospatial targets based on visual saliency modeling and discriminative learning of sparse coding. ISPRS J. Photogramm. Remote Sens. 89, 37–48.

Hariharan, B., Malik, J., Ramanan, D., 2012. Discriminative decorrelation for clustering and classification. In: Proc. Eur. Conf. Comput. Vis., pp. 459–472.

Hay, G.J., Blaschke, T., Marceau, D.J., Bouchard, A., 2003. A comparison of three image-object methods for the multiscale analysis of landscape structure. ISPRS J. Photogramm. Remote Sens. 57, 327–345.

Hay, G.J., Castilla, G., Wulder, M.A., Ruiz, J.R., 2005. An automated object-based approach to the multiscale image segmentation of forest scenes. Int. J. Appl. Earth Observ. Geoinform. 7, 339–359.

Heitz, G., Koller, D., 2008. Learning spatial context: using stuff to find things. In: Proc. Eur. Conf. Comput. Vis., pp. 30–43.

Hinton, G.E., Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks. Science 313, 504–507.

Hofmann, A.D., Maas, H.-G., Streilein, A., 2002. Knowledge-based building detection based on laser scanner data and topographic map information. Int. Arch. Photogramm. Remote Sens. Spat. Inform. Sci. 34, 169–174.

Hofmann, P., Blaschke, T., Strobl, J., 2011. Quantifying the robustness of fuzzy rule sets in object-based image analysis. Int. J. Remote Sens. 32, 7359–7381.

Hu, J., Razdan, A., Femiani, J.C., Cui, M., Wonka, P., 2007. Road network extraction and intersection detection from aerial images by tracking road footprints. IEEE Trans. Geosci. Remote Sens. 45, 4144–4157.

Huang, X., Zhang, L., 2009. Road centreline extraction from high-resolution imagery based on multiscale structural features and support vector machines. Int. J. Remote Sens. 30, 1977–1987.

Huang, Y., Wu, Z., Wang, L., Tan, T., 2014. Feature coding in image classification: a comprehensive study. IEEE Trans. Pattern Anal. Mach. Intell. 36, 493–506.

Huertas, A., Nevatia, R., 1988. Detecting buildings in aerial images. Comput. Vis. Graph. Image Process. 41, 131–152.

Hung, C., Bryson, M., Sukkarieh, S., 2012. Multi-class predictive template for tree crown detection. ISPRS J. Photogramm. Remote Sens. 68, 170–183.

Hussain, M., Chen, D., Cheng, A., Wei, H., Stanley, D., 2013. Change detection from remotely sensed images: from pixel-based to object-based approaches. ISPRS J. Photogramm. Remote Sens. 80, 91–106.

Im, J., Jensen, J.R., Tullis, J.A., 2008. Object-based change detection using correlation image analysis and image segmentation. Int. J. Remote Sens. 29, 399–423.

Inglada, J., 2007. Automatic recognition of man-made objects in high resolution optical remote sensing images by SVM classification of geometric image features. ISPRS J. Photogramm. Remote Sens. 62, 236–248.

Irvin, R.B., McKeown, D.M., 1989. Methods for exploiting the relationship between buildings and their shadows in aerial imagery. IEEE Trans. Syst., Man, Cybern. 19, 1564–1575.

Jain, A.K., Ratha, N.K., Lakshmanan, S., 1997. Object detection using Gabor filters. Pattern Recognit. 30, 295–309.

Jain, A.K., Zhong, Y., Dubuisson-Jolly, M.-P., 1998. Deformable template models: a review. Signal Process. 71, 109–129.

Janssen, L., Middelkoop, H., 1992. Knowledge-based crop classification of a Landsat Thematic Mapper image. Int. J. Remote Sens. 13, 2827–2837.

Jin, X., Davis, C.H., 2007. Vehicle detection from high-resolution satellite imagery using morphological shared-weight neural networks. Image Vis. Comput. 25, 1422–1431.

Jing, Y., An, J., Liu, Z., 2011. A novel edge detection algorithm based on global minimization active contour model for oil slick infrared aerial image. IEEE Trans. Geosci. Remote Sens. 49, 2005–2013.

Jungho, I., Quackenbush, L.J., Li, M., Fang, F., 2014. Optimum scale in object-based image analysis. In: Weng, Q. (Ed.), Scale Issues in Remote Sensing. John Wiley & Sons Inc, pp. 197–214.

Karantzalos, K., Paragios, N., 2009. Recognition-driven two-dimensional competing priors toward automatic and accurate building detection. IEEE Trans. Geosci. Remote Sens. 47, 133–144.

Kasetkasem, T., Varshney, P.K., 2002. An image change detection algorithm based on Markov random field models. IEEE Trans. Geosci. Remote Sens. 40, 1815–1823.

Kembhavi, A., Harwood, D., Davis, L.S., 2011. Vehicle detection using partial least squares. IEEE Trans. Pattern Anal. Mach. Intell. 33, 1250–1265.

Kim, M., Madden, M., Warner, T., 2008. Estimation of optimal image object size for the segmentation of forest stands with multispectral IKONOS imagery. In: Blaschke, T., Lang, S., Hay, G.J. (Eds.), Object-Based Image Analysis-Spatial Concepts for Knowledge Driven Remote Sensing Applications. Springer, Berlin, Heidelberg, pp. 291–307.

Kim, M., Warner, T.A., Madden, M., Atkinson, D.S., 2011. Multi-scale GEOBIA with very high spatial resolution digital aerial imagery: scale, texture and image objects. Int. J. Remote Sens. 32, 2825–2850.

Kim, T., Park, S.-R., Kim, M.-G., Jeong, S., Kim, K.-O., 2004. Tracking road centerlines from high resolution remote sensing images by least squares correlation matching. Photogramm. Eng. Remote Sens. 70, 1417–1422.

Kumar, S., Hebert, M., 2003. Discriminative random fields: a discriminative framework for contextual interaction in classification. In: Proc. IEEE Int. Conf. Comput. Vis., pp. 1150–1157.

Lafferty, J., McCallum, A., Pereira, F.C., 2001. Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: Proc. Int. Conf. Mach. Learn., pp. 282–289.

Laptev, I., Mayer, H., Lindeberg, T., Eckstein, W., Steger, C., Baumgartner, A., 2000. Automatic extraction of roads from aerial images based on scale space and snakes. Mach. Vis. Appl. 12, 23–31.

Lazebnik, S., Schmid, C., Ponce, J., 2006. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: Proc. IEEE Int. Conf. Comput. Vision Pattern Recognit., pp. 2169–2178.

Le Hégarat-Mascle, S., André, C., 2009. Use of Markov random fields for automatic cloud/shadow detection on high resolution optical images. ISPRS J. Photogramm. Remote Sens. 64, 351–366.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.

Lefèvre, S., Weber, J., Sheeren, D., 2007. Automatic building extraction in VHR images using advanced morphological operators. In: Proc. Urban Remote Sens. Joint Event, pp. 1–5.

Lei, Z., Fang, T., Huo, H., Li, D., 2012. Rotation-invariant object detection of remotely sensed images based on texton forest and Hough voting. IEEE Trans. Geosci. Remote Sens. 50, 1206–1217.

Lei, Z., Fang, T., Huo, H., Li, D., 2014. Bi-temporal texton forest for land cover transition detection on remotely sensed imagery. IEEE Trans. Geosci. Remote Sens. 52, 1227–1237.

Leitloff, J., Hinz, S., Stilla, U., 2010. Vehicle detection in very high resolution satellite images of city areas. IEEE Trans. Geosci. Remote Sens. 48, 2795–2806.

Leninisha, S., Vani, K., 2015. Water flow based geometric active deformable model for road network. ISPRS J. Photogramm. Remote Sens. 102, 140–147.

Leon, J., Woodroffe, C.D., 2011. Improving the synoptic mapping of coral reef geomorphology using object-based image analysis. Int. J. Geogr. Inform. Sci. 25, 949–969.

Lhomme, S., He, D.C., Weber, C., Morin, D., 2009. A new approach to building identification from very-high-spatial-resolution images. Int. J. Remote Sens. 30, 1341–1354.

Li, E., Femiani, J., Xu, S., Zhang, X., Wonka, P., 2015a. Robust rooftop extraction from visible band images using higher order CRF. IEEE Trans. Geosci. Remote Sens. 53, 4483–4495.

Li, F.F., Perona, P., 2005. A bayesian hierarchical model for learning natural scene categories. In: Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., pp. 524–531.

Li, X., Cheng, X., Chen, W., Chen, G., Liu, S., 2015b. Identification of forested landslides using LiDar data, object-based image analysis, and machine learning algorithms. Remote Sens. 7, 9705–9726.

Li, X., Myint, S.W., Zhang, Y., Galletti, C., Zhang, X., Ii, B.L.T., 2014. Object-based land-cover classification for metropolitan Phoenix, Arizona, using aerial photography. Int. J. Appl. Earth Obs. Geoinf. 33, 321–330.

Li, X., Shao, G., 2013. Object-based urban vegetation mapping with high-resolution aerial photography as a single data source. Int. J. Remote Sens. 34, 771–789.

Li, X., Zhang, S., Pan, X., Dale, P., Cropp, R., 2010. Straight road edge detection from high-resolution remote sensing images based on the ridgelet transform with the revised parallel-beam Radon transform. Int. J. Remote Sens. 31, 5041–5059.

Li, Y., Wang, S., Tian, Q., Ding, X., 2015c. Feature representation for statistical-learning-based object detection: a review. Pattern Recognit. 48, 3542–3559.

Li, Z., Itti, L., 2011. Saliency and gist features for target detection in satellite images. IEEE Trans. Image Process. 20, 2017–2029.

Li, Z., Khananian, A., Fraser, R.H., Cihlar, J., 2001. Automatic detection of fire smoke using artificial neural networks and threshold approaches applied to AVHRR imagery. IEEE Trans. Geosci. Remote Sens. 39, 1859–1870.

Lienhart, R., Kuranov, A., Pisarevsky, V., 2003. Empirical analysis of detection cascades of boosted classifiers for rapid object detection. In: Proc. DAGM Sympos., pp. 297–304.

Lin, C., Nevatia, R., 1998. Building detection and description from a single intensity image. Comput. Vis. Image Understand. 72, 101–121.

Lin, Y., He, H., Yin, Z., Chen, F., 2015. Rotation-invariant object detection in remote sensing images based on radial-gradient angle. IEEE Geosci. Remote Sens. Lett. 12, 746–750.

Liow, Y.-T., Pavlidis, T., 1990. Use of shadows for extracting buildings in aerial images. Comput. Vis. Graph. Image Process. 49, 242–277.

Lisita, A., Sano, E., Durieux, L., 2013. Identifying potential areas of Cannabis sativa plantations using object-based image analysis of SPOT-5 satellite data. Int. J. Remote Sens. 34, 5409–5428.

Liu, G., Sun, X., Fu, K., Wang, H., 2013a. Aircraft recognition in high-resolution satellite images using coarse-to-fine shape prior. IEEE Geosci. Remote Sens. Lett. 10, 573–577.

Liu, G., Sun, X., Fu, K., Wang, H., 2013b. Interactive geospatial object extraction in high resolution remote sensing images using shape-based global minimization active contour model. Pattern Recognit. Lett. 34, 1186–1195.

Liu, L., Shi, Z., 2014. Airplane detection based on rotation invariant and sparse coding in remote sensing images. Optik-Int. J. Light Electron Opt. 125, 5327–5333.

Liu, Q., Liao, X., Carin, L., 2008. Detection of unexploded ordnance via efficient semisupervised and active learning. IEEE Trans. Geosci. Remote Sens. 46, 2558–2567.

Lizarazo, I., 2014. Accuracy assessment of object-based image classification: another STEP. Int. J. Remote Sens. 35, 6135–6156.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 60, 91–110.

Ma, L., Crawford, M.M., Tian, J., 2010. Local manifold learning-based-nearest-neighbor for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 48, 4099–4109.

Macfaden, S.W., O'Neil-Dunne, J.P.M., Royar, A.R., Lu, J.W.T., Rundle, A.G., 2012. High-resolution tree canopy mapping for New York City using LIDAR and object-based image analysis. J. Appl. Remote Sens. 6, 1–23.

MacLean, M.G., Congalton, R.G., 2012. Map accuracy assessment issues when using an object-oriented approach. In: Proceedings of the American Society for Photogrammetry and Remote Sensing 2012 Annual Conference, pp. 1–5.

Maillard, P., Cavayas, F., 1989. Automatic map-guided extraction of roads from SPOT imagery for cartographic database updating. Int. J. Remote Sens. 10, 1775–1787.

Malek, S., Bazi, Y., Alajlan, N., AlHichri, H., Melgani, F., 2014. Efficient framework for palm tree detection in UAV images. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 7, 4692–4703.

Mallinis, G., Koutsias, N., Tsakiri-Strati, M., Karteris, M., 2008. Object-based classification using Quickbird imagery for delineating forest vegetation polygons in a Mediterranean test site. ISPRS J. Photogramm. Remote Sens. 63, 237–250.

Martha, T.R., Kerle, N., Jetten, V., van Westen, C.J., Kumar, K.V., 2010. Characterising spectral, spatial and morphometric properties of landslides for semi-automatic detection using object-oriented methods. Geomorphology 116, 24–36.

Martha, T.R., Kerle, N., Van Westen, C.J., Jetten, V., Kumar, K.V., 2011. Segment optimization and data-driven thresholding for knowledge-based landslide detection by object-based image analysis. IEEE Trans. Geosci. Remote Sens. 49, 4928–4943.

Martha, T.R., Kerle, N., Westen, C.J.V., Jetten, V., Kumar, K.V., 2012. Object-oriented analysis of multi-temporal panchromatic images for creation of historical landslide inventories. ISPRS J. Photogramm. Remote Sens. 67, 105–119.

Mayer, H., 1999. Automatic object extraction from aerial imagery—a survey focusing on buildings. Comput. Vis. Image Understand. 74, 138–149.

Mayer, H., Hinz, S., Bacher, U., Baltsavias, E., 2006. A test of automatic road extraction approaches. Int. Arch. Photogramm. Remote Sens. Spat. Inform. Sci. 36, 209–214.

McGlone, J.C., Shufelt, J., 1994. Projective and object space geometry for monocular building extraction. In: Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., pp. 54–61.

McKeown, D.M., Denlinger, J.L., 1988. Cooperative methods for road tracking in aerial imagery. In: Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., pp. 662–672.

Mena, J.B., 2003. State of the art on automatic road extraction for GIS update: a novel classification. Pattern Recognit. Lett. 24, 3037–3058.

Mikolajczyk, K., Schmid, C., 2001. Indexing based on scale invariant interest points. In: Proc. IEEE Int. Conf. Comput. Vis., pp. 525–531.

Ming, D., Li, J., Wang, J., Zhang, M., 2015. Scale parameter selection by spatial statistics for GeOBIA: using mean-shift based multi-scale segmentation as an example. ISPRS J. Photogramm. Remote Sens. 106, 28–41.

Mishra, N.B., Crews, K.A., 2014. Mapping vegetation morphology types in a dry savanna ecosystem: integrating hierarchical object-based image analysis with random forest. Int. J. Remote Sens. 35, 1175–1198.

Mokhtarzade, M., Zoej, M.V., 2007. Road detection from high-resolution satellite images using artificial neural networks. Int. J. Appl. Earth Observ. Geoinform. 9, 32–40.

Moon, H., Chellappa, R., Rosenfeld, A., 2002. Performance analysis of a simple vehicle detection algorithm. Image Vis. Comput. 20, 1–13.

Moskal, L.M., Styers, D.M., Halabisky, M., 2011. Monitoring urban tree cover using object-based image analysis and public domain remotely sensed data. Remote Sens. 3, 2243–2262.

Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: a review. ISPRS J. Photogramm. Remote Sens. 66, 247–259.

Moustakidis, S., Mallinis, G., Koutsias, N., Theocharis, J.B., Petridis, V., 2012. SVM-based fuzzy decision trees for classification of high spatial resolution remote sensing images. IEEE Trans. Geosci. Remote Sens. 50, 149–169.

Movaghati, S., Moghaddamjoo, A., Tavakoli, A., 2010. Road extraction from satellite images using particle filtering and extended Kalman filtering. IEEE Trans. Geosci. Remote Sens. 48, 2807–2817.

Myint, S.W., Gober, P., Brazel, A., Grossman-Clarke, S., Weng, Q., 2011. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. Remote Sens. Environ. 115, 1145–1161.

Nebiker, S., Lack, N., Deuber, M., 2014. Building change detection from historical aerial photographs using dense image matching and object-based image analysis. Remote Sens. 6, 8310–8336.

Niu, X., 2006. A semi-automatic framework for highway extraction and vehicle detection based on a geometric deformable model. ISPRS J. Photogramm. Remote Sens. 61, 170–186.

Ojala, T., Pietikäinen, M., Mäenpää, T., 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. 24, 971–987.

Ok, A.O., 2013. Automated detection of buildings from single VHR multispectral images using shadow information and graph cuts. ISPRS J. Photogramm. Remote Sens. 86, 21–40.

Ok, A.O., Senaras, C., Yuksel, B., 2013. Automated detection of arbitrarily shaped buildings in complex environments from monocular VHR optical satellite imagery. IEEE Trans. Geosci. Remote Sens. 51, 1701–1717.

Pacifici, F., Chini, M., Emery, W.J., 2009. A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification. Remote Sens. Environ. 113, 1276–1292.

Peng, J., Liu, Y., 2005. Model and context-driven building extraction in dense urban aerial images. Int. J. Remote Sens. 26, 1289–1307.

Peng, J., Zhang, D., Liu, Y., 2005. An improved snake model for building detection from urban aerial images. Pattern Recognit. Lett. 26, 587–595.

Phinn, S.R., Roelfsema, C.M., Mumby, P.J., 2012. Multi-scale, object-based image analysis for mapping geomorphic and ecological zones on coral reefs. Int. J. Remote Sens. 33, 3768–3797.

Qian, Y., Ye, M., Zhou, J., 2013. Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features. IEEE Trans. Geosci. Remote Sens. 51, 2276–2291.

Radoux, J., Bogaert, P., Fasbender, D., Defourny, P., 2011. Thematic accuracy assessment of geographic object-based image classification. Int. J. Geogr. Inform. Sci. 25, 895–911.

Schapire, R.E., Singer, Y., 1999. Improved boosting algorithms using confidence-rated predictions. Mach. Learn. 37, 297–336.

Senaras, C., Ozay, M., Yarman Vural, F.T., 2013. Building detection with decision fusion. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 6, 1295–1304.

Shi, Z., Yu, X., Jiang, Z., Li, B., 2014. Ship detection in high-resolution optical imagery based on anomaly detector and local shape feature. IEEE Trans. Geosci. Remote Sens. 52, 4511–4523.

Shufelt, J., 1999. Performance evaluation and analysis of monocular building extraction from aerial imagery. IEEE Trans. Pattern Anal. Mach. Intell. 21, 311–326.

Shufelt, J.A., 1996. Exploiting photogrammetric methods for building extraction in aerial images. Int. Arch. Photogramm. Remote Sens. 31, B6.

Sirmaçek, B., Ünsalan, C., 2009. Urban-area and building detection using SIFT keypoints and graph theory. IEEE Trans. Geosci. Remote Sens. 47, 1156–1167.

Sirmacek, B., Ünsalan, C., 2011. A probabilistic framework to detect buildings in aerial and satellite images. IEEE Trans. Geosci. Remote Sens. 49, 211–221.

Solberg, A.H.S., 1999. Contextual data fusion applied to forest map revision. IEEE Trans. Geosci. Remote Sens. 37, 1234–1243.

Song, M., Civco, D., 2004. Road extraction using SVM and image segmentation. Photogramm. Eng. Remote Sens. 70, 1365–1371.

Stankov, K., He, D.-C., 2013. Building detection in very high spatial resolution multispectral images using the hit-or-miss transform. IEEE Geosci. Remote Sens. Lett. 10, 86–90.

Stankov, K., He, D.-C., 2014. Detection of buildings in multispectral very high spatial resolution images using the percentage occupancy hit-or-miss transform. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 7, 4069–4080.

Stilla, U., Geibel, R., Jurkiewicz, K., 1997. Building reconstruction using different views and context knowledge. Int. Arch. Photogramm. Remote Sens. 32, 129–136.

Stumpf, A., Kerle, N., 2011. Object-oriented mapping of landslides using Random Forests. Remote Sens. Environ. 115, 2564–2577.

Sugiyama, M., 2007. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. J. Mach. Learn. Res. 8, 1027–1061.

Sun, H., Sun, X., Wang, H., Li, Y., Li, X., 2012. Automatic target detection in high-resolution remote sensing images using spatial sparse coding bag-of-words model. IEEE Geosci. Remote Sens. Lett. 9, 109–113.

Sun, X., Wang, H., Fu, K., 2010. Automatic detection of geospatial objects using taxonomic semantics. IEEE Geosci. Remote Sens. Lett. 7, 23–27.

Tang, J., Deng, C., Huang, G.-B., Zhao, B., 2015. Compressed-domain ship detection on spaceborne optical image using deep neural network and extreme learning machine. IEEE Trans. Geosci. Remote Sens. 53, 1174–1185.

Tanner, F., Colder, B., Pullen, C., Heagy, D., Eppolito, M., Carlan, V., Oertel, C., Sallee, P., 2009. Overhead imagery research data set – an annotated data library & tools to aid in the development of computer vision algorithms. In: Proc. IEEE Appl. Imagery Pattern Recognit. Workshop, pp. 1–8.

Tao, C., Tan, Y., Cai, H., Tian, J., 2011. Airport detection from large IKONOS images using clustered SIFT keypoints and region information. IEEE Geosci. Remote Sens. Lett. 8, 128–132.

Tao, L., Sun, F., Yang, S., 2012. A fast and robust sparse approach for hyperspectral data classification using a few labeled samples. IEEE Trans. Geosci. Remote Sens. 50, 2287–2302.

Tchoku, C., Karnieli, A., Meisels, A., Chorowicz, J., 1996. Detection of drainage channel networks on digital satellite images. Int. J. Remote Sens. 17, 1659–1678.

Trinder, J.C., Wang, Y., 1998. Knowledge-based road interpretation in aerial images. Int. Arch. Photogramm. Remote Sens. 32, 635–640.

Tuermer, S., Kurz, F., Reinartz, P., Stilla, U., 2013. Airborne vehicle detection in dense urban areas using HoG features and disparity maps. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 6, 2327–2337.

Tzotsos, A., Karantzalos, K., Argialas, D., 2011. Object-based image analysis through nonlinear scale-space filtering. ISPRS J. Photogramm. Remote Sens. 66, 2–16.

Ünsalan, C., Sirmacek, B., 2012. Road network detection using probabilistic and graph theoretical methods. IEEE Trans. Geosci. Remote Sens. 50, 4441–4453.

Vapnik, V.N., Vapnik, V., 1998. Statistical Learning Theory. Wiley, New York.

Viola, P., Jones, M., 2001. Rapid object detection using a boosted cascade of simple features. In: Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., pp. 511–518.

Walker, J.S., Blaschke, T., 2008. Object-based land-cover classification for the Phoenix metropolitan area: optimization vs. transportability. Int. J. Remote Sens. 29, 2021–2040.

Walker, J.S., Briggs, J.M., 2007. An object-oriented approach to urban forest mapping in Phoenix. Photogramm. Eng. Remote Sens. 73, 577–583.

Walter, V., 2004. Object-based classification of remote sensing data for change detection. ISPRS J. Photogramm. Remote Sens. 58, 225–238.

Wang, F., 1993. A knowledge-based vision system for detecting land changes at urban fringes. IEEE Trans. Geosci. Remote Sens. 31, 136–145.

Wang, F., Newkirk, R., 1988. A knowledge-based system for highway network extraction. IEEE Trans. Geosci. Remote Sens. 26, 525–531.

Wang, H., Nie, F., Huang, H., Ding, C., 2013. Heterogeneous visual features fusion via sparse multimodal machine. In: Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., pp. 3097–3102.

Wang, J., Song, J., Chen, M., Yang, Z., 2015. Road network extraction: a neural-dynamic framework based on deep learning and a finite state machine. Int. J. Remote Sens. 36, 3144–3169.

Wang, M., Zhang, S., 2011. Road extraction from high-spatial-resolution remotely sensed imagery by combining multi-profile analysis and extended Snakes model. Int. J. Remote Sens. 32, 6349–6365.

Weber, J., Lefèvre, S., 2008. A multivariate hit-or-miss transform for conjoint spatial and spectral template matching. In: Proc. IEEE Int. Conf. Image Signal Process., pp. 226–235.

Weber, J., Lefèvre, S., 2012. Spatial and spectral morphological template matching. Image Vis. Comput. 30, 934–945.

Wegner, J.D., Hänsch, R., Thiele, A., Soergel, U., 2011a. Building detection from one orthophoto and high-resolution InSAR data using conditional random fields. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 4, 83–91.

Wegner, J.D., Soergel, U., Rosenhahn, B., 2011b. Segment-based building detection with conditional random fields. In: Proc. Joint Urb. Remote Sens. Event, pp. 205–208.

Weidner, U., Förstner, W., 1995. Towards automatic building extraction from high-resolution digital elevation models. ISPRS J. Photogramm. Remote Sens. 50, 38–49.

Wen, X., Shao, L., Fang, W., Xue, Y., 2015. Efficient feature selection and classification for vehicle detection. IEEE Trans. Circuits Syst. Video Technol. 25, 508–517.

Weng, Q., 2009. Remote Sensing and GIS Integration-Theories, Methods, and Applications. McGraw-Hill, New York.

Weng, Q., 2011. Advances in Environmental Remote Sensing: Sensor, Algorithms, and Applications. CRC Press, Boca Raton, FL, USA.

Woodcock, C.E., Strahler, A.H., 1987. The factor of scale in remote sensing. Remote Sens. Environ. 21, 311–332.

Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y., 2009. Robust face recognition via sparse representation. IEEE Trans. Pattern Anal. Mach. Intell. 31, 210–227.

Xie, Z., Roberts, C., Johnson, B., 2008. Object-based target search using remotely sensed data: a case study in detecting invasive exotic Australian Pine in south Florida. ISPRS J. Photogramm. Remote Sens. 63, 647–660.

Xu, C., Duan, H., 2010. Artificial bee colony (ABC) optimized edge potential function (EPF) approach to target recognition for low-altitude aircraft. Pattern Recognit. Lett. 31, 1759–1772.

Xu, S., Fang, T., Li, D., Wang, S., 2010. Object classification of aerial images with bag-of-visual words. IEEE Geosci. Remote Sens. Lett. 7, 366–370.

Yang, J.-M., Yu, P.-T., Kuo, B.-C., 2010. A nonparametric feature extraction and its application to nearest neighbor classification for hyperspectral image data. IEEE Trans. Geosci. Remote Sens. 48, 1279–1293.

Yang, L., Bi, G., Xing, M., Zhang, L., 2015. Airborne SAR moving target signatures and imagery based on LVD. IEEE Trans. Geosci. Remote Sens. 53, 5958–5971.

Yang, L., Xing, M., Wang, Y., Zhang, L., Bao, Z., 2013. Compensation for the NsRCM and phase error after polar format resampling for airborne spotlight SAR raw data of high resolution. IEEE Geosci. Remote Sens. Lett. 10, 165–169.

Yang, Y., Newsam, S., 2010. Bag-of-visual-words and spatial extensions for land-use classification. In: Proc. ACM SIGSPATIAL Int. Conf. Adv. Geogr. Inform. Syst., pp. 270–279.

Yang, Y., Newsam, S., 2011. Spatial pyramid co-occurrence for image classification. In: Proc. IEEE Int. Conf. Comput. Vis., pp. 1465–1472.

Yang, Y., Newsam, S., 2013. Geographic image retrieval using local invariant features. IEEE Trans. Geosci. Remote Sens. 51, 818–832.

Yao, X., Han, J., Guo, L., Bu, S., Liu, Z., 2015. A coarse-to-fine model for airport detection from remote sensing images using target-oriented visual saliency and CRF. Neurocomputing 164, 162–172.

Yin, J., Li, H., Jia, X., 2015. Crater detection based on gist features. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 8, 23–29.

Yokoya, N., Iwasaki, A., 2015. Object detection based on sparse representation and hough voting for optical remote sensing imagery. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 8, 2053–2062.

Yu, Q., Gong, P., Clinton, N., Biging, G., Kelly, M., Schirokauer, D., 2006. Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery. Photogramm. Eng. Remote Sens. 72, 799–811.

Zhan, Q., Molenaar, M., Tempfli, K., Shi, W., 2005. Quality assessment for geo-spatial objects derived from remotely sensed data. Int. J. Remote Sens. 26, 2953–2974.

Zhang, D., Han, J., Cheng, G., Liu, Z., Bu, S., Guo, L., 2015a. Weakly supervised learning for target detection in remote sensing images. IEEE Geosci. Remote Sens. Lett. 12, 701–705.

Zhang, J., Lin, X., Liu, Z., Shen, J., 2011a. Semi-automatic road tracking by template matching and distance transformation in urban areas. Int. J. Remote Sens. 32, 8331–8347.

Zhang, L., Zhang, L., Tao, D., Huang, X., 2011b. A multifeature tensor for remote-sensing target recognition. IEEE Geosci. Remote Sens. Lett. 8, 374–378.

Zhang, L., Zhang, L., Tao, D., Huang, X., 2014a. Sparse transfer manifold embedding for hyperspectral target detection. IEEE Trans. Geosci. Remote Sens. 52, 1030–1043.

Zhang, Q., Couloigner, I., 2006. Benefit of the angular texture signature for the separation of parking lots and roads on high resolution multi-spectral imagery. Pattern Recogit. Lett. 27, 937–946.

Zhang, W., Sun, X., Fu, K., Wang, C., Wang, H., 2014b. Object detection in high-resolution remote sensing images using rotation invariant parts based model. IEEE Geosci. Remote Sens. Lett. 11, 74–78.

Zhang, W., Sun, X., Wang, H., Fu, K., 2015b. A generic discriminative part-based model for geospatial object detection in optical remote sensing images. ISPRS J. Photogramm. Remote Sens. 99, 30–44.

Zhang, Y., Du, B., Zhang, L., 2015c. A sparse representation-based binary hypothesis model for target detection in hyperspectral images. IEEE Trans. Geosci. Remote Sens. 53, 1346–1354.

Zhang, Y., Zhang, L., Du, B., Wang, S., 2015d. A nonlinear sparse representation-based binary hypothesis model for hyperspectral target detection. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 8, 2513–2522.

Zhao, T., Nevatia, R., 2003. Car detection in low resolution aerial images. Image Vis. Comput. 21, 693–703.

Zhao, Y.-Q., Yang, J., 2015. Hyperspectral image denoising via sparse representation and low-rank constraint. IEEE Trans. Geosci. Remote Sens. 53, 296–308.

Zhen, Z., Quackenbush, L.J., Stehman, S.V., Zhang, L., 2013. Impact of training and validation sample selection on classification accuracy and accuracy assessment when using reference polygons in object-based classification. Int. J. Remote Sens. 34, 6914–6930.

Zheng, Z., Zhou, G., Wang, Y., Liu, Y., Li, X., Wang, X., Jiang, L., 2013. A novel vehicle detection method with high resolution highway aerial image. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 6, 2338–2343.

Zhong, P., Wang, R., 2007. A multiple conditional random fields ensemble model for urban area detection in remote sensing optical images. IEEE Trans. Geosci. Remote Sens. 45, 3978–3988.

Zhou, J., Bischof, W.F., Caelli, T., 2006. Road tracking in aerial images based on human–computer interaction and Bayesian filtering. ISPRS J. Photogramm. Remote Sens. 61, 108–124.

Zhou, P., Cheng, G., Liu, Z., Bu, S., Hu, X., 2015a. Weakly supervised target detection in remote sensing images based on transferred deep features and negative bootstrapping. Multidimension. Syst. Signal Process. http://dx.doi.org/10.1007/s11045-015-0370-3.

Zhou, P., Zhang, D., Cheng, G., Han, J., 2015b. Negative bootstrapping for weakly supervised target detection in remote sensing images. In: Proc. IEEE Int. Conf. Multimedia Big Data, pp. 318–323.

Zhou, W., 2013. An object-based approach for urban land cover classification: integrating LiDAR height and intensity data. IEEE Geosci. Remote Sens. Lett. 10, 928–931.

Zhou, W., Huang, G., Troy, A., Cadenasso, M.L., 2009. Object-based land cover classification of shaded areas in high spatial resolution imagery of urban areas: a comparison study. Remote Sens. Environ. 113, 1769–1777.

Zhou, W., Troy, A., 2008. An object-oriented approach for analysing and characterizing urban landscape at the parcel level. Int. J. Remote Sens. 29, 3119–3135.

Zhu, C., Shi, W., Pesaresi, M., Liu, L., Chen, X., King, B., 2005. The recognition of road network from high-resolution satellite remotely sensed data using image morphological characteristics. Int. J. Remote Sens. 26, 5493–5508.

Zhu, C., Zhou, H., Wang, R., Guo, J., 2010. A novel hierarchical method of ship detection from spaceborne optical image based on shape and texture features. IEEE Trans. Geosci. Remote Sens. 48, 3446–3456.

Zhu, F., Shao, L., 2014. Weakly-supervised cross-domain dictionary learning for visual recognition. Int. J. Comput. Vis. 109, 42–59.

Zhu, H., Basir, O., 2005. An adaptive fuzzy evidential nearest neighbor formulation for classifying remote sensing images. IEEE Trans. Geosci. Remote Sens. 43, 1874–1889.