Prior Knowledge-Based Automatic Object-Oriented Hierarchical Classification for Updating Detailed Land Cover Maps

Tianjun Wu · Jiancheng Luo · Liegang Xia · Zhanfeng Shen · Xiaodong Hu

Abstract Automatic information extraction from optical remote sensing images is still a challenge for large-scale remote sensing applications. For instance, artificial sample collection cannot achieve an automatic remote sensing imagery classification. Based on this, this paper resorts to the technologies of change detection and transfer learning, and further proposes a prior knowledge-based automatic hierarchical classification approach for detailed land cover updating. To establish this method, an automatic sample collection scheme for object-oriented classification is presented. Unchanged landmarks are first located. Prior knowledge of these categories from previously interpreted thematic maps is then transferred to the new target task. The knowledge is utilized to rebuild the relationship between landmark classes and their spatial-spectral features for land cover updating. A series of high-resolution remote sensing images are experimented for validating the effectiveness of the proposed approach in rapidly updating detailed land cover. The results show that, with the assistance of preliminary thematic maps, the approach can automatically obtain reliable object samples for object-oriented classification. Detailed land cover information can be excellently updated with a competitive accuracy, which demonstrates the practicability and effectiveness of our method. It creates a novel way for employing the technologies of knowledge discovery into the field of information extraction from optical remote sensing images.

T. Wu (✉) · J. Luo · L. Xia · Z. Shen · X. Hu
State Key Laboratory of Remote Sensing Sciences, Institute of
Remote Sensing and Digital Earth, Chinese Academy of Sciences,
Beijing 100101, China
e-mail: wutianjun1986@163.com

T. Wu · L. Xia
University of Chinese Academy of Sciences, Beijing 100049, China

Introduction

Land Use and Cover Change (LUCC) is the most prominent symbol of the Earth surface. It forcefully affects the ecological environment and climate change. As an important means to monitor global change, optical remote sensing can simultaneously observe the change on the Earth surface within a short period (Green et al. 1994). Remote sensing has played a huge role in large-scale monitoring environmental change and assessing the rational use of land resources. Many research projects of LUCC were completed based on MODIS, AVHR R, TM and other earth observing data (Franklin and Wulder 2002; Wardlow and Egbert 2003). The key issue of its applications is how to effectively translate observed data into valuable spatial and temporal information in a rapid way. Thus, land cover information extraction of large areas from optical images is one of the most fundamental problems for applying remote sensing in the field of land resource surveillance (Cihlar 2000).

With the development of computer science, especially artificial intelligence, remote sensing image classification has gradually shifted from the visual interpretation to computer-aided methods. The classification accuracies of the computer-aided methods are also improved after years of research (Lu and Weng 2007; Tso and Mather 2009; Richards and Jia 2012). However, it remains a challenge because of many factors, such as the complexity of the landscape, the uncertainty of imaging and the reliability of classification approach (Lu

and Weng 2007). Particularly, from the perspective of image understanding, the land cover classification methods previously reported in the literatures have not yet reached a high level of automation. Although there is an early exploration of automatic classification method, they placed more emphasis on theoretical formulations rather than practical applications. Current classification methods still require some human intervention and control (Lu and Weng 2007). For instance, in majority of published work by various researchers so far, most of the methods need manually mark the label of land cover types. Few methods achieve satisfactory classified results without human participant in selecting training samples. The process of manually training samples collection makes the automatic level of classification far from realistic requirements. Therefore, it is still challenging to quickly obtain land cover information from massive remote sensing data.

Under these situations, on the one hand, people need to further investigate the transmission mechanism of remote sensors, geological laws, imaging principles. On the other hand, theories and methods of artificial intelligence need to be further developed to increase the prompt application capabilities of remote sensing. Many limitations restricting the automation of classification are urgent to be improved in technical practice, especially the process of training sample collection. Rapid collection of effective training samples has been the bottleneck of the classification with large-scale remote sensing images (Demir et al. 2014).

Actually, large-scale landscapes (e.g. the landscape covered by 5–10 Landsat TM images) rarely change in a short time (e.g. two months or one season, which is related to the growth cycle of main vegetation), and the same land cover types always have same features in regional areas (e.g. the area with tens of square kilometers in size). This continuity in time and space can assume that most of the features of land cover types, such as the spectrum and spatiality, can be learned from prior knowledge, and be quantitatively expressed. Therefore, previously collected samples or produced thematic maps can provide amounts of prior knowledge to a new classification task. It is advantageous for automatically extracting information from optical remote sensing imageries by mining historical auxiliary data. Most researches have used these data as some types of classification features (Friedl et al. 2010; Homer et al. 2004), others are used as constraints for post-processing (Loveland et al. 2000). However, existing algorithms could not well consider how to effectively utilize this prior knowledge to find samples, and then improve the automation of classification. This limits the breadth and depth of optical remote sensing applications. Based on these, this paper arm to break the traditional bottleneck of manually collecting samples and presents a change detection and transfer learning-based automatic sample collection method. With this sample collection scheme, an automatic object-oriented hierarchical classification (AOHC) approach is further proposed for detailed land cover updating. The experiment results with high-resolution optical remote sensing images show the algorithm can automatically obtain reliable object samples for object-oriented classification. Performances of the method are demonstrated for land cover updating in an automatic manner.

Related Works and Background

Land Cover Classification

For accessing to land cover information, computer-aided interpretation of remote sensing images has been greatly developed in the joint effort of the experts in the fields of pattern recognition and remote sensing. Numerous methods were proposed in succession from the improvements of automation, accuracy and effectiveness (Lu and Weng 2007; Tso and Mather 2009; Richards and Jia 2012). Explicit research ideas can be roughly divided into two ways. One is considering the classification algorithms from the perspective of pattern recognition. That is, how to effectively apply the methods of pattern recognition, such as minimum distance method, maximum likelihood (ML) method (Luo et al. 2002), support vector machine (SVM) method (Foody and Mathur 2004), decision tree method (Pal and Mather 2003; Barros and Basgalupp 2012), random forest method (Gislason et al. 2006), artificial neural network method (Foody et al. 1995), and multi-classifier ensemble method (Briem et al. 2002), into the field of optical remote sensing image classification. Another is how to effectively use experts' prior knowledge, such as visual interpretation in the past, to enhance the degree of automation and intelligence of classification (Mennis and Guo 2009).

From the point of handling scale, current image classification algorithms can be divided into two categories: pixel-based and object-oriented. The former views the characteristics of each pixel as a vector. They classify the pixels of entire images by matching the feature vectors and pre-selected samples (Togi et al. 2013). Such methods use only spectral characteristics of pixels and then relatively simple. With the rapid development of spaceflight, sensors and computer technologies, the spatial resolution of optical remote sensing has improved significantly. High spatial resolution optical remote sensing imageries are obtained with voluminous data, more ground details, and more serious spectral confusion. It makes traditional pixel-based classification only relying on image spectrum in applicable, as a host of additional kinds of characteristics are ignored within pixel-based methods. While object-oriented methods, through clustering pixels into objects via

image segmentation, can take the spectral characteristics, texture, shape, and topology features into account. Multiple types of information are utilized together (Mennis and Guo 2009), which is more in line with the principles and process of human visual interpretation. Hence, they gradually become the development trend of classification of high-resolution images (Togi et al. 2013). In this paper, we will employ the object-oriented classification as a rudimentary framework.

Sample Collection

Among the fore-mentioned supervised classification methods, the quantity and quality of training samples are of great importance. Existing land cover classification methods always require adequate training samples to learn a predictive model. Most previous work collected samples manually because of the importance and complexity of samples. Many studies have shown that the accuracy of a classification varies as a function of the range of training set, like number of samples, space distribution of samples (Fardanesh and Ersoy 1998; Tsai and Philpot 2002). Moreover, the training samples of a classification typically need the use of a lot of randomly selected pure pixels in order to characterize the classes. However, human annotation is a time consuming work, which makes labelled data costly to be obtained in practice. Therefore, Maulik and Chakraborty (2011) proposed a semi-supervised support vector machine that uses self-training approach to decrease the number of samples. Tuia et al. (2009) introduced an active learning framework for land cover classification, permitting people to select samples as less as possible. All these approaches can reduce the training set size without significant influence on the classification accuracy. However, they didn't fundamentally change the human involvement in the process of sample collection. It is still the main contradiction in the land cover classification, since the type labels are difficult or expensive to be obtained. This motivates us to improve the process of sample collection.

Transfer Learning

At present, various optical remote sensing satellites are producing images with high time resolution. A lot of time-series new images are constantly received, which constitutes a time-intensive spatial data. For classification tasks on these new images, however, previously collected training samples in the libraries often outdated. Then, generally speaking, the assumption of traditional machine learning theory cannot be achieved. That is, the training samples and test samples should be identically distributed with a same probability distribution. That is because the spectral data obtained from satellite sensors often change over time due to the factors of atmospheric absorption and scattering, sensor calibration, solar elevation angle, azimuth, phenological phase, and data processing. These factors make that the spectral values of "training samples

from pervious old images" and "test samples from current new images" cannot obey a same statistical probability distribution. In this circumstance, traditional machine learning theory cannot be carried out and a large number of new samples are needed to be re-labeled. Then, the collection of new training samples becomes a necessary step for an accurate classification. While it is a time, manpower, material resource–consuming process. However, the historical thematic land cover maps are rich in amount of prior knowledge. It is economical to discover beneficial knowledge from these data to assist current classification tasks. Hence, collecting samples by transferring historical knowledge to current classification can be a breakthrough point for automatic classification.

As a new kind of machine learning theory, transfer learning (TL) was proposed to solve the problem of transferring existing prior knowledge learned from an environment to help learning in a new environment. It emphasizes to transfer of knowledge across domains, tasks, and distributions that are similar but not the same. Therefore, it relaxes the strict assumptions of traditional machine learning theory, namely the same statistical probability distribution of training data and test data (Dai et al. 2008; Pan and Yang 2010). In recent years, researches in this area have been increasingly applied in real life, including text processing, LAN positioning, emotional classification. This theory is believed to be suitable for the classification tasks of updating land cover with historical thematic maps. However, its application in the classification of remote sensing images is rare. This paper aims to use the idea of transfer learning to partly solve the problem of manual sample collection.

In the field of transfer learning, the data set D is divided into two parts: source date set $D_s = \{(x_s^i, l_s^i)\}_{i=1}^m$ and target data set $D_t = \{(x_t^i, l_t^i)\}_{i=1}^n$ (x are the values of data and l are their labels), i.e. $D = D_s \cup D_t$. And the target data set is further divided into training set $D_{t-train} = \{(x_t^i, l_t^i)\}_{i=1}^k$ and test set $D_{t-test} = \{(x_t^i, l_t^i)\}_{i=k+1}^n$, i.e. $D_t = D_{t-train} \cup D_{t-test}$. Here all data in D are in the same feature space, i.e., all the characteristics can be described in the feature space, while the data in D_s and D_t are generated from different but similar domains. On the one hand, the data set $D_{t-train}$ is generally obtained from the manual tagging in practical applications. As a small amount of data in $D_{t-train}$ is often not sufficient to train a good classifier, the task of manual tagging (i.e. human annotation) is tremendously huge for the requirement of a high-performance classifier. On the other hand, there are a large amount of labelled data in D_s. However, they cannot be directly used to train D_t due to their different distributions. Some schemes need to be completed on D_s for feature subset selection. The goal of transfer learning is to label the data in D_t by transferring the existing priori knowledge in source domain D_s to target domain D_t.

In the framework of transfer learning, there is no strict assumptions on the same distribution of training data and test data, and knowledge in source fields can be transferred through different ways. Hence, three issues are researched in

the theory, namely what to transfer, how to transfer and when to transfer. Based on the issue of "what to transfer", current methods for solving knowledge transfer can be summarized into four categories (Pan and Yang 2010), namely instance transfer, feature transfer, parameter transfer, and relational knowledge transfer. Within the framework of relational knowledge transfer, this paper will propose a change detection-based transfer method to incorporate ancillary data for interpreting new remote sensing images.

Methodology

According to the characteristics of remote sensing images and the demand of automatic land cover updating, an automatic object-oriented hierarchical classification (AOHC) method is designed in this section. The methodology is a cluster of a series of algorithms and focuses on automatic object sample collection under the guidance of prior knowledge. The flow chart is illustrated in Fig. 1, including the following steps.

(1) First, pre-processing technologies, including automatic registration and radiometric normalization, are carried out on the multi-phase optical remote sensing images. Then, segmentation based mean-shift technology is applied to the second-phase target image, which converts the handing scale from pixels to objects. Features of extracted objects then can be calculated for pattern learning.

(2) Second, after change detection between the images from different phases, object samples for rough land cover classification are automatically selected with the assistance of prior knowledge (i.e. the thematic maps in first-phase and spectral library data). Then, rough land cover classified map is updated utilizing a supervised object-oriented classification algorithm. In the methodology, all the information achieved from the thematic maps and spectral library data, including land cover types and locations, is defined as prior knowledge.

(3) Finally, based on the results of step (2), object samples for detailed land cover classification are further automatically selected with the auxiliary data under a series of strict rules described in subsection 3.2. Hierarchical object-oriented classifications for detailed land cover are then carried out for each rough land cover type. The second-phase classified map of detailed land cover is then updated.

According to above pre-set processes, the procedure of this methodology can automatically complete except setting the region of interest. Following subsections will give more details about the implementation of several key steps during the procedure.

Segmentation and Feature Extraction

Segmentation Based on Mean-Shift

Given the robustness of segmentation, we resort a mean-shift (MS)-based algorithm to extract objects from optical remote
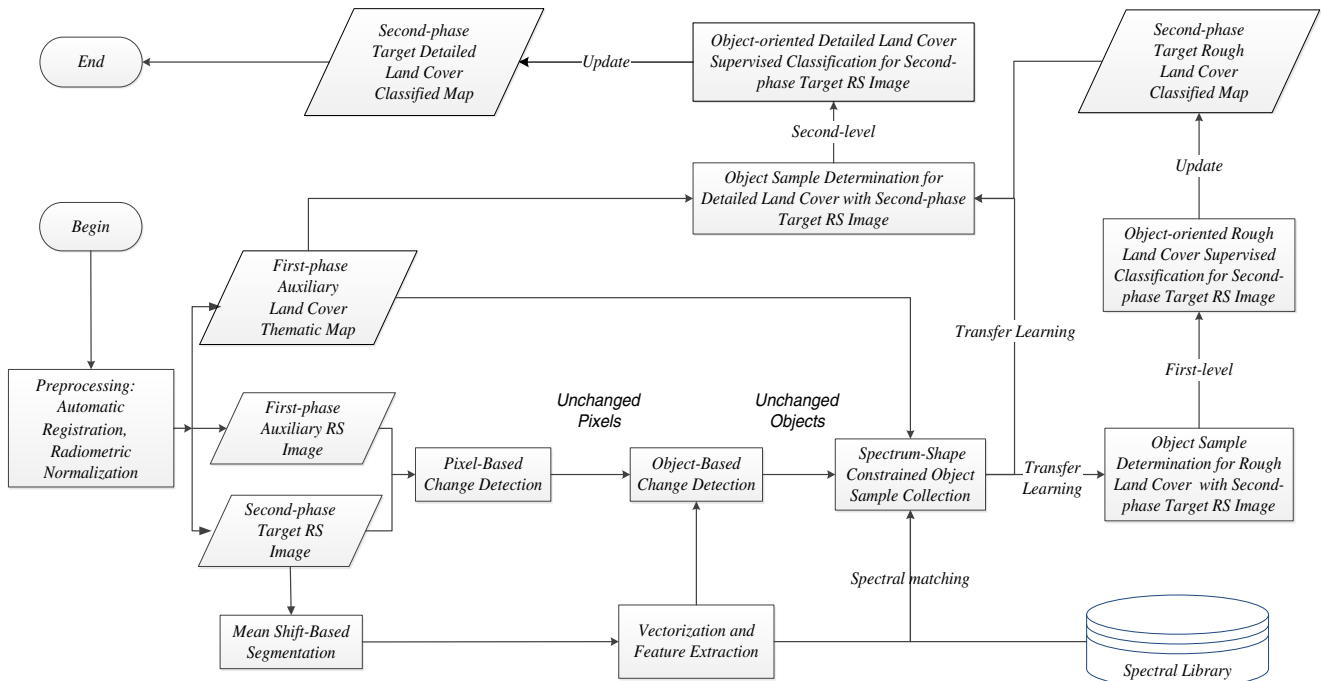
Fig. 1 Flow chart of the automatic object-oriented hierarchical classification for updating detailed land cover

sensing images. As a kind of region-based segmentation method, this algorithm completes non-parametric density function estimation and automatic clustering through iteratively shift means to the local maxima of density functions in the feature space (Comaniciu and Meer 2002).

Given n points $x_i(i=1,\ldots,n)$ in d-dimensional space, the kernel density at point x can be written as

$$\widehat{f}_{h,K}(x) = \frac{c_{k,d}}{nh^d}\sum_{i=1}^{n} k\left(\left\|\frac{x-x_i}{h}\right\|^2\right), \tag{1}$$

where $c_{k,d}$ is a normalization constant, h is the bandwidth, and $k(\bullet)$ is the kernel profile for the estimation. The local maxima of the density will be searched in feature space by locating among the zeroes of the gradient $\nabla\widehat{f}_{h,K}(x) = 0$. The gradient of density can be obtained by

$$\nabla\widehat{f}_{h,K}(x) = \frac{2}{h^2 c}\widehat{f}_{h,G}(x) \times m_{h,G}(x), \tag{2}$$

where the profile of kernel G is defined as $g(x)=-k'(x), c=c_{g,d}/c_{k,d}$ is the normalization constant with $c_{g,d}$ as its normalization parameter. In Eq. (2), the first term is the density estimation at x with the kernel G

$$\widehat{f}_{h,G}(x) = \frac{c_{k,d}}{nh^d}\sum_{i=1}^{n} g\left(\left\|\frac{x-x_i}{h}\right\|^2\right), \tag{3}$$

and the second term is the MS

$$m_{h,G}(x) = \frac{\sum_{i=1}^{n} x_i \times g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} -x. \tag{4}$$

From (4), it can be found that the MS is the difference between the weighted mean, using the kernel G for weights, and x, the center of the kernel. According to Eq. (2), the MS can be written as

$$m_{h,G}(x) = \frac{1}{2}h^2 c\frac{\nabla\widehat{f}_{h,K}(x)}{\widehat{f}_{h,G}(x)}. \tag{5}$$

Equation (5) shows that the MS vector at point x with kernel G is proportional to the normalized density gradient estimate obtained with kernel K, and it thus always points toward the direction of maximum increase in the density. In other words, the local mean is shifted toward the region in which the majority of the points reside (Comaniciu and Meer 2002).

Optical remote sensing images are typically represented as a spatial range joint feature space. The dimensionality of the joint domain is $d=2+p$ (two for spatial domain and p for spectral domain). That is, for a point $x=(x^s,x^r)$, the spatial domain $x^s=(l_x,l_y)^T$ denotes the coordinates and locations for different pixels, and the range domain $x^r=(r_1,\ldots,r_p)^T$ represents the spectral signals for different channels. Then, the multivariate kernel is defined for joint density estimation

$$K_{h_s,h_r}(x) = \frac{C}{h_s^2 h_r^p}k\left(\left\|\frac{x^s}{h_s}\right\|^2\right)k\left(\left\|\frac{x^r}{h_r}\right\|^2\right), \tag{6}$$

where C is a normalization parameter and h_s and h_r are the kernel bandwidths for spatial and range sub-domains. After determined kernel function and its bandwidths, image clustering can be achieved through mean filtering. Regions are then merged according to the minimum merging parameter M. Finally, the technology of vectorization is employed to extract the boundaries of objects. More detailed description of MS algorithm can be found in Comaniciu and Meer (2002).

Feature Extraction

After segmentation, homogenous pixels are merged into objects. Multiple features of objects are then calculated to structure feature thematic layers. There are three kinds of spatial-spectral features of objects, namely spectrum features, shape features and texture features, to be extracted. The spectrum features include objects' spectral signals of each band (i.e. average and standard deviation of spectral signals from all the pixels located in the objects), objects' brightness (i.e. average of spectral signals from all the bands), objects' maximum differences (i.e. maximum variation between the spectral signals of bands), objects' indices (i.e. specific values of assembling spectral signals of various bands, including normalized difference water index (NDWI), normalized difference vegetation index (NDVI)). The shape features consist of geometric features of objects such as length-width ratio, main direction and shape index (Lu and Weng 2007). The texture features comprise the measures of Gray-Level Co-occurrence Matrices (GLGM) (Marceau et al. 1990). As listed in Table 1, by assembling these three kinds of features, 27 features are extract to construct a high-dimensional feature space. Note that the architecture of objects' feature calculation is open for object-oriented classification. Any other features can be easily added to improve the performances of our method.

Object Sample Automatic Collection Based on Change Detection and Transfer Learning

For supervised classification, influence training samples in distinguishing land cover types should be selected for a classification system. The key problem in this step is how to

Table 1 List of "spatial-spectral" features of objects

Spectrum features		Shape features		Texture features	
No.	Feature name	No.	Feature name	No.	Feature name
01	Mean of spectrum values in band 1	13	Length-width ratio	22	Homogeneity
02	Mean of spectrum values in band 1	14	Length of geometry	23	Contrast
03	Mean of spectrum values in band 3	15	Width of geometry	24	Dissimilarity
04	Mean of spectrum values in band 4	16	Compactness	25	Second moment
05	Standard deviation of spectrum values in band 1	17	Main direction of geometry	26	Entropy
06	Standard deviation of spectrum values in band 2	18	Number of points	27	Correlation
07	Standard deviation of spectrum values in band 3	19	Length of border	–	–
08	Standard deviation of spectrum values in band 4	20	Shape index	–	–
09	Brightness	21	Number of corner points	–	–
10	Maximum differences	–	–	–	–
11	NDVI (Normalized Difference Vegetation Index)	–	–	–	–
12	NDWI (Normalized Difference Water Index)	–	–	–	–

automatically collect object samples on the target images. In this subsection, a prior knowledge-based object sample automatic selection scheme is proposed after 27-dimensional features are extracted from the dataset.

Note that the interested land cover types of a regional area change little in a short period. We suppose that most of the types are unchanged. The positions and features of the unchanged landmarks can be extracted as prior knowledge. Under this assumption, an automatically collect object samples scheme based on change detection and transfer learning is designed with data mining technologies. Its main idea is to view pre-interpreted thematic maps and outdated spectral library data as important references for sample collection. Unchanged landmarks' locations, land cover types, and features are combined together to guide object sample collection on target images.

The concrete implementation is described in Fig. 2. First, change detection is carried out between the optical remote sensing images of two phases, and a large number of unchanged landmarks can be located. In this paper, change vector analysis (CVA) and Otsu's thresholding methods (Lu et al. 2004) are jointly applied to automatically detect unchanged pixels and their spatial locations. Locations of unchanged pixels will be marked on the second-phase target image. Second, prior interpretation knowledge of categories (i.e. labels) from the first-phase interpreted map is transferred to the new target image. Third, four rules with thresholds for filtering objects are employed to select trustworthy samples. That is, the sizes of objects, the percentages of pixels labeled in objects, the proportion of pixels marked by a same label, and the matching degree of objects' average spectrum must be larger than given thresholds. Under these constraints, the pixel's
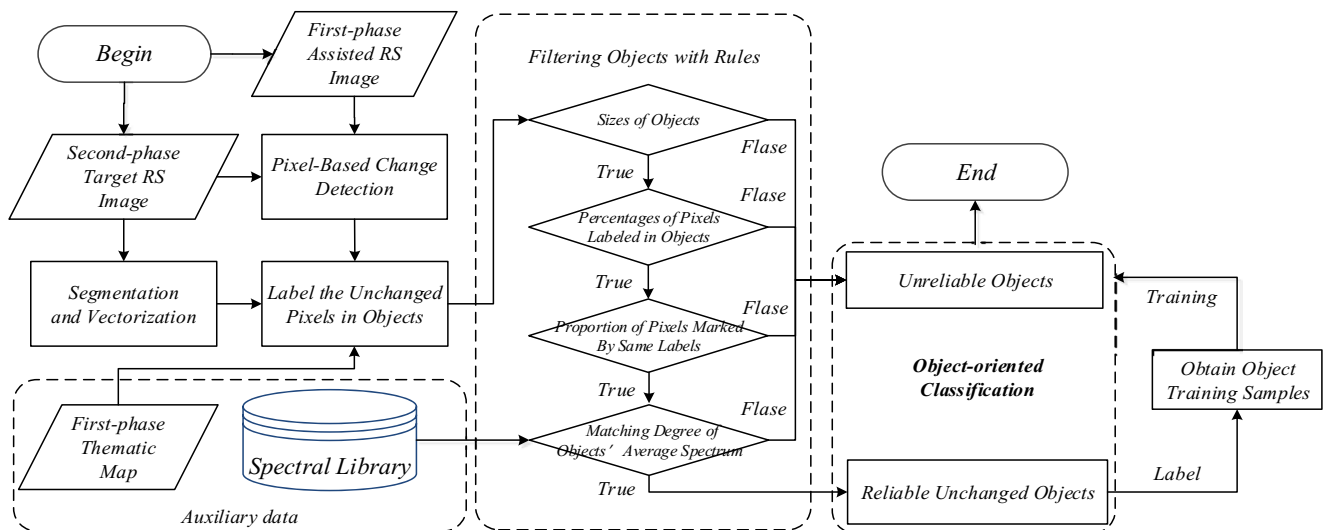
Fig. 2 Diagram of automatic collection of object samples with the assistance of auxiliary data

label with the highest proportion in an object is employed as this objects' label.

Among these thresholding rules of filtering objects, spectral matching algorithm is used to determine the degrees of membership of auto-selected objects, which indicates the probability of the objects to be distinguished as a specific land cover type. To determine which land cover type a specific object belongs to, a spectral matching algorithm called spectral similarity is applied, as given below

$$d^2(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{p} (x_i - y_i)^{\mathrm{T}} \mathbf{A}(x_i - y_i) + \left(1 - \frac{\sum_{i=1}^{p}(x_i - \mu_x)(y_i - \mu_y)}{p\,\sigma_x\sigma_y}\right)^2, \quad (7)$$

Here $\mathbf{x} = (x_1, \ldots, x_p)^{\mathrm{T}}$ is the average spectrum vector of an object whose type is unknown, and $\mathbf{y} = (y_1, \ldots, y_p)^{\mathrm{T}}$ is the average spectrum vector of a known land cover type from the spectral library. μ_x, μ_y are the mean values and σ_x, σ_y are the standard deviations. The first term in the right part of Eq. (7) is Mahalanobis distance rather than Euclidean distance. And $\mathbf{A}$ denotes a positive semi-definite matrix which is learned from prior knowledge such as relevance and neighborhood (Xing et al. 2002). This distance metric decreases the distance between similar vectors and increases the distance between dissimilar vectors. The second term is the correlation coefficient by which the overall shape difference between the reflectance curves can be exploited. As a result, the resulting distance matches to the total difference between $\mathbf{x}$ and $\mathbf{y}$. Both of these two terms range in [0,1], and $d \in [0, \sqrt{2}]$. The smaller d is, the larger matching degree of objects' average spectrum is.

Additionally, note that, there are still incorrect labels among these object samples satisfying mentioned thresholding rules. The undesirable samples need to be further eliminated for acquiring robust learning rules. Therefore, finally, the synthetic minority over-sampling technique (SMOTE) and neighborhood cleaning rule (NCR) (Jerzy and Szymon 2008) are employed to handle the imbalanced skew-slash distribution of automatically collected object samples. Thus, reliable objects with labels are filtered out as highly credible training samples for pattern learning. Then, through training these collected samples, the relationship between landmark classes and their spatial-spectral features then can be established to classify other unlabeled objects. Apparently,

the entire procedure of object sample collection is automatically executed with the assistance of previous thematic maps and spectral library data.

Object-Oriented Hierarchical Classification

To accurately obtain detailed land cover, a hierarchical classification is developed in this field. In this paper, a two-level tree-structured hierarchy is employed to stratify land cover types, namely rough land cover and detailed land cover. An example of tree-based hierarchical class structure is shown in Fig. 3.

As shown in Fig. 3, classification will be performed in a hierarchical two-level land cover types, one running with rough land cover types and the other running with detailed land cover types. After automatic acquisition of object samples for rough land cover classification, supervised object-oriented classification method is adopted to extract rough land cover information. Taking into account the efficiency of learning, we employ C5.0 decision tree algorithm (Barros and Basgalupp 2012) to produce a rough land cover classified map. This procedure is referred as first-level rough land cover classification. Subsequently, for each type of rough land cover, object samples for detailed land cover classification are automatically collected via the scheme in Fig. 2. At this level, samples establish a link between the rough and detailed land cover types, which can be constructed as a pattern to further wipe out some unreliable samples. That is, if the rough land cover types of some labeled objects are different from their classified rough land cover type, they will be automatically wiped out for training. Suitable discrimination rules are then formatted via carrying out C5.0 decision tree on the selected object samples. Decision trees are then constructed to complete the classification for each rough land cover type. We call this procedure a second-level detailed land cover classification. Through this two-level hierarchical classification, detailed land cover information of the second-phase target map is finally updated. This automatic object-oriented classification with this two-level hierarchical classifier is referred to *automatic object-oriented hierarchical classification* (AOHC) method in this paper. By contrast, the term "*automatic object-oriented global classification* (AOGC)" is used to define the method directly using an ordinary global classifier on the level of detailed land cover types. For a comprehensive
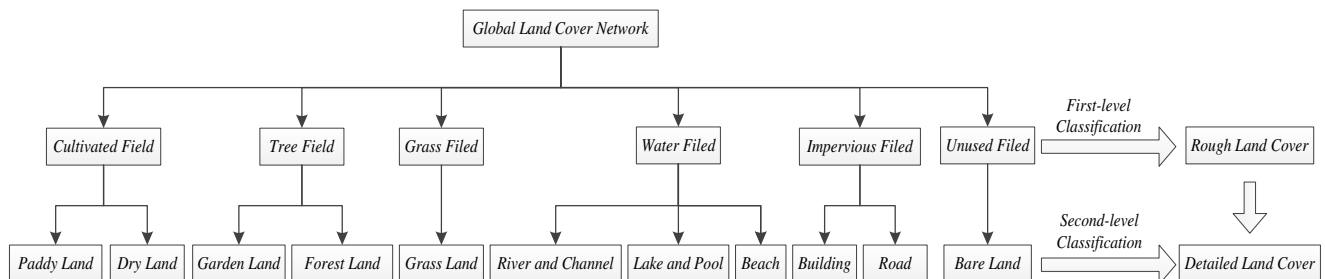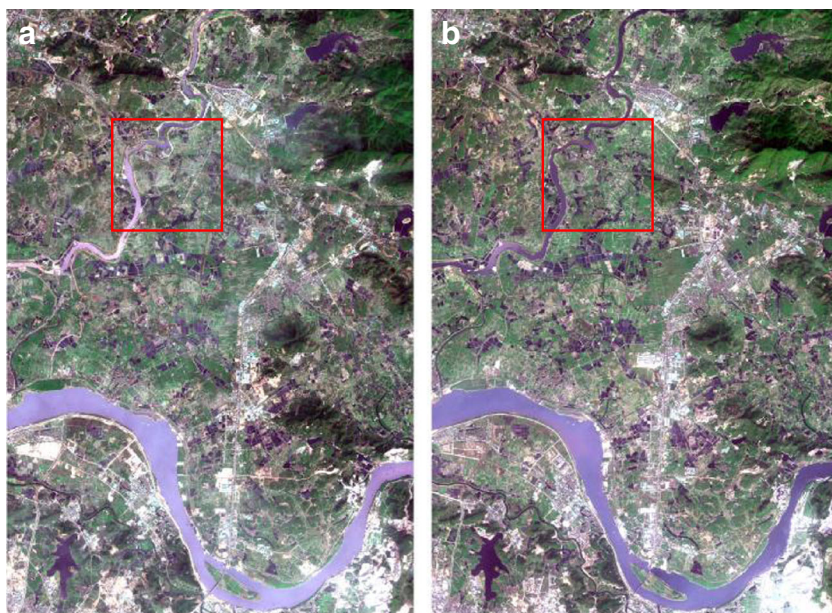

Fig. 3 An example of a tree-based hierarchical class structure for land cover types

Fig. 4 Previous auxiliary and current target SPOT5 optical remote sensing images for the experiment with data set A: (**a**) previous auxiliary image ((Hinton et al. 2006)-04-13); (**b**) current target image (2008-05-15)

introduction to hierarchical classification theory, algorithms and some of its applications, we refer to (Carlos and Alex 2011).

Experiments

Data Set Description

Data Set A

Two Systeme Probatoire d'Observation dela Tarre 5 (SPOT5) images are first selected to evaluate the proposed method. Figure 4 shows their true color compositions by fusing

panchromatic band and multi-spectral bands. They were acquired over Dongguan City in South China, on April 13, 2006 and May 15, 2008, respectively. They are 2.5 m in spatial-resolution and 4977×7369 pixels in size. The image in 2006 year (Fig. 4a) is referred as the first-phase auxiliary image, while the image in 2008 year (Fig. 4b) as the second-phase target image. They are certainly different in spectral signals due to inconsistent phase and then physical radiation, which inevitably makes the spectral signals of two images subject to different statistical probability distributions. Thus, the samples from Fig. 4a cannot be directly adopted as training samples for the classification of Fig. 4b. But they can be viewed as spectrum library data for spectral matching. To match the inputs of our approach, a thematic map in
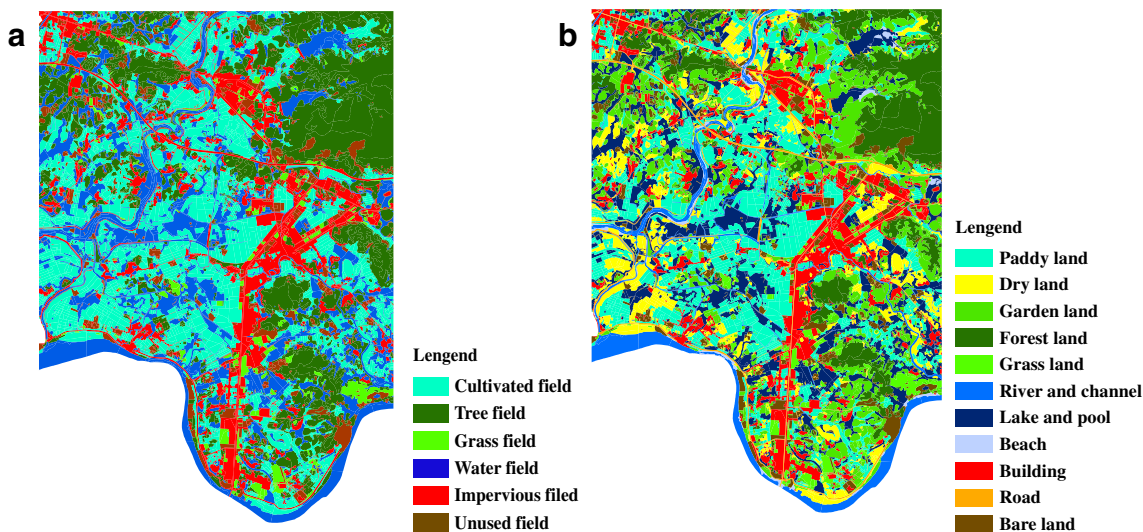
Fig. 5 Previous (first-phase) auxiliary thematic map for data set A (2006-04-13): **a** Auxiliary thematic map of rough land cover type; **b** auxiliary thematic map of detailed land cover type

Fig. 6 Previous auxiliary and current target ZY3 optical remote sensing images for the experiment with data set B: **a** previous auxiliary image (2012-11-05); **b** current target image (2013-03-18)
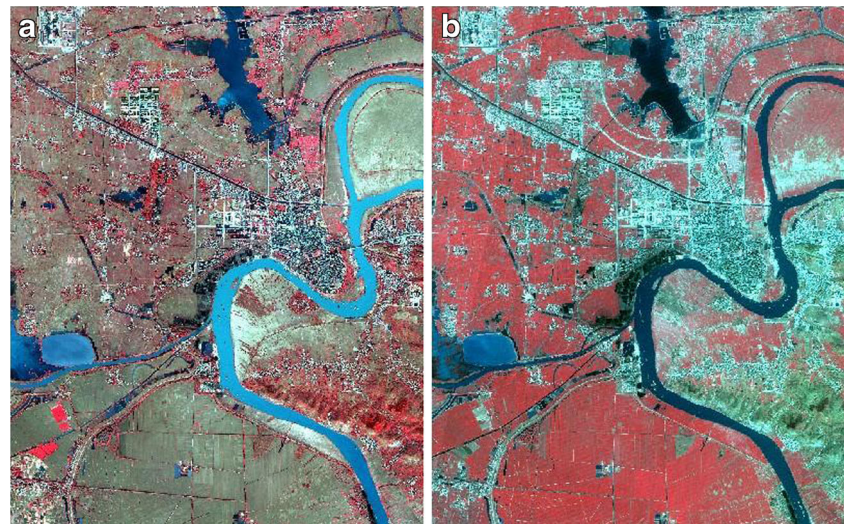
accordance with Fig. 4a, is collected as an auxiliary data. As shown in Fig. 5a, this region contains many typical land cover categories. Each type of land cover has obvious visual features and thus can be interpreted. Therefore, this selected data set is suitable for experiment to analyze our proposed algorithm.
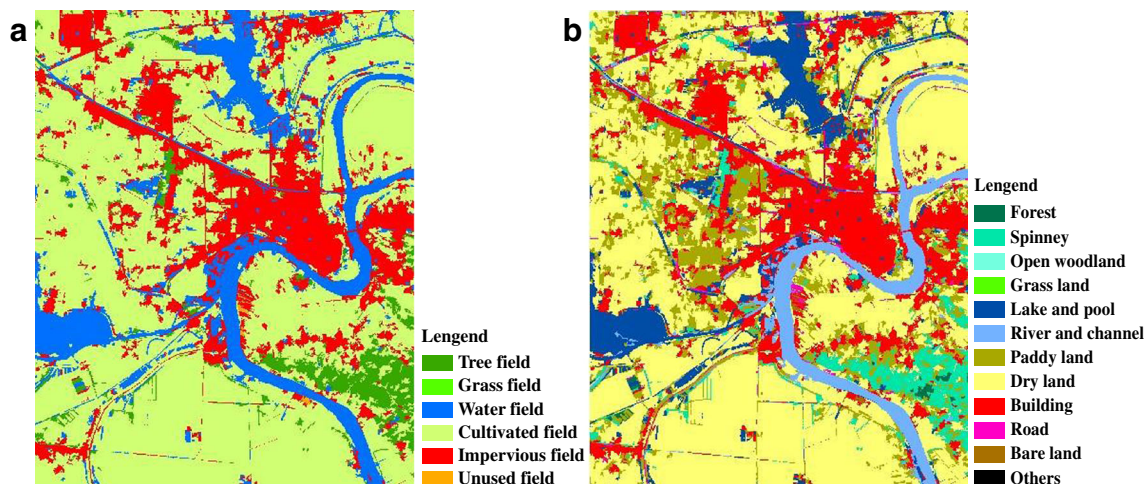
Data Set B

Demonstrate works on data set B are carried out as another case. Figure 6 presents two Zi Yuan 3 (ZY3) satellite images from Satellite Surveying and Mapping Application Center (SBSM). They were obtained over Huainan in central China, on Nov. 05, 2012 and Mar. 18, 2013, respectively. The size is 5035×6338 pixels and the spatial-resolution is 2.1 m. Figure 6a is referred as the first-phase auxiliary image, while Fig. 6b as the second-phase target image. Due to inconsistent phase, the images also have different statistical probability distributions in spectral signals. Auxiliary thematic maps are

shown in Fig. 7, including the maps of rough land cover type and detailed land cover type extracted according to Fig. 6a.
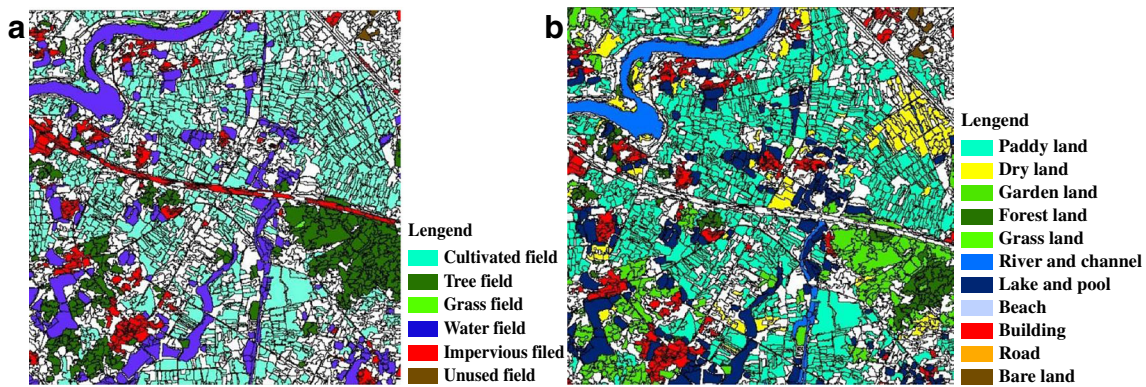
Results

Data Set A

By comparing the features of Fig. 4a and b, the result of pixel-based change detection is calculated through jointly using CVA and Otsu's thresholding methods. According to the characteristics of this data set, the kernel bandwidths and minimum merging parameter are set as $h_s=7$, $h_r=6.5$, $M=150$, respectively, for MS-based segmentation. Figure 8 presents the segmentation results of the marked area utilizing the red rectangles in Fig. 4. In addition, with the assistance of Fig. 5 and pre-reserved spectral library data, a large number of object samples are automatically collected under the scheme of Fig. 2. The constraint rules in the scheme consist of: (1) the

Fig. 7 Previous (first-phase) auxiliary thematic map for data set B (2012-11-05): **a** Auxiliary thematic map of rough land cover type; **b** auxiliary thematic map of detailed land cover type

Fig. 8 Results of MS-based segmentation and distributions of object samples for data set A using the AOHC approach: **a** distribution of object samples for rough land cover; **b** distribution of object samples for detailed land cover

sizes of objects must be larger than 50 pixels; (2) the percentages of pixels labeled in objects must be larger than 90 %; (3) the proportion of pixels marked by same labels must be larger than 90 %; and (4) the matching degree of objects' average spectrum based Eq. (7) must be larger than 90 %. The distributions of the selected training object samples are also partly shown in Fig. 8. Figure 9 presents the updated thematic maps for different land cover levels through learning decision trees based on the collected object samples.

To quantitatively investigate the performance of the proposed algorithm, validation points are collected to evaluate the accuracy of these results. The procedure of accuracy assessment is illustrated as follows. First, to ensure the uniformity of the validation points' distribution, 10×10 regular grids are generated according to the size of the images. Ten randomly generated points are then located in each grid to guarantee the randomness of the points. Second, through the visual interpretation, the land cover types of these points are recorded. Third, to keep the validation points balance in quantity, 600 verification points are picked out to verify the effectiveness of the

method. Four, classified types of each point are obtained from Fig. 9. Finally, by comparison of the results of artificial interpretation and automatic updating, accuracies of the proposed algorithm can be calculated using confusion matrices.

The updating accuracies of rough land cover type using AOHC method are summarized in Table 2 and 3. It can be seen from these tables that major of rough land cover types can update accurately by the AOHC method. Although the updating accuracies of cultivated field and grass field are not high (81.81 % and 77.59 %, respectively), the accuracies of water field and impervious fields are relative high (92.63 % and 87.50 %, respectively). Since there are no rough land cover types with particularly low recognition rates, the overall accuracy and kappa coefficient can reach to 84.50 % and 0.8179, respectively. In addition, accuracy evaluation of detailed land cover updating using the AOHC method is subsequently presented in Table 4. As shown in this table, the proposed method can automatically extract detailed land cover information with an acceptable performance. Furthermore, except the types of dry land and road, other land cover types
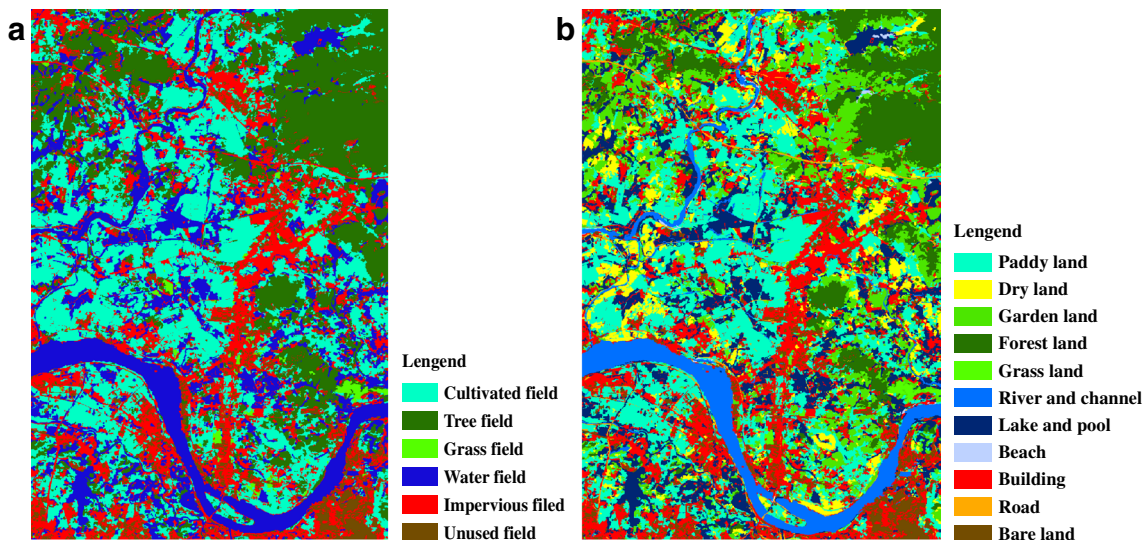
Fig. 9 Results of land cover updating for data set A using the AOHC approach: **a** result of rough land cover updating; **b** result of detailed land cover updating

Table 2 Confusion matrix of rough land cover updating using the AOHC approach (data set A)

Rough land cover type	Number of artificially interpreted samples						Producer accuracy (%)
	Cultivated field	Tree field	Grass field	Water field	Impervious field	Unused field	
Cultivated field	81	13	6	0	0	0	81.00
Tree field	7	89	4	0	0	0	89.00
Grass field	3	0	90	7	0	0	90.00
Water field	0	3	9	88	0	0	88.00
Impervious field	8	0	0	0	77	15	77.00
Unused field	0	0	7	0	11	82	82.00
User accuracy (%)	81.81	84.76	77.59	92.63	87.50	84.50	–
Overall measures	Overall accuracy (%): 84.50			Kappa coefficient: 0.8179			

can be updated accurately. Overall accuracy of detailed land cover updating using the AOHC method can reach to 79.50 %, which suggests the updated results of detailed land cover are highly consistent with their actual categories.

A key scheme of the proposed method is the use of hierarchical classification for updating. A comparative experiment is conducted to investigate on the role of the two-level hierarchical classifier. Table 5 shows itemized accuracy of each detailed land cover type for the AOGC approach. As shown in the Table 6 and 7, each algorithm can achieve a high performance on the validation data via automatic sample collection. However, by comparing with the two-level hierarchical classifier, global classification in the AOGC method results in a considerable drop in performance. The performance of our proposed method can reach 79.50 % at level of detailed land cover types while the AOGC only achieves 76.67 % at the same level. That is, 2.83 % improvement in accuracy can be produced for AOHC method with data set A, which illustrates the hierarchical strategy in AOHC method outperforms the one-level global strategy. This is due to hierarchical classification in the AOHC method constructs a top-down classification to make the types less as the level gets deeper. Another factor for the marginal improvements is the use of global information compensates for the local training sample collection within the second-level classification.

Data Set B

With the same parameter setup in the experiment of data set A, the segmentation results of data set B are shown in Fig. 10. The distributions of the selected training object samples are also presented in Fig. 10. Figure 11 shows the updated thematic maps for data set B. It can be found that the changes over polygons are obviously updated, which suggests our method partly overcomes the limitation of small separability between different detailed land cover types with similar spectrum signals.

Similar to the accuracy assessment in the experiment of data set A, the results of data set B are also validated using uniformly collected verification points. 600 verification points are resorted to calculate the corresponding updating accuracy. Table 6 and 7 summarize the updating accuracies of rough land cover type. It can be seen from these tables, the updated results of rough land cover are acceptable as the overall accuracy and kappa coefficient reach to 91.83 % and 0.8764, respectively. As shown in Table 7, the classification accuracies of tree fields, grass field, water field and impervious field are also relatively high (93.8, 90.5, 98.0 and 92.7 %, respectively). Furthermore, accuracy evaluation of detailed land cover updating is presented in Table 8. We find that most of detailed land cover types can be accurately updated via utilizing a two-

Table 3 Statistical table of itemized accuracy of each rough land cover type for the AOHC approach (data set A)

Rough land cover type	Number of artificially interpreted samples	Number of correctly classified samples	Accuracy (%)	Main misclassification
Cultivated field	99	81	81.81	Tree field; Impervious field
Tree field	105	89	84.76	Cultivated field
Grass field	116	90	77.59	Water field
Water field	95	88	92.63	Grass field
Impervious field	88	77	87.50	Unused field
Unused field	97	82	84.50	Impervious field
Total	600	507	84.50	–

Table 4 Statistical table of itemized accuracy of each detailed land cover type for the AOHC approach (data set A)

Detailed land cover type	Number of artificially interpreted samples	Number of correctly classified samples	Accuracy (%)	Main misclassification
Paddy land	66	52	78.79	Dry land
Dry land	33	22	66.67	Paddy land
Garden land	31	28	90.32	Grass land
Forest land	74	58	78.38	Grass land
Grass land	116	90	77.59	Garden land
River & channel	65	50	76.92	Grass field
Lake & pool	22	17	77.27	River & channel
Beach	8	6	75.00	Grass field
Building	56	48	85.71	Road
Road	32	21	65.63	Building; dry land
Bare land	97	85	87.63	Dry land
Total	600	477	79.50	–

level hierarchical classification. For instance, the updating accuracies of open woodland, grass land, river & channel, lake & pool and building can reach to 91.67, 90.48, 96.55, 93.18 and 95.24 %, respectively. Updating errors of these detailed land cover types are effectively prevented as spectrum, shape and texture features are jointly considered during the process of object-oriented classification. It suggests that the performance of detailed land cover updating is acceptable in an automatic manner.

In order to compare our method against the AOGC algorithm, we then implement the two algorithms on date set B. The overall performances for the AOGC algorithm are listed in Table 9. Similarly, comparing the results in Table 8, the AOHC algorithm can achieve consistent improvement over the AOGC algorithm at the level of detailed land cover types. As shown in Table 8 and 9, the performance of our proposed algorithm can reach 86.33 % at level of detailed land cover types while the AOGC only achieves 82.33 % at the same level. That is, our top-down approach can get about 4 %

improvements over the global approach on data set B. The performance of our two-level classification achieves higher and that of one-level global classification is significantly reduced over the multiple level categories. As a result, with a pruned two-stage hierarchy, the proposed algorithm can make a more accurate updating performance. Its top-down structure of the hierarchy can be effectively utilized to improve the updating efficiency.

Discussions

Misclassification

Note that the accuracies of the experiments with data set A are lower than those of the experiments with data set B. The reasons is that error propagation exists in several procedures of the proposed algorithm. It is a key issue for the designed algorithm and generalized from the following aspects. First, as shown in Fig. 1, the samples are selected from the unchanged

Table 5 Statistical table of itemized accuracy of each detailed land cover type for the AOGC approach (data set A)

Detailed land cover type	Number of artificially interpreted samples	Number of correctly classified samples	Accuracy (%)	Main misclassification
Paddy land	66	50	75.76	Dry land
Dry land	33	21	63.64	Paddy land; bare land
Garden land	31	26	83.87	Paddy land; grass land
Forest land	74	57	77.03	Grass land;
Grass land	116	85	73.28	Garden land
River & channel	65	49	75.38	Grass field
Lake & pool	22	16	72.73	River & channel
Beach	8	6	75.00	Grass field
Building	56	47	83.93	Road; dry land
Road	32	21	65.63	Building; dry land
Bare land	97	82	84.54	Dry land
Total	600	460	76.67	–

Table 6 Confusion matrix of rough land cover updating using the AOHC approach (data set B)

| Rough land cover type	Number of artificially interpreted samples						Producer accuracy (%)
	Cultivated field	Tree field	Grass field	Water field	Impervious field	Unused field	
Cultivated field	89	6	5	0	0	0	89.00
Tree field	9	91	0	0	0	0	91.00
Grass field	3	0	95	2	0	0	95.00
Water field	0	0	0	100	0	0	100.00
Impervious field	1	0	0	0	89	10	89.00
Unused field	1	0	5	0	7	87	87.00
User accuracy (%)	86.41	93.81	90.48	98.04	92.71	89.69	–
Overall measures	Overall accuracy (%): 91.83			Kappa coefficient: 0.8764			

positions, while the results of change detection may contain errors. Second, diagram of automatic collection of object samples in Fig. 2 achieves the purposes of sample purification. However, it cannot weed out all the undesirable objects from the training samples. Third, as the labels of samples are directly transferred from the auxiliary data, the reliability of the AOHC depends upon the accuracies of auxiliary thematic data. While the auxiliary data is partly inaccuracy (i.e. the thematic maps have been classified in advance may contain a certain degree of errors). For instance, in the experiments, the accuracies of the auxiliary thematic map in data set A (Fig. 5) are lower than that in data set B (Fig. 7), which makes the updating accuracies have the difference. From the factors presented above, samples with incorrect labels will be collected to learn the classification rules. Then, errors will be propagated to the final classification results. Finally, misclassification of rough land cover types will be propagated downwards to all its descendant detailed classes. Consequently, the accuracy of update results depend on the accuracy of change detection, the first classified thematic map, automatic sample collection and the robustness of classifiers. These essential factors could be further improved. For example, error detection can be considered in sample collection strategy to obtain "high purity" samples, and error tolerance in robust classifiers can be carried out to obtain more reliable classification results. In addition, by examining the misclassified sample points in

the experiments, we find that a part of misclassification can be summarized as the following circumstances. (1) The verification points are located on the boundary of two different classes. Therefore, we cannot determine their categories according to visual interpretation. For instance, the different results of classification and artificial interpretation usually occur for the validation points on the fuzzy boundaries of water fields and grass fields. (2) The verification points are located in the easily confused classes, like cultivated fields and grass fields, garden lands and forest lands, dry lands and bare lands. Hence, parts of the occurrence of misclassification are produced by erroneous artificial interpretation. If the errors caused by artificial interpretation can be completely excluded, the updating accuracies of the proposed AOFC should be theoretically larger than the aforementioned accuracies.

It should be noted that the fundamental step of the proposed AOHC is object sample automatic collection based on prior knowledge. As described in subsection 3.2, this procedure combines the processes of change detection and classification. Through change detection, spatial location and spectrum matching between historical data and current data, the accumulated knowledge in the historical data is explored and further transferred to the new tasks. Then, the problem of inconsistent distributions between images from different phases can be solved by transferring this prior knowledge, which contributes to improve the effectiveness of land cover updating.

Table 7 Statistical table of itemized accuracy of each rough land cover type for the AOHC approach (data set B)

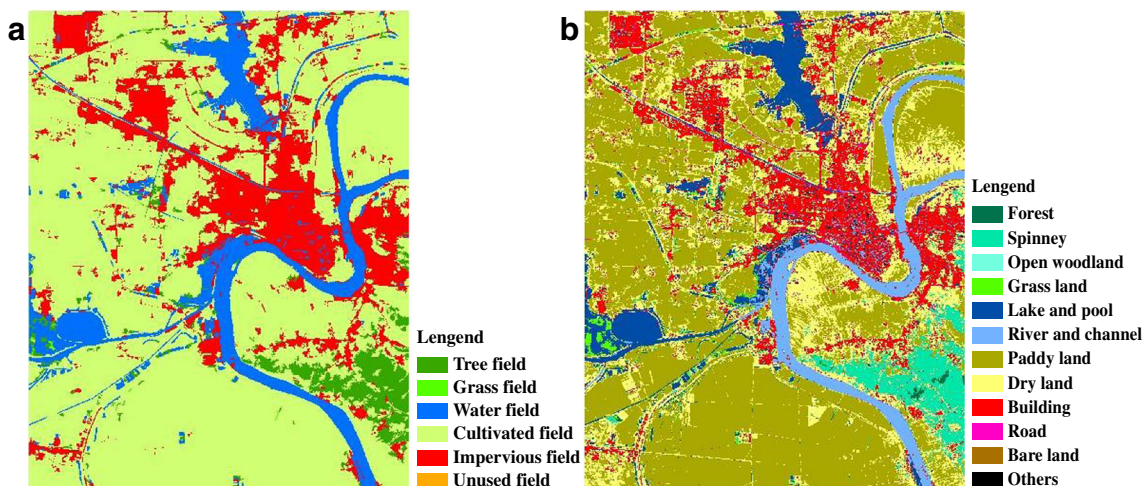| Rough land cover type	Number of artificially interpreted samples	Number of correctly classified samples	Accuracy (%)	Main misclassification
Cultivated field	103	89	86.41	Tree field; Grass field
Tree field	97	91	93.81	Cultivated field
Grass field	105	95	90.48	Cultivated field
Water field	102	100	98.04	Grass field
Impervious field	96	89	92.71	Unused field
Unused field	97	87	89.69	Impervious field
Total	600	551	91.83	–
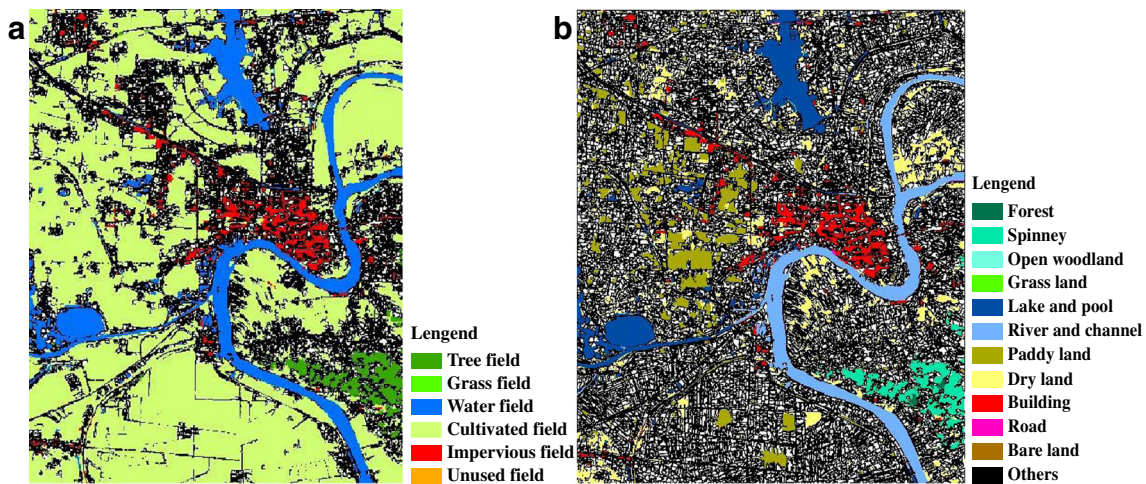
Fig. 10 Results of MS-based segmentation and distributions of object samples for data set B using the AOHC approach: **a** distribution of object samples for rough land cover; **b** distribution of object samples for detailed land cover

Another notable innovation of the AOHC is the two-level tree-structured land cover updating. Traditional classification algorithms typically predicate only classes at the leaf nodes of the tree (Bishop 2006). As shown in Fig. 3, however, there is an obvious class hierarchy between the types of rough and detailed land cover. Traditional approaches have the serious disadvantage of having to build a classifier to discriminate among a large number of classes (all leaf classes), without exploring information about parent–child class relationships present in the class hierarchy (Carlos and Alex 2011). To avoid ignoring this hierarchy information, this paper employed the hierarchical classifier to seek a better separability among the detailed land cover types. As shown in the experiments, more efficient performances are achieved by applying the hierarchical scheme. The results prove that it is necessary to perform the classification stage for deep classification algorithm, which can lead to more precise results for the

deep hierarchy. This suggests that local information in each rough land cover is useful for detailed land cover updating. However, as a representative top-down approach, the presented AOHC has the disadvantage that an error at the level of rough land cover classification is going to be propagated downwards to the level of detailed land cover classification. Misclassification of first-level nodes will be propagated downwards to all its descendant classes. This confirms the initial conjecture that selecting the right branch of the tree at the highest level plays an essential role in determining the class of land cover. Therefore, in the near future, some procedures for the problems of structuring a reasonable hierarchical tree and avoiding error propagation need to be further developed.

Furthermore, through change detection and transfer learning, prior knowledge is certified to be effective for automatically collecting samples in current updating tasks. However,

Fig. 11 Results of land cover updating for data set B using the AOHC approach: **a** result of rough land cover updating; **b** result of detailed land cover updating

Table 8 Statistical table of itemized accuracy of each detailed land cover type for the AOHC approach (data set B)

Detailed land cover type	Number of artificially interpreted samples	Number of correctly classified samples	Accuracy (%)	Main misclassification
Paddy land	73	62	84.93	Dry land
Dry land	30	25	83.33	Paddy land
Spinney	44	31	70.45	Forest land
Open woodland	36	33	91.67	Forest land
Forest land	17	13	76.47	Paddy land
Grass land	105	95	90.48	Paddy land
River & channel	58	56	96.55	Lake & pool
Lake & pool	44	41	93.18	River & channel
Building	42	40	95.24	Road
Road	54	43	79.63	Building; Bare land
Bare land	69	57	82.61	Building
Others	28	22	78.57	Lake & pool
Total	600	518	86.33	–

other steps in the AOHC are finished using simple methods to illustrate the effectiveness of the approach. More effective theories and methods can be employed to further improve its performance. For instance, multiple analysis detection (MAD) has been validated as a more effective technology to detect changes (Nielsen 2007). Multi-instance learning is expected as a new framework for object-oriented classification (Zhou et al. 2012). Deep learning recently attracts increasing attention in the field of feature self-learning (Hinton et al. 2006). Semi-supervised learning is devoted to effectively making use of unlabeled samples or unreliable labeled samples (Zhu 2008). Active learning focuses on solving informative samples selection (Burr 2009). Distance metric learning may be helpful in integrating all kinds of features extracted from prior knowledge (Liu and Metaxas 2008). These innovative machine learning theories are worthy of further applying in the

field of land cover updating. Besides, by integrating more domain knowledge of optical remote sensing, the rules of filtering objects in Fig. 2 can be optimized to further improve the reliability of object samples.

Conclusions

To meet the urgent needs of automatic information extraction for optical remote sensing images with high spatial resolution, this paper presents a prior knowledge-based AOHC to rapidly update detailed land cover maps. In the proposed method, MS-based segmentation technology is firstly employed to extract objects. Multiple features, including spectrum, shape and texture features, are then extracted together for supervised

Table 9 Statistical table of itemized accuracy of each detailed land cover type for the AOGC approach (data set B)

Detailed land cover type	Number of artificially interpreted samples	Number of correctly classified samples	Accuracy (%)	Main misclassification
Paddy land	73	60	82.19	Dry land
Dry land	30	25	83.33	Paddy land; bare land
Spinney	44	28	63.64	Forest land
Open woodland	36	31	86.11	Forest land; paddy land
Forest land	17	12	70.59	Paddy land
Grass land	105	91	86.67	Paddy land
River & channel	58	56	96.55	Lake & pool
Lake & pool	44	40	90.91	River & channel
Building	42	37	88.10	Road; bare land
Road	54	42	77.78	Building; bare land
Bare land	69	53	76.81	Building
Others	28	19	67.86	Lake & pool
Total	600	494	82.33	–

pattern classification. With the assistance of historical auxiliary data, a novel adaptive method for object sample automatic collection scheme is designed by combining change detection with transfer learning. In this scheme, prior knowledge from previous interpreted thematic map is transferred to the new target images and then used to rebuild the relationship between landmark classes and their multiple features. Based on the automatic collected samples, a two-level AOHC is carried out on different levels of rough land cover types and detailed land cover types. From the whole process of the methodology, land cover updating tasks can be automatically solved under a supervised learning framework. Based on prior knowledge, all the steps are automatically achieved after extracting and fusing the information from the steps forward.

The results of the experiments show that the approach can automatically obtain reliable object samples with the assistance of auxiliary data. The performance of AOHC in accuracies of rough and detailed land cover classification were also validated. Furthermore, the method did not require any artificial intervention, thereby increasing the degree of automation and applicability of land cover updating. Consequently, the approach creates a novel way for employing the technologies of knowledge discovery into the field of remote sensing information extraction without manual intervention. It is a technical reference for the image automatically understanding using prior knowledge.

References

Barros, R. C., & Basgalupp, M. P. (2012). A Survey of evolutionary algorithms for decision-tree induction. *IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications and Reviews, 42*(3), 291–312.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer.

Briem, G. J., Benediktsson, J. A., & Sveinsson, J. R. (2002). Multiple classifiers applied to multisource remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing, 40*(10), 2291–2299.

Burr, S. (2009). Active learning literature survey. *Computer Sciences Technical Report 1648*, University of Wisconsin Madison.

Carlos, N. S. J., & Alex, A. F. (2011). A survey of hierarchical classification across different application domains. *Data Mining and Knowledge Discovery, 22*(1), 31–72.

Cihlar, J. (2000). Land cover mapping of large areas from satellites: Status and research priorities. *International Journal of Remote Sensing, 21*(6–7), 1093–1114.

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

Dai, W. Y., Chen, Y. Q., Xue, G. R., Yang, Q, & Yu, Y. (2008). Translated learning: transfer learning across different feature spaces. *Processing of 2008 Advances in Neural Information Processing Systems 12 (NIPS 2008)*, Canada, pp. 353–360.

Demir, B., Minello, L., & Bruzzone, L. (2014). Definition of effective training sets for supervised classification of remote sensing images by a novel cost-sensitive active learning method. *IEEE Transactions on Geoscience and Remote Sensing, 52*(2), 1272–1284.

Fardanesh, M. T., & Ersoy, O. K. (1998). Classification accuracy improvement of neural network classifiers by using unlabeled data. *IEEE Transactions on Geoscience and Remote Sensing, 36*(3), 1020–1025.

Foody, G. M., Mcculloch, M. B., & Yates, W. B. (1995). Classification of remote sensing data by an artificial neural network: issues related to training data characteristics. *Photogrammetric Engineering and Remote Sensing, 61*(4), 391–401.

Foody, G. M., & Mathur, A. (2004). A relative evaluation of multiclass image classification by support vector machines. *IEEE Transactions on Geoscience and Remote Sensing, 42*(6), 1336–1343.

Franklin, S. E., & Wulder, M. A. (2002). Remote sensing methods in medium spatial resolution satellite data land cover classification of large areas. *Photogrammetric Engineering and Remote Sensing, 26*(2), 173–205.

Friedl, M. A., Sulla-Menashe, D., Tan, B. A., Schneider, N., Ramankutty, A. S., & Huang, X. (2010). MODIS collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment, 114*(1), 168–182.

Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random forests for land cover classification. *Pattern Recognition Letters, 27*(4), 294–300.

Green, K., Kempka, D., & Lackey, L. (1994). Using remote sensing to detect and monitor land-cover and land-use change. *Photogrammetric Engineering and Remote Sensing, 60*(3), 331–337.

Hinton, G., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation, 18*(7), 1527–1554.

Homer, C. C., Huang, L., Yang, B. W., & Coan, M. (2004). Development of a 2001 national landcover database for the United States. *Photogrammetric Engineering and Remote Sensing, 70*(7), 829–840.

Jerzy, S, & Szymon, W. (2008). Selective pre-processing of imbalanced data for improving classification performance. *In Proceedings of the 10th International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2008)*, Italy, pp. 283–292.

Liu, Q. S., & Metaxas, D. N. (2008). Unifying subspace and distance metric learning with bhattacharyya coefficient for image classification. *Proceedings of the Emerging Trends in Visual Computing (ETVC 2008)*, France, pp. 254–267.

Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, Z., Yang, L., & Merchant, J. W. (2000). Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *International Journal of Remote Sensing, 21*(6), 1303–1330.

Lu, D., Mausel, P., Brondizio, E., & Moran, E. (2004). Change detection techniques. *International Journal of Remote Sensing, 25*(12), 2365–2407.

Lu, D., & Weng, Q. (2007). A Survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing, 28*(5), 823–870.

Luo, J. C., Wang, Q., Ma, J. H., Zhou, C. H., & Liang, Y. (2002). The EM-based maximum likelihood classifier for remote sensing data. *Acta Geodaetica et Cartographica Sinica, 31*(3), 234–239 (in Chinese with English abstract).

Marceau, D. J., Howarth, P. J., Dubois, J. M., & Gratton, D. J. (1990). Evaluation of the grey-level co-occurrence matrix method for land-cover classification using SPOT imagery. *IEEE Transactions on Geoscience and Remote Sensing, 28*(4), 513–519.

Maulik, U., & Chakraborty, D. (2011). A self-trained ensemble with semi-supervised SVM: An application to pixel classification of remote sensing imagery. *Pattern Recognition, 44*(3), 615–623.

Mennis, J., & Guo, D. (2009). Spatial data mining and geographic knowledge discovery: An introduction. *Computers, Environment and Urban Systems, 33*(6), 403–408.

Nielsen, A. A. (2007). The regularized iteratively reweighted MAD method for change detection in multi- and hyperspectral data. *IEEE Transaction on Image Processing, 16*(2), 463–478.

Pal, M., & Mather, P. M. (2003). An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment, 86*(4), 554–565.

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering, 22*(10), 1345–1359.

Richards, J. A., & Jia, X. P. (2012). *Remote sensing digital image analysis: An introduction* (4th ed.). Berlin: Springer.

Togi, T., Khiruddin, B. A., & Lim, H. S. (2013). Comparison of pixel and object based approaches using Landsat data for land use and land cover classification in coastal zone of Medan, Sumatera. *International Journal of Tomography and Simulation, 24*(3), 86–94.

Tsai, F., & Philpot, W. D. (2002). A derivative-aided hyperspectral image analysis system for land-cover classification. *IEEE Transactions on Geoscience and Remote Sensing, 40*(2), 416–425.

Tso, B., & Mather, P. (2009). *Classification methods for remote sensing data* (2nd ed.). New York: Taylor & Francis.

Tuia, D., Ratle, F., Pacifici, F., Kanevski, M. F., & Emery, W. J. (2009). Active learning methods for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing, 47*(7), 2218–2232.

Wardlow, B. D., & Egbert, S. L. (2003). A state-level comparative analysis of the GAP and NLCD land-cover data sets. *Photogrammetric Engineering and Remote Sensing, 69*(12), 1387–1397.

Xing, E. P., Ng, A. Y., Jordan, M. I., & Russell, S. (2002). Distance metric learning with application to clustering with side-information. *Proceedings of the 15th Advances in Neural Information Processing Systems (NIPS 2002)*, USA, pp. 521–528.

Zhou, Z. H., Zhang, M. L., Huang, S. J., & Li, Y. F. (2012). Multi-instance multi-label learning. *Artificial Intelligence, 176*(1), 2291–2320.

Zhu, X. J. (2008). Semi-supervised learning literature survey. *Computer Sciences Technical Report 1530*, University of Wisconsin Madison.