ISPRS Journal of Photogrammetry and Remote Sensing 98 (2014) 133–144

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Domain adaptation for land use classification: A spatio-temporal knowledge reusing method

Yilun Liu, Xia Li *

School of Geography and Planning, Sun Yat-sen University, 135 West Xingang Rd., 510275 Guangzhou, People's Republic of China
Guangdong Key Laboratory for Urbanization and Geo-simulation, Sun Yat-sen University, 135 West Xingang Rd., 510275 Guangzhou, People's Republic of China

ARTICLE INFO

Article history:
Received 8 April 2014
Received in revised form 25 September 2014
Accepted 26 September 2014
Available online 9 November 2014

Keywords:
Domain adaptation
Transfer learning
Land use classification
k-Nearest neighbors
TrAdaBoost
TrCbrBoost

ABSTRACT

Land use classification requires a significant amount of labeled data, which may be difficult and time consuming to obtain. On the other hand, without a sufficient number of training samples, conventional classifiers are unable to produce satisfactory classification results. This paper aims to overcome this issue by proposing a new model, *TrCbrBoost*, which uses old domain data to successfully train a classifier for mapping the land use types of target domain when new labeled data are unavailable. *TrCbrBoost* adopts a fuzzy *CBR* (Case Based Reasoning) model to estimate the land use probabilities for the target (new) domain, which are subsequently used to estimate the classifier performance. Source (old) domain samples are used to train the classifiers of a revised *TrAdaBoost* algorithm in which the weight of each sample is adjusted according to the classifier's performance. This method is tested using time-series SPOT images for land use classification. Our experimental results indicate that *TrCbrBoost* is more effective than traditional classification models, provided that sufficient amount of old domain data is available. Under these conditions, the proposed method is 9.19% more accurate.

1. Introduction

Land use classification using remote sensing data can be based on supervised learning methods, which typically require a set of new field investigation samples for new study areas (Li et al., 2014; Weng, 2012). However, collecting extensive amount of labeled data is extremely expensive and time consuming, and in some cases may be impossible to perform (Jansen and Gregorio, 2002; Robinove, 1981). In order to overcome the issue of the lack of labeled data, numerous methods have been proposed, such as unsupervised learning (Hegarat-Mascle et al., 1997) and semi-supervised learning (Bennett and Demiriz, 1999; Chapelle et al., 2002). While these approaches aim to reduce the quantity of training labeled data required, they are not as effective as the supervised methods. If the supervised methods are preferred, the cost for collecting samples will be reduced if previously collected labeled data can be reused to classify a new image. This strategy assumes that old labeled data, even if obsolete, can provide useful information for a new classification task. In many situations, old data remain beneficial for training a new classifier (Dai et al., 2007). While the process of evaluating and selecting useful old data can be problematic, this issue can be overcome by using transfer learning methods, which can be employed to transfer the knowledge learned by the classification model from an old domain to a new one.

Traditional classification methods assume that the training and test data sets are drawn from the same distribution (Pan and Yang, 2010). Thus, any distribution changes require most classification models to be rebuilt using newly collected training data. Transfer learning methods that transfer knowledge learned by performing one or more source tasks to the target task can be employed to address this issue (Pan and Yang, 2010; Torrey and Shavlik, 2009).

Since 2005, transfer learning has become an increasingly important topic in computer sciences (Rosenstein et al., 2005). According to the extant literature, transfer learning is a powerful tool that can be applied to many fields (Pan and Yang, 2010). Several simple and effective algorithms have been employed to modify traditional machine learning methods to suit various domain problems. For example, Dai et al. (2007) proposed a boosting transfer learning method (*TrAdaBoost*), which utilizes old data, in combination with a small amount of new data, to train an ensemble classifier. More recently, Pan et al. (2009) developed an unsupervised transfer

* Corresponding author at: Yat-sen Chair Professor, School of Geography and Planning, Sun Yat-sen University, 135 West Xingang Rd., 510275 Guangzhou, People's Republic of China. Tel.: +86 13924203023; fax: +86 (020) 84115833.
E-mail addresses: lixia@mail.sysu.edu.cn, lixia@graduate.hku.hk (X. Li).
URL: http://www.geosimulation.cn/ (X. Li).

http://dx.doi.org/10.1016/j.isprsjprs.2014.09.013
0924-2716/

learning method (Transfer Component Analysis or *TCA*) in which a dimensionality reduction framework for transfer learning is used to compensate for absence of target domain knowledge. Transfer learning has been successfully applied in various fields to solve numerous real-world classification or regression problems (Pan and Yang, 2010). However, thus far, the use of transfer learning techniques for land use classification has received limited attention from researchers and industry practitioners. One exception is the work of Rajan et al. (2008), who examined the knowledge transfer problem of hyperspectral data classification using an active learning method. By adding new samples to existing classifiers, in this study, the classifiers are corrected to fit the classification of target images. Similarly, Matasci et al. (2011) used *TCA* to solve the knowledge transfer problem for classifying hyperspectral images. However, in both studies, only transfer learning algorithms were employed to handle spectral properties, thus failing to consider the spatio-temporal knowledge of the study areas.

This paper aims to overcome the aforementioned shortcomings by proposing a novel transfer learning method named Transfer Case Based Reasoning Boosting (*TrCbrBoost*). This method is based on the integration of fuzzy case-based reasoning (fuzzy *CBR*) and ensemble learning methods. The objective of this method is to use the samples of multi-temporal images to train a classifier for a target image. The multi-temporal data are analyzed to identity both stable and unstable feature distributions. These features are subsequently weighted in accordance with the divergence between distributions pertaining to different domains, before being employed in the fuzzy *CBR* model to produce land use probability maps. This allows the land use probability maps to be used as a constraint to adjust the sample weight of each boosting iteration. The increased number of boosting iterations can thus reduce the distribution divergence between the training data and the target domain data. The final classification is obtained by majority vote of the last half of boosting classifiers.

2. Methodology

Traditional classification methods usually classify land use types by employing machine learning models that are applied to the collected data (which can be labeled or unlabeled) typically obtained from the same image, defined as domain D. In the same vein, the classification task can be defined as task T. A domain D consists of two components, namely, feature space X and marginal probability distribution $P(X)$, where $X = \{x_1, \ldots, x_n\} \in X$. For example, if our task is the classification of SPOT images with four bands, X is the space of all four bands and x_i is the i_{th} band of the image.

Given a specific domain $D = \{X, P(X)\}$, task T also consists of two components, namely label space Y and predictive objective function $f(\cdot)$ (denoted by $T = \{Y, f(\cdot)\}$). These components are derived from the training data, where $x_i \in X$ and $y_i \in Y$. Thus, function $f(\cdot)$ can be used to predict land use type.

When the task is land use classification in a different scene (pertaining to a different region or taken at a different time) remote sensing images, this task can be considered as a different domain task for several reasons. First, numerous studies have shown that the atmospheric, surface variation, solar incident angle, and other factors the effects of which cannot be predicted with certainty, cause discrepancies in the spectral and texture data distributions

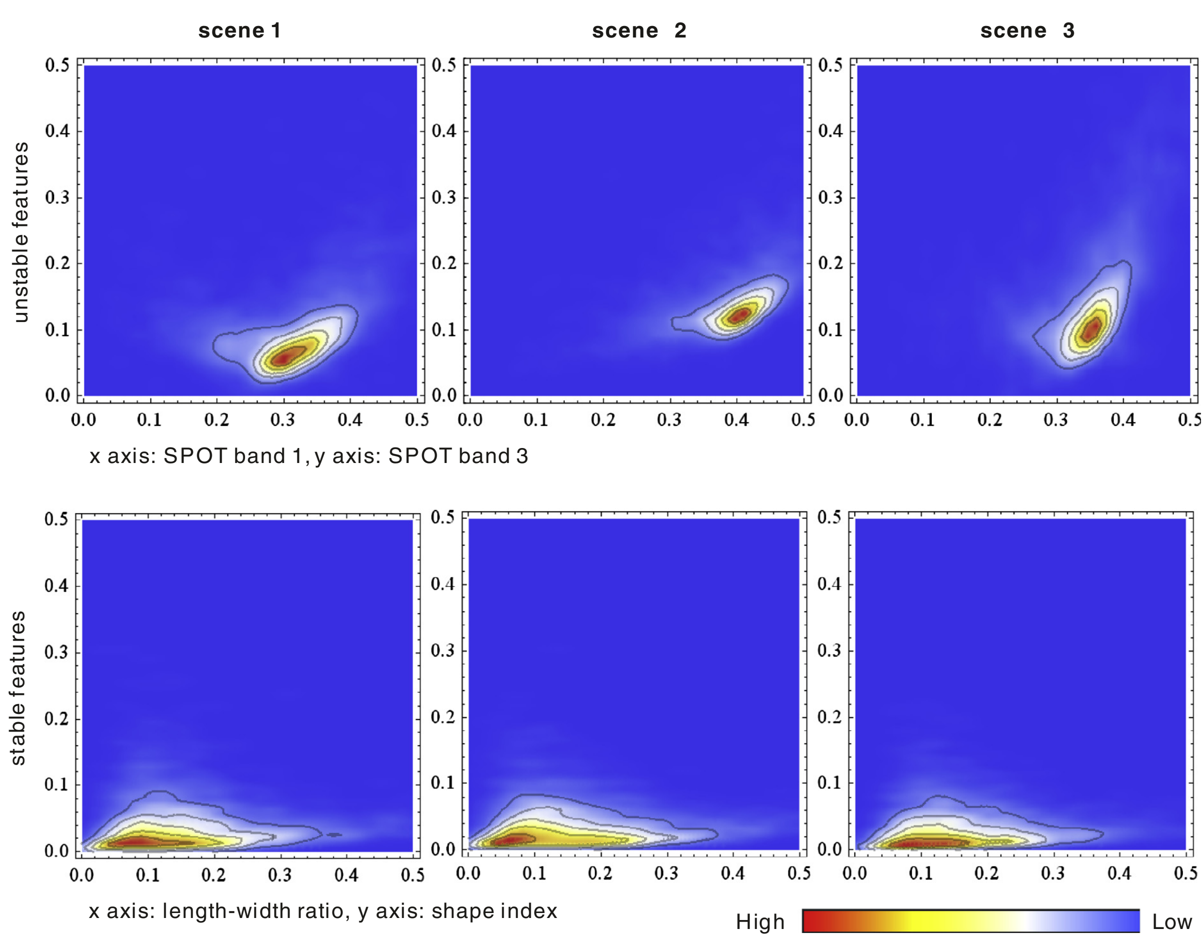

Fig. 1. The sample density pertaining to the agricultural land use category in different scene remote sensing images.

of discrete scene images, even when these are obtained by the same sensor (Benz et al., 2004; Wicks et al., 2002). Second, the spectral and texture feature distributions of the same land use type may be changed in different scene images due to instability in the land use objects' character. For example, the images taken in different years may not have been acquired during the same time of the crop growth cycles, or the soil moisture has changed across time (Li and Yeh, 2004). Third, the geometric characteristic of the land parcel may be changed because of human intervention. For example, road network development will cause landscape fragmentation, whereby a large plot of land may be divided into several smaller sections. These situations can result in discrepant land use classification problems, which present different tasks and belong to different domains.

The feature distributions for the same land use type can also vary in different scene images. This is exemplified in Fig. 1 showing the sample density of agricultural land use for three different scene images. The *x* and *y* axes in Fig. 1a correspond to SPOT band 2 and band 3, while the *x* and *y* axes in Fig. 1b denote length–width ratio and shape index of the land use patch, respectively. When features (in this case, the length–width ratio and shape index) have similar distribution in different scenes, they are considered as stable features. Thus, they can be used directly for training the new classifier for a different scene image. In contrast, features that have discrepant distributions (in this case, the SPOT band 2 and band 3) are deemed unstable. However, as shown in Fig. 1, even the discrepant distributions of unstable feature from different scenes can have some overlaps. Thus, capitalizing on this characteristic, the objective is to use the old samples that are within this overlapping part to train the classifier according to the proposed knowledge transfer method.

The knowledge transfer process involves a target domain image, which needs to be classified, and several source domain images that have already been classified. Moreover, both images must correspond to the same region, and are typically taken at different points in time. In the approach described in this paper, the labeled data from the source domains and the unlabeled data from the target domain are required to classify the target domain image. We propose a paradigm called *TrCbrBoost* to solve this problem.

In our method, remote sensing images are first segmented into land parcels (objects) that can be used in the classification of land use types. This enables the features (e.g., spectral and texture information) of each land parcel to be extracted for subsequent analysis. It should be noted that the classification model cannot be trained using old data directly. Moreover, the same cannot be directly input into *CBR* model in order to retrieve an accurate land use map. These limitations are imposed by feature distributions, which are unstable in different domains for remote sensing imagery. Therefore, the divergence of feature distribution is analyzed, allowing the stable features to be assigned a greater weight than the unstable features. As shown in Fig. 2, these weighted features serve as input into a fuzzy *CBR*. In fact, the *CBR* method yields the probabilities for the target domain land use. The land use probabilities of target domain obtained in this manner are reliable because the higher-weight (stable) features contribute most of the information for the *CBR* retrieval process. These probabilities will also be used as criterion in the following modified *TrAdaBoost* algorithm. In each iteration of *TrAdaBoost*, the sample weights are adjusted according to the base classifier performance. Thus, the distribution divergence between the selected training data set and the target domain data set is reduced with each successive iteration.

This proposed *TrCbrBoost* method consists of the following three major components (Fig. 3):

(a) *Feature selection and divergence analysis:* In this study, four types of features (spectral, texture, geometrical characteristic and spatio-temporal relationship information) are extracted from the remote sensing images. The feature divergence between different source domains is analyzed and measured. The features are then weighted (feature weight v) according to the divergence.
(b) *Case-based reasoning for land use probabilities:* These features will be imported into a *CBR* model. Features and labels pertaining to source domains will be represented as a case library. This enables the use of fuzzy *k-NN* method to predict the land use possibilities of the target domain according to the information stored in this case library.
(c) *Boosting for transfer learning:* The source domain features are imported into a modified *TrAdaBoost* algorithm. *TrAdaBoost* is an ensemble learning method that generates a diverse ensemble of classifiers by manipulating the training data

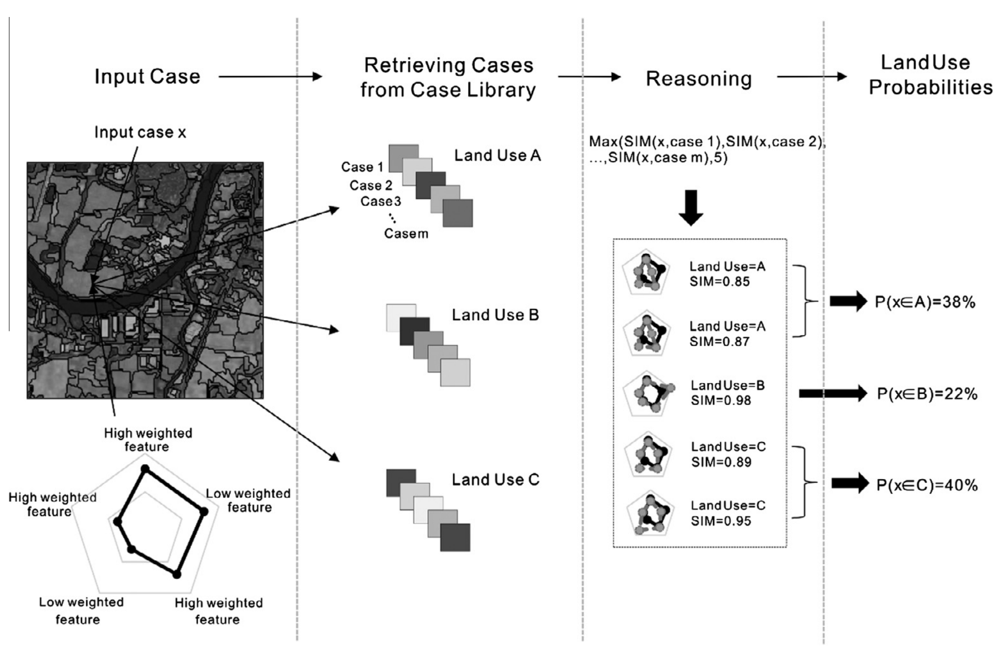

Fig. 2. Calculation of land use probabilities using fuzzy CBR.

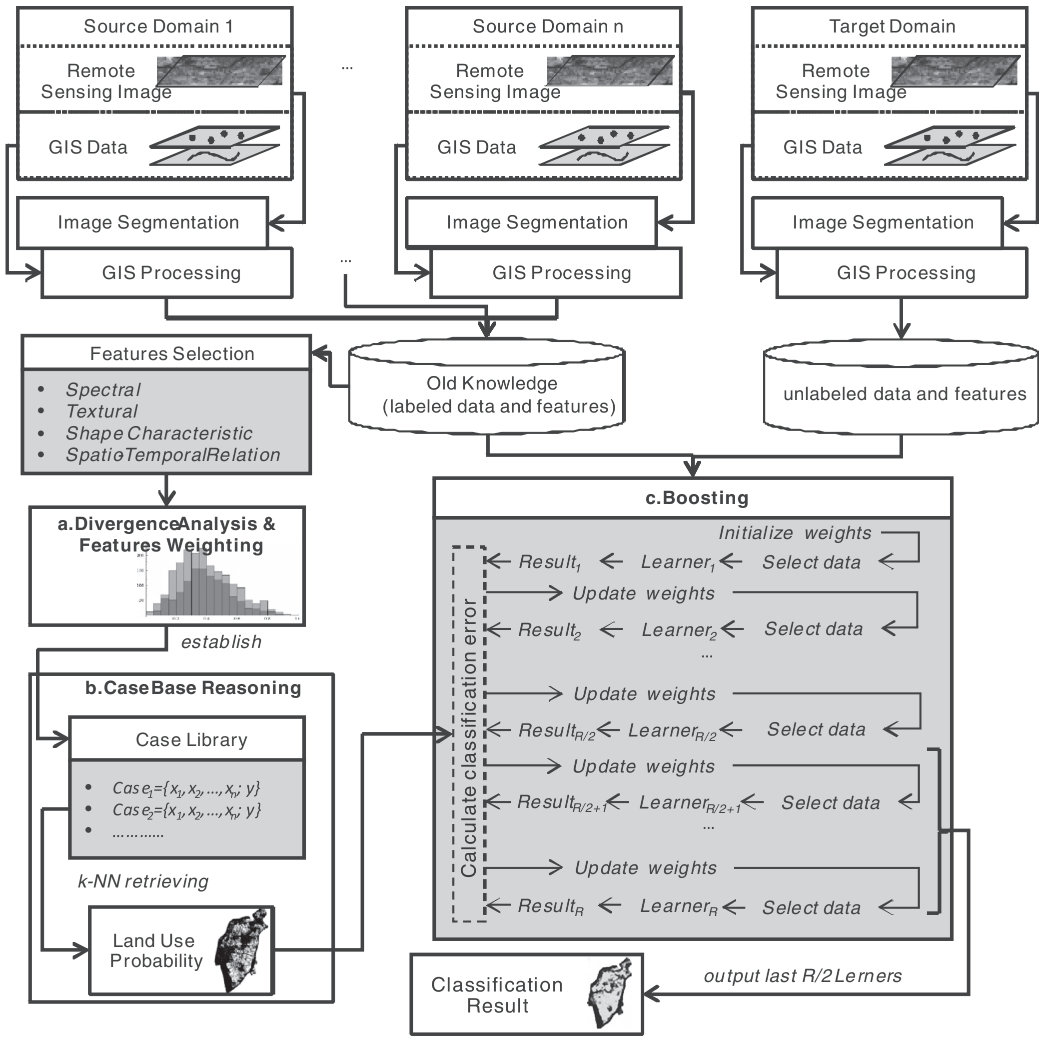

Fig. 3. *TrCbrBoost* method flowchart.

provided to a "base" learning algorithm. Here, the land use possibilities produced by *CBR* are used to evaluate each classifier's performance. The weights (sample weight *w*) of the source domain samples are adjusted in each iteration, according to the error associated with this classifier. Thus, with each subsequent iteration, the source domain samples that are assigned greater weights should provide a better fit to the target domain distribution. Finally, the classification results are obtained from the model trained using these higher-weight samples.

2.1. Feature selection and divergence analysis

2.1.1. Feature selection

Four types of features are selected for *TrCbrBoost*, namely spectral, texture, geometrical characteristic and spatio-temporal relationship information. According to the extant literature, spectral information is considered the most important feature in image classification (Awrangjeb et al., 2010; Robinove, 1981). It can, however, exhibit a profound divergence among different domains because of the uncertainties caused by atmospheric, solar incident angle, and other effects on satellite data.

Texture is another important characteristic of remote sensing data, as it is required by many classification or detection algorithms (Awrangjeb et al., 2010, 2012). The texture distribution is affected by similar issues to those pertaining to spectral features and thus may also be different for different domains.

The third feature type involves the geometrical characteristics of a land use object, such as perimeter, length–width ratio, and other shape indices. Previous studies indicate that the shape information can be employed effectively to improve classification performance (Xia, 1996). However, for this process to be successful, the shape distribution of different land segments should be similar, even if these correspond to different regions or are taken at different points in time. Here, it is worth noting some extreme cases, such as habitat fragmentation, which will split up the continuous land parcels.

The final feature type of interest for this work is spatio-temporal relationship. The first law of geography states that features or objects that are spatially closer are more highly related (Tobler, 1965). Authors of numerous studies in this field reported that land use change is affected by spatial relationship, such as proximity factors, topological relationship, and surrounding land use types (Du et al., 2012; Li and Yeh, 2000). Extant studies also indicate presence of temporal correlations between land use types at different periods (Lu and Weng, 2007). For example, the probability that a plot of agricultural land located next to main roads would be converted into built-up land is relatively high. Conversely, probability of converting a built-up area into forestlands is virtually zero.

2.1.2. Distribution divergence analysis

In the analysis, the features should be weighted according to the distribution divergence between the different domains, as this allows the land use probabilities to be produced using the fuzzy *CBR* model. Here, features characterized by lower divergence will be assigned higher weights because, due to their greater stability and thus greater value for classification.

The Kullback–Leibler divergence (*KL*) (Kullback and Leibler, 1951) is a well-known approach to measuring the distribution similarity among different source domains. *KL* is a non-symmetric measure of the closeness of the two probability distributions, *P* and *Q*. The *KL* for a discrete distribution is defined as:

$$KL(P\|Q) = \sum_{x=1} P(x) \log \frac{P(x)}{Q(x)} \tag{1}$$

Thus, the closer the two distributions are to one another, the smaller the *KL*. Moreover, as *KL* is non-symmetric, the divergence D_{KL} of distributions *P* and *Q* can be expressed as:

$$D_{KL} = \frac{1}{2} KL(P\|Q) + \frac{1}{2} KL(Q\|P) \tag{2}$$

2.2. CBR for land use probabilities

CBR is an artificial intelligence technique, which relies on knowledge from previous experiences (e.g., old cases) to solve a new problem (Du et al., 2002). This method has been widely adopted with the aim to classify or predict land use. However, as traditional *CBR* is an intra-domain analogy method (Aamodt and Plaza, 1994), the knowledge on old cases and new problems must be in the same domain and fit the same distribution. For this reason, when the feature distribution is unstable, *CBR* is unreliable (Li and Yeh, 2004). In order to avoid this defect, in this work, we modify the traditional *CBR* method by weighting the features according to their divergence, before using them as input into the fuzzy *CRB* method to produce the map of land use probabilities for the target domain.

2.2.1. Case representation

The case representation and of case library construction are the main components of the *CBR* approach. The basic unit of a case is the feature resulting from features selection and distribution divergence analysis. A case consists of two components, namely problem description (imported features, e.g., spectral, textual, geometrical characteristic or other relationship information) and solution (land use type).

A land use case related to the land parcel (object) can be defined using the following expression:

$$Case_i = \{id, x_1^i, x_2^i, \ldots, x_n^i;\ y^i\} \tag{3}$$

where x_j^i is input feature *j* of land object *i*, and y_i is the land use type of the source domain object *i*.

2.2.2. Case retrieving algorithm

After establishing the case library, case retrieval is conducted to estimate the probability of each land use type for each unknown case. The retrieval is performed according to the similarity between an unknown case (unlabeled target domain samples) and known cases (labeled source domain samples) in the case library.

The *k-Nearest Neighbors* (*k-NN*) algorithm is widely adopted for *CBR* retrieval process. However, when a traditional *k-NN* is used, the land use type of each new queried case is determined by the cumulative similarity of its *k* nearest neighbors. Consequently, the case will be assigned to the major land use type among these neighbors (Dasarathy, 1990). *CBR* is used in this study to calculate the probability of each land use of an unknown case. Thus, we adopt the fuzzy *k-NN* algorithm proposed by Keller et al. (1985) to calculate the probabilities. For the i^{th} case, the probability of a land use type is calculated using the following equation:

$$\Phi(i,s) \leftarrow \sum_{j}^{k} w_{ij} \cdot \delta(s, \eta(j)) \begin{Bmatrix} \delta(s,\eta(j)) = 1, & if\ s = \eta(j) \\ \delta(s,\eta(j)) = 0, & if\ s \neq \eta(j) \end{Bmatrix} \tag{4}$$

where $\eta(j)$ is a target function of known case (indicating its land use type), *k* is the total number of nearest neighbors, *s* is the finite set of target class values, and w_{ij} is the feature-distance weight proportional to the inverse cumulative similarity:

$$w_{ij} = \frac{1}{SIM(i,j)^2} \tag{5}$$

The similarity (distance) between objects *i* and *j* is based on the weighted Euclidian distance given below:

$$SIM(i,j) = \sqrt{\sum_{n=1}^{N} v_{cn} (a_{in} - a_{jn})^2} \tag{6}$$

$$v_{cn} = 1 - \frac{KL_{cn} - min\left(\sum_{n=1}^{N} KL_{cn}\right)}{max\left(\sum_{n=1}^{N} KL_{cn}\right) - min\left(\sum_{n=1}^{N} KL_{cn}\right)} \tag{7}$$

where v is the feature weight, *a* is the normalized feature value ($a \in (0,1)$), *c* is the land use class, and *N* is the total number of features.

Once an unknown case is matched with *k* known cases, the probabilities of the target domain land use are calculated using Eq. (4). Here, the probability of a land use type for each case should be normalized using the following equation:

$$\overline{\Phi}(i,s) = \frac{\Phi(i,s)}{\sum_{j=1}^{S} \Phi(i,j)} \tag{8}$$

After normalization, the sum of probabilities pertaining to all land use types for each case is equal to 1. This enables the map of land use probabilities of the target domain to be obtained based on the previously discussed *CBR* reasoning. The next step is to input these probability maps into a modified *TrAdaBoost* algorithm, described in the following section.

2.3. Boosting for transfer learning

Boosting is a machine learning algorithm, which improves the accuracy of a base learner (an ordinary classification algorithm, such as *Decision Tree* or *SVM*) by adjusting the weights assigned to the training data and allowing the classifier to learn accordingly (Schapire, 1999). However, similar to most traditional learning methods, boosting assumes that the training and the test data follow the same distribution. Dai et al. (2007) attempted to overcome the issue of different distributions by developing the so-called *TrAdaBoost* algorithm, which is based on *boosting*. In our study, a revised *TrAdaBoost* algorithm is proposed, and is applied to the labeled data collected in source domains for domain adaptation of land use classification.

Fig. 3c illustrates this transfer learning procedure. The detailed methodology for *TrAdaBoost* is described below.

Step 1: Preparing the inputs.

In this step, the input data is prepared and the maximum number of iterations (*R*) for the *TrAdaBoost* algorithm defined. The labeled data from the source domain provide the empirical information for training the base learner. The maximum number of iterations that defines the set (ensemble) of base learners is determined according to the error rate iteration curve (Dai et al., 2007). Here, the improvement of prediction accuracy is assumed to stabilize after a certain number of iterations.

The labeled data are usually obtained by classifying remote sensing images or carrying out field investigations. The labeled data set $S = \{(x_1^S, y_1^S), \ldots, (x_m^S, y_m^S)\}$ (where m is the number of old data points) is collected from the source domains. In each labeled data set (sample), the x and y variables represent the site (object) features at a location and its land use type, respectively. The unlabeled data set $T = \{x_1^T, x_2^T, \ldots, x_n^T\}$ (where n is the number of new data) is collected from the target domain.

Step 2: Initializing the weights for the labeled data.

The weights for the labeled data are initialized before calculating the weight decay factor. While it is assumed that all available data can be used for training the classifier, the contribution of each sample to the classification corresponds to the weight assigned to it.

At the beginning, all weights are assigned equal value, defined as follows:

$$w_i^1 = \frac{1}{m} \quad \textit{for } i = 1, \ldots, m \tag{9}$$

Step 3: For $r = 1, \ldots, R$, the base learner is run, while adjusting the weight of each sample according to its classification performance.

TrAdaBoost consists of a number of base learners that are generated using different combinations of labeled data. In each iteration, a part of the labeled data set is selected to train a base learner according to the sample weight.

First, the weight of each sample is normalized using the following equation:

$$w_i^r = w_i^r \Bigg/ \sum_{i=1}^{m} w_i^r \tag{10}$$

Only those samples whose weights are greater than a dynamic threshold defined below will be selected to train a base learner f_r.

$$\alpha_r = mean(w_1^r, \ldots, w_m^r) \cdot \gamma, \quad 0 \leqslant \gamma \leqslant 1 \tag{11}$$

where γ is a random variable and r is the current number of iterations.

The f_r is applied on the unlabeled data set T to obtain the classification land use results. This allows estimating the classifier error of this base learner f_r by comparing the previously obtained classification output with the probability map, as shown below:

$$\varepsilon_r = \sum_{i=1}^{n} \sum_{s=1}^{Y} \overline{\Phi}(i,s)\left(1 - \delta(s, f_r(x_i^T))\right)/n \tag{12}$$

where ε_r is the model error of f_r, $\overline{\Phi}(i,s)$ is the probability of land use type s for case i, $\delta(s, f_r(x_i^T))$ is a sign function, and $\delta(s, f_r(x_i^T)) = 0$ when $s = f_r(x_i^T)$ and $\delta(s, f_r(x_i^T)) = 1$ when $s \neq f_r(x_i^T)$.

The weight decay factors pertaining to s are defined as follows:

$$\beta_r = \varepsilon_r/(2\varepsilon_1 - \varepsilon_r) \text{ and } \beta = 1 \Bigg/ \left(1 + \sqrt{2 \ln m/R}\right) \tag{13}$$

The weights are dynamically updated according to the following equation:

$$w_i^{r+1} = \begin{cases} w_i^r \cdot \beta_r^{-1} & \textit{if } f_r(x_i^S) = y_i^S \\ w_i^r \cdot \beta & \textit{if } f_r(x_i^S) \neq y_i^S \end{cases} \tag{14}$$

We can consider f_r as a satisfactory classifier when the error associated with it is less than that pertaining to the initial classifier f_1 ($\varepsilon_t < \varepsilon_1$), as this implies that a correctly predicted source domain sample is more similar to the target domain data. Therefore, the weight of this sample will be increased ($\beta_{r-1} > 1$), enhancing its effect on the next iteration training. Irrespective of the accuracy of f_r, the weight of the incorrectly predicted source domain data will be decreased evenly by multiplying their individual weights by β ($\beta \in (0,1]$). Thus, as the source domain data fit the target domain distribution better, their weights will be greater after several iterations. In addition, the data dissimilar to those in the target domain will have lower weights. Thus, only the data assigned significant weights will be used to assist the learner in training better performing classifiers in the subsequent iteration.

Step 4: Generating the final classification result according to the ensembles of base learners.

The weight adjusting strategy described above decreased the classification error with each subsequent iteration. Consequently, the classifiers built at later iterations will be better than their precedents in terms of classification accuracy (Dai et al., 2007). Finally, the classifiers built in the last $R/2$ iterations will be used to obtain the final result, based on their respective errors. Here, the following equation is used to estimate the land use type for an unknown case:

$$\arg\max_{y \in Y} \left(\sum_{r=R/2}^{R} \log \frac{1}{\beta_r} \right) \tag{15}$$

3. Case study

3.1. Study area and data set

The middle part of the Pearl River Delta in China (latitude 23°02′N, longitude 113°32′E) is chosen as the study area (Fig. 4). This location is in the east estuary of the Pearl River, covering an area of approximately 91 km^2. Before the economic reform that took place in 1978, a large part of the study area was dedicated to agricultural activities. However, a significant amount of agricultural land was converted to residential and industrial land use in the last three decades.

The labeled data of the source domains were obtained from the classification of SPOT-5 HRG images (10 m spatial resolution). These images were acquired on November 25, 2004, October 29, 2006, and November 10, 2008 from Scene No. 285-304 found in the China Remote Sensing Ground Station reference system. The image of the target domain, which needs to be classified, was acquired on November 09, 2010. Absolute atmospheric correction of the images was not performed because simultaneously acquired ground based spectral data or appropriate meteorological data was lacking for the study area. Thus, the images were radiometrically corrected instead, using the dark object subtract tool of ENVI. These images were rectified to UTM zone 49, WGS1984 according to ground-control points (Fig. 5). The root mean square errors were less than 0.5 pixels for each image. These images were finally transformed using the projection of Transverse Mercator.

To assess the classification accuracy of the target domain, a field investigation in the study area was undertaken shortly before acquiring the satellite image. The field measurement is based on a Continuous Operational Reference System (CORS), which uses local reference stations to provide up to centimeter-level accuracy (Li et al., 2013). The high-precision handheld GPS obtains detailed geometric information for each confirmed site of *in situ* data. As shown in Fig. 5, we collected 380 *in situ* data points, as well as recorded the coordinates of each *in situ* location and its related land use type using CORS.

The classification of the source domain images was conducted using an object-based supervised machine learning method. The classification scheme comprised eight classes, namely agriculture, lawn/grass, built-up area, transportation, bare land, orchards, fishpond, and water (Table 1). We first step is to identified the image

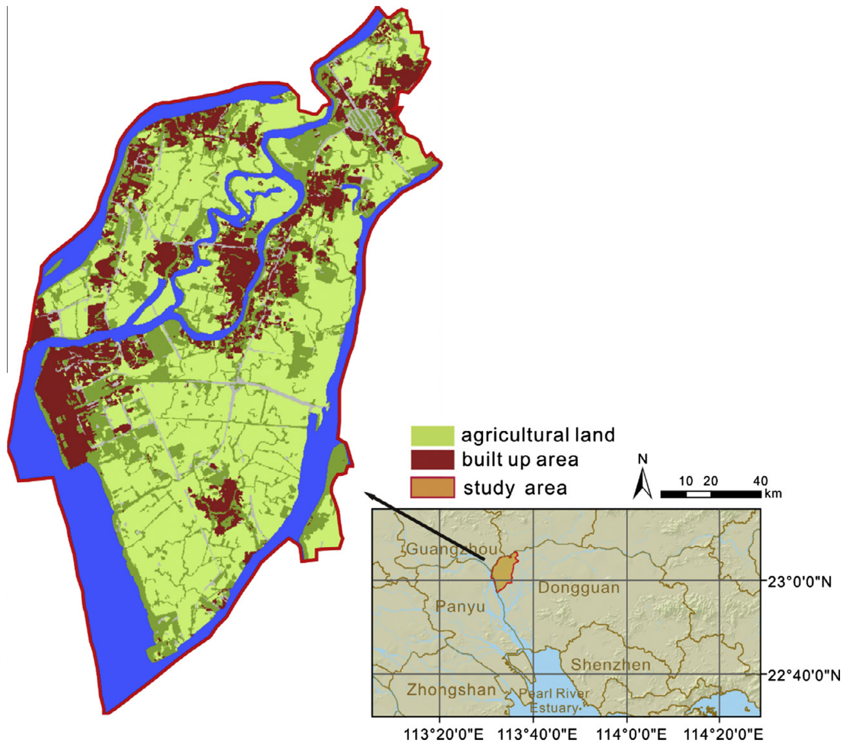

Fig. 4. Study area location in the Pearl River Delta.

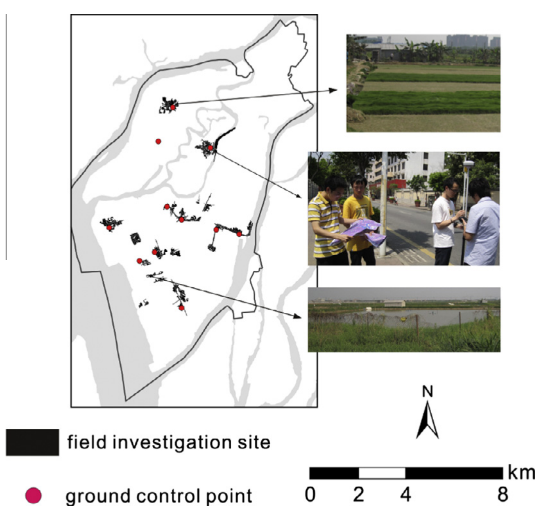

Fig. 5. Field investigation sites and ground control points.

Table 1
Land use classification scheme.

Land use types	Description
Built-up area (BU)	Residential, commercial and industrial, mixed urban or build-up land
Transportation (TL)	Highway, road and railway
Bare land (BL)	Sand, gravel and bare soil
Water (RI)	Permanent open water, lake, river and wetland
Fishpond (FP)	Freshwater pond with fish
Orchard (OC)	Garden consisting of a small cultivated wood and fruits area
Lawn/grass (LG)	Golf courses, lawns and sod fields
Agricultural land (AL)	Crop fields, pasture and fallow field

objects, whereby image segmentation was implemented within eCognition 8.7 software. The land use map was produced according to the spectral signature, texture features, and geometric characteristics of the labeled objects. Based on field checking, we achieved 85–90% classification accuracy.

3.2. Feature selection and analysis

Eleven features were chosen and extracted from each object, including the four major types (spectral, texture, geometrical and spatial relationship) mentioned in Section 2.1:

- Four features are related to the spectral signature of SPOT, including mean values of band 1 (B_1), band 2 (B_2), band 3 (B_3), and band 4 (B_4). These features are the basic information provided by satellite data and have been used in several previous studies (Gong et al., 1992).
- One feature is related to the texture, which refers to the mean value of gray-level co-occurrence matrix (*GLCM*), extracted from satellite data. Extant literature reports (Marceau et al., 1990; Zhang et al., 2003) indicate that *GLCM* is helpful in improving the accuracy of SPOT data classification.
- Three features are related to geometrical characteristics, which include object area (*A*), length–width ratio (*LWR*), and shape index (*SI*).
- Three features are related to spatial relationship and include distance to the nearest existing main road (D_1), distance to the nearest water (D_2), and distance to the nearest built-up land (D_3). These proximity factors significantly affect potential land use change (Li and Yeh, 2002; Shrestha and Zinck, 2001).

The average *KL-divergence* of the distribution of these features among the three source domains was calculated using Wolfram Mathematica 8.0. As can be seen in Table 2, the *KL* value of the spectral and texture features exceeds those of the shape and spatial relationship features.

Table 2
Average *KL*-divergence of features.

	B1	B2	B3	B4	GLCM	LWR	A	SI	D1	D2	D3
BL	0.55	1.01	0.06	0.59	0.42	0.00	0.01	0.00	1.16	0.32	0.03
RI	0.16	5.82	0.51	0.27	1.05	0.02	0.01	0.01	0.59	0.20	0.09
FP	0.41	4.84	0.46	0.53	1.09	0.10	0.14	0.01	0.83	0.15	0.04
OC	1.71	3.11	0.11	1.46	1.28	0.52	0.07	0.03	0.84	0.10	0.04
LG	0.69	3.86	0.25	0.47	1.09	0.03	0.02	0.01	0.70	0.12	0.17
TL	0.57	1.26	0.07	0.58	0.49	0.00	0.00	0.00	0.93	0.16	0.03
BU	0.85	1.71	0.13	0.66	0.76	0.03	0.02	0.01	0.42	0.08	0.02
AL	1.19	2.53	0.61	1.38	2.49	0.00	0.01	0.00	1.63	0.07	0.00

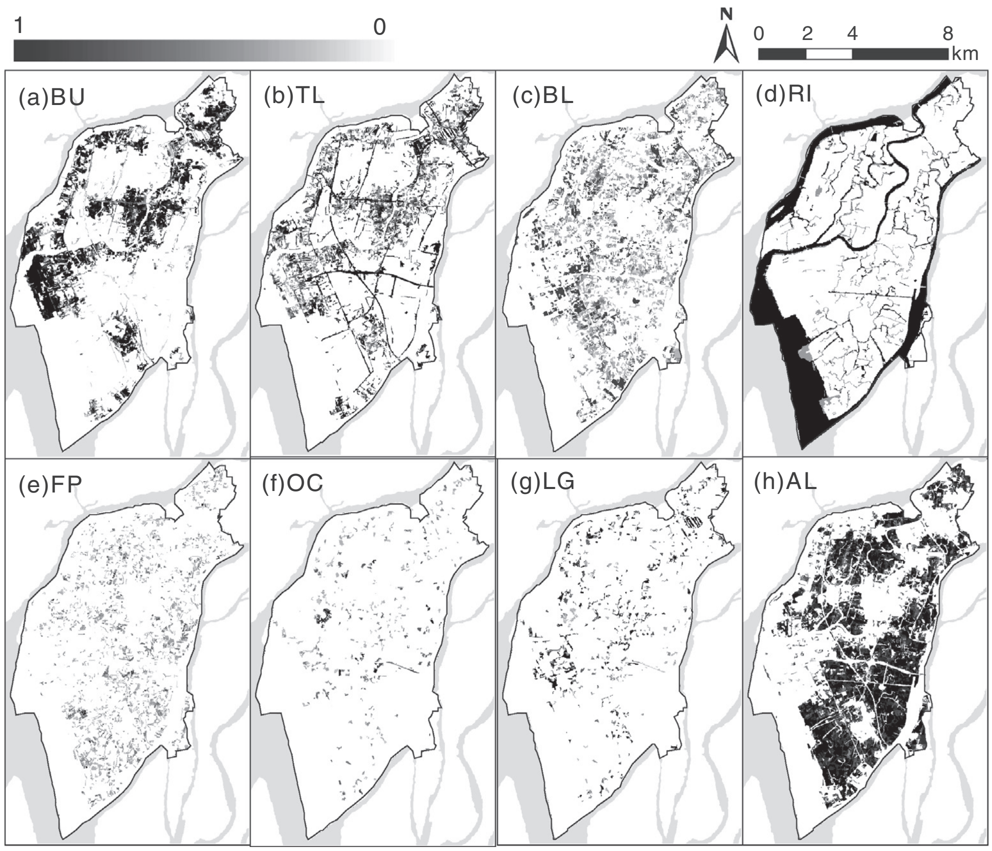

Fig. 6. Land use probability maps produced by fuzzy CBR.

3.3. CRB for land use probabilities

The features described above were imported into the fuzzy *CBR* model, which is implemented using Matlab 7.0 fuzzy logic tool box, to produce land use probabilities of the target domain. In addition to these 11 features, the *CBR* method also requires the temporal relationship information. In this work, three periods of land use are treated as three source domains. The time intervals between adjacent time domains (including source and target domain) are nearly equal (2 years). The land use type of the former period (LU_{former}) is considered as a feature of this period. For example, LU_{former} of the target domain (year 2010) is represented by the land use type information pertaining to 2008. Based on Eq. (3), a case can be represented using the following expression:

$$Case = \{id, B_1, B_2, B_3, B_4, GLCM, A, LWR, SI, D_1, D_2, D_3, LU_{former};\ LU\} \tag{16}$$

As LU_{former} is a discrete value, distance calculation is obtained using the following expression:

Table 3
Average *KL*-divergence between training and reference data sets.

Iteration	1	5	20	100	Divergence reduction (%)
B1	1.47	1.41	1.23	1.06	27.89
B2	0.26	0.24	0.20	0.18	30.77
B3	0.51	0.42	0.31	0.23	54.90
B4	1.12	1.04	0.96	0.86	23.21
GLCM	1.44	1.26	1.09	1.05	27.08

$$LU^{unknown}_{former} - LU^{known}_{former} = \begin{cases} 0 \text{ if } LU^{unknown}_{former} = LU^{known}_{former} \text{ or } LU^{known} \\ 1 \text{ if } LU^{unknown}_{former} \neq LU^{known}_{former} \text{ or } LU^{known} \end{cases} \tag{17}$$

where $LU^{unknown}_{former}$ is the former land use type of the unknown case, LU^{known}_{former} is the former land use type of the known case, and LU^{known} is the current land use type of the known case.

A case library comprising 31,921 cases of source domains was established by using the classified remote sensing images. Similar old cases were retrieved using the *k-NN* method for the 16,933

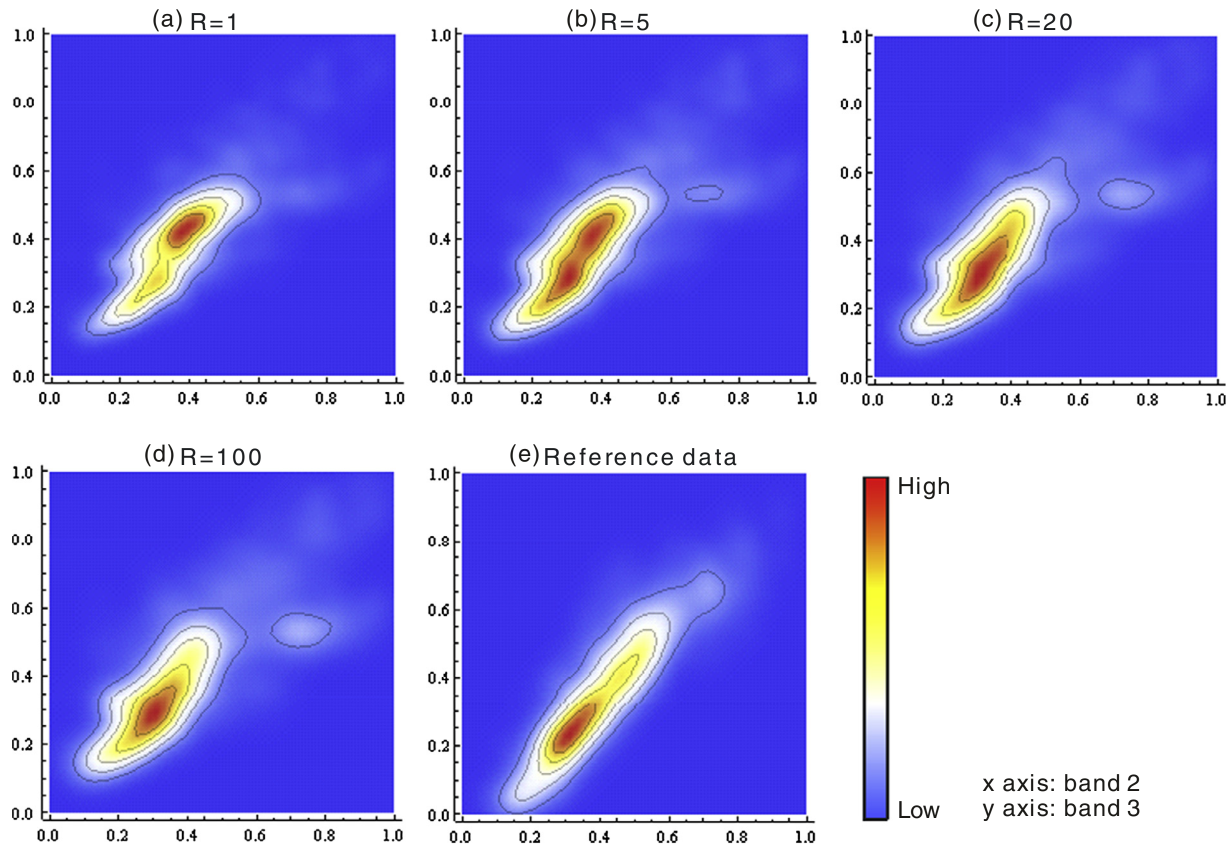

Fig. 7. Density of training samples for the built-up area category in band 2 and band 3.

Table 4
Error matrix of the SVM_{DS} for land use classification.

Classified data (pixel)	Reference data								UA (%)
	BL	RI	FP	OC	LG	TL	BU	AL	
BL	300	0	0	0	0	2	5	0	97.72
RI	0	184	6	0	0	3	0	0	95.34
FP	1	6	86	0	0	1	1	6	85.15
OC	1	0	0	22	86	2	0	0	19.82
LG	3	0	0	8	215	1	1	0	94.30
TL	10	5	0	0	8	232	268	5	43.94
BU	26	1	0	1	2	64	258	8	71.67
AL	2	5	69	0	3	8	2	497	84.81
Total	343	201	161	31	314	313	535	516	
PA (%)	87.46	91.54	53.42	70.97	68.47	74.12	48.22	96.32	

UA = the user's accuracy, PA = the producer's accuracy.
Overall classification accuracy = 74.32%.
Overall kappa statistics = 0.69.

cases of the target domain. The k of k-NN was set to 10 and the probability of each unknown case was calculated using Eq. (4). This enables generating the land use probability of each land use type (as shown in Fig. 6). Finally, the probability map was used to assess the performance of each boosting classifier.

3.4. Transfer boosting for land use classification

In this study, Support Vector Machines (*SVM*) were used to develop the basis Learner for *TrCbrBoost*. We used *Libsvm* (Chang and Lin, 2011), for which the source code is available at http://www.csie.ntu.edu.tw/~cjlin/libsvm, while we implemented the *SVM* classifier using *RBF* kernel. We employed a 10-fold cross-validation strategy to select the two *SVM* free parameters (the penalization parameter and variance of the considered *RBF* kernels). The number of iterations R was set to 100. As noted above, 31,921 labeled data pertaining to the source domains and 16,933 unlabeled data representing the target domain with 11 features and the temporal relationship were selected and imported into the boosting algorithm. The revised *TrAdaBoost* algorithm was implemented using Matlab 7.0. Interested readers can find the complementary material (Matlab source code and demos) of *TrAdaBoost* algorithm at https://github.com/Jiaolong/Boosting_DA.

3.4.1. Domain divergence changing

To determine the *TrCbrBoost* efficiency in selecting useful old data (source domain data), in each iteration, we calculated the *KL-divergence* between the reference data and the selected source

Table 5
Error matrix of the *k-NN* for land use classification.

Classified data (pixel)	Reference data								UA (%)
	BL	RI	FP	OC	LG	TL	BU	AL	
BL	298	0	0	1	1	23	8	6	88.43
RI	0	190	7	0	0	5	0	0	94.06
FP	0	5	24	0	0	0	0	9	63.16
OC	1	0	5	19	36	1	0	0	30.65
LG	1	0	0	10	273	24	0	6	86.94
TL	5	3	0	1	2	238	128	2	62.80
BU	7	0	0	0	1	22	399	11	90.68
AL	31	3	125	0	1	0	0	482	75.08
Total	343	201	161	31	314	313	535	516	
PA (%)	86.88	94.53	14.91	61.29	86.94	76.04	74.58	93.41	

UA = the user's accuracy, PA = the producer's accuracy.
Overall classification accuracy = 79.66%.
Overall kappa statistics = 0.75.

Table 6
Error matrix of the *TrCbrBoost* for land use classification.

Classified data (pixel)	Reference data								UA (%)
	BL	RI	FP	OC	LG	TL	BU	AL	
BL	268	0	0	0	3	5	8	4	93.06
RI	0	185	5	0	3	3	2	4	91.58
FP	1	1	111	0	2	5	2	8	85.38
OC	0	0	1	25	6	2	0	1	71.43
LG	2	1	0	4	251	3	3	3	94.01
TL	12	2	3	1	16	225	32	11	74.50
BU	50	4	13	1	24	53	480	14	75.12
AL	10	8	28	0	9	17	8	471	85.48
Total	343	201	161	31	314	313	535	516	
PA (%)	78.13	92.04	68.94	80.65	79.94	71.88	89.72	91.28	

UA = the user's accuracy, PA = the producer's accuracy.
Overall classification accuracy = 83.51%.
Overall kappa statistics = 0.80.

domain data. We used Eqs. (1) and (2) to measure the divergence of each learner's training data from the reference data. As can be seen in Table 3, the divergence is significantly reduced (by more than 20%).

Fig. 7 shows the examples of distribution variation found in the training data for band 2 and band 3. As can be seen, the difference between the training samples and reference data is particularly pronounced in the high-density region (marked in red[1] in Fig. 7) at the start of the iteration (Fig. 7(a): $R = 1$). The high-density region of training samples resides in approximately 0.35–0.45 of the x axis and 0.37–0.48 of the y axis, whereas the corresponding values for the same region of the reference data are 0.25–0.37 and 0.18–0.32, respectively. However, as expected, the divergence decreases as the number of iterations increases. The source domain data that are less similar to that of the target domain will be eliminated by this procedure, due to their negative effect on the training of good classifiers. At the end of the iteration, the high-density region of the training samples was fitted to approximately 0.25–0.35 of the x axis and 0.22–0.36 of the y axis (Fig. 7(d): $R = 100$). Compared to the initial values, this result is closer to the high-density region of the reference data.

3.4.2. Accuracy assessment

In order to evaluate its performance in terms of land use classification accuracy, the *TrCbrBoost* model was compared to three traditional models. These are *SVM* trained with source domain labeled data (SVM_{DS}), fuzzy *k-NN* model trained with source domain labeled data (mentioned in Section 2.2.2), and *SVM* trained with the target domain labeled data (SVM_{DT}). Prior to commencing the comparison, the 11 features of labeled data were imported into the *SVM* to train the classification model.

For assessing the accuracy of classification, 2414 reference data points were used, including 535 built-up pixels, 313 transportation pixels, 343 bare land pixels, 201 water pixels, 161 fishpond pixels, 31 orchard pixels, 314 lawn pixels, and 516 agricultural land pixels. In Tables 4–7, the confusion matrices representing the classification quality of each model are given, while the land use classification map produced by each model is given in Fig. 8(a)–(d). For the SVM_{DT}, about half of the reference data points were used as training data, while the remaining set was employed in the accuracy assessment.

At 74.32%, the accuracy of the SVM_{DS}, which is trained with source domain data, was the lowest (Fig. 8(a) and Table 4). The large distribution divergence of spectral and textual features causes *SVM* to misclassify land use types with similar spectral and textual information, such as built-up area and transportation, orchard and lawn.

At 79.71%, the accuracy of the *k-NN* model is somewhat better than that of the *SVM*, as shown in Fig. 8(b) and Table 5. This is likely due to the fact that, while both models employ the source domain labeled data to train the classifier, the temporal land use information is used only in the *k-NN* model.

Compared to the above, *TrCbrBoost* is a superior approach, with an overall accuracy of 83.51% (Fig. 8(c)). As shown in Table 6,

[1] For interpretation of color in Fig. 7, the reader is referred to the web version of this article.

Table 7
Error matrix of the SVM_{DT} for land use classification.

Classified data (pixel)	Reference data								UA (%)
	BL	RI	FP	OC	LG	TL	BU	AL	
BL	178	0	0	0	1	8	1	1	94.18
RI	0	107	0	0	0	1	0	1	98.17
FP	0	0	69	0	0	2	0	2	94.52
OC	0	0	0	16	2	2	0	1	76.19
LG	0	0	0	2	164	2	1	2	95.91
TL	4	2	2	0	2	100	19	0	77.52
BU	6	1	5	0	5	59	273	0	78.22
AL	2	2	14	0	1	2	0	277	92.95
Total	190	112	90	18	175	175	294	284	
PA (%)	93.68	95.54	76.67	88.89	93.71	56.82	92.86	97.54	

UA = the user's accuracy, PA = the producer's accuracy.
Overall classification accuracy = 88.49%.
Overall kappa statistics = 0.87.

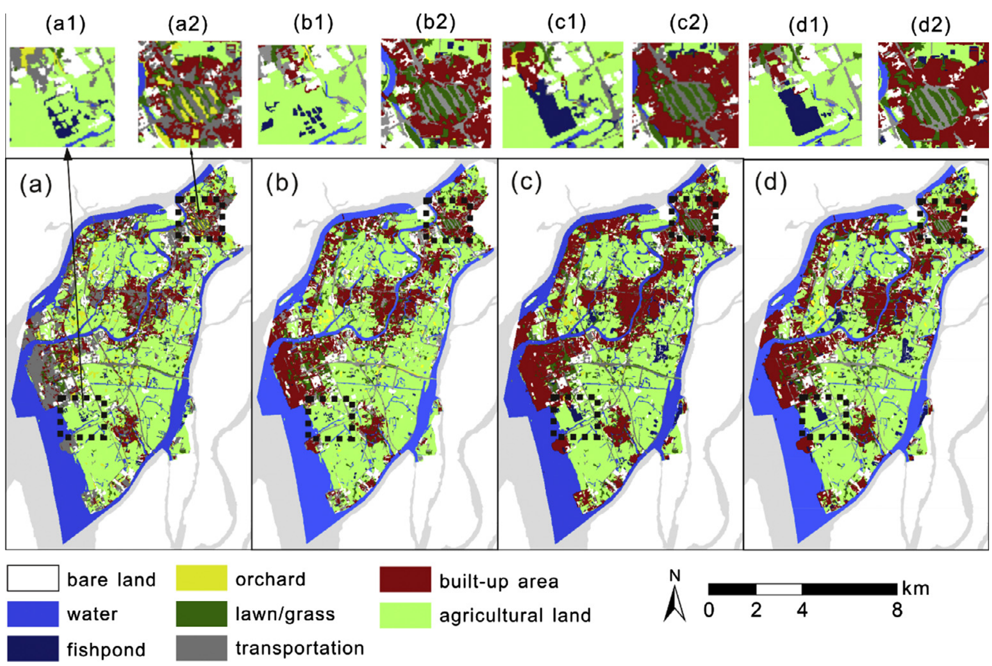

Fig. 8. Land use maps produced by classification approaches. (a) SVM model train with source domain data, overall accuracy = 74.32%; (b) Fuzzy *k-NN* model, overall accuracy = 79.66%; (c) *TrCbrBoost*, overall accuracy = 83.51%; and (d) SVM model train with target domain data, overall accuracy = 88.49%.

because of their similar spatial relationship, the specific accuracies pertaining to fishpond (*FP*), lawn (*LG*), and transportation (*TL*) classification are lower, compared to those achieved for other land use types. The similar spatial relationship information directly influenced the land use probabilities produced by *CBR*.

The results shown above clearly demonstrate that the *SVM* model can produce satisfactory results when sufficient quantity of good quality training data is available. Fig. 8(d) and Table 7 pertain to the classification output of SVM_{DT}, which is trained using the target domain labeled data. Predictably, this *SVM* model produces the highest accuracy (88.49%), as shown in Fig. 8(d). As this model is trained with the target domain data, its performance is better than the model is trained with the source domain data and improved upon by *TrCbrBoost*.

Fig. 8(a1), (b1), (c1) and (d1) depicts the four models' classification results pertaining to the same site. It should be noted that some of the farmland owners in Pearl River Delta enjoy a special agricultural production arrangement, allowing them to periodically designate their land as either fishpond or agricultural plot. Thus, for the *k-NN* model, which uses the temporal land use information and assigns a low weight to the spectral feature, it is easy to mistakenly classify the fishpond to the agricultural land category (as shown in Tables 4–7, the fishpond accuracy of *k-NN* is the lowest of the four models). The *TrCbrBoost* model overcomes this problem by using the domain adaptive strategy to find the useful old labeled data with which to train the classifier. If the training samples are chosen so that they fit the distribution of the target domain data well, it is easy to distinguish fishpond and agricultural land by using spectral and texture information. As a result, the *TrCbrBoost* and SVM_{DT} produce satisfactory results when detecting the fishpond land use type (Fig. 8(c1) and (d1)).

Fig. 8(a2), (b2), (c2) and (d2) shows the four models' classification results obtained at another site. As can be seen, in some cases, lawn is mistakenly classified as orchard in Fig. 8(a1), because the spectral and texture distribution of these two land use types is similar. Thus, these unstable distribution features have negative effect on the *SVM* model training success. If target domain labeled data is used as training samples instead, *SVM* model can classify these two land use types very well, as shown in Fig. 8(d2). On the other hand, the *k-NN* model can correctly classify the lawn

(Fig. 8(b2)), as it relies on the source domains temporal land use information, in which is classified as such. Similarly, as the useful old data are selected by the domain adaptive strategy in the *TrCbrBoost*, it too can classify the lawn with less errors (Fig. 8(c2)).

4. Conclusions

This paper presents a new approach to using knowledge transfer, called *TrCbrBoost*, as the aim is to improve classification performance. As we have shown, this method is effective in the domain adapting of land use classification by using various types of source (old) training data. Its main advantage is that it does not require any newly collected labeled data. This can significantly save the labor costs because collecting labeled data is very expensive and time consuming and may be impossible in some cases.

TrCbrBoost adopts three techniques to achieve the domain adapting purpose. First, the features pertaining the labeled data are extracted, before being evaluated using *KL-divergence* to identify the stable and unstable ones. Second, a fuzzy *CBR* is further applied to these features. Here, these features are weighted according to their divergences to generate land use probability maps of the target domain. Third, by using the maps of land use probabilities to assess the quality of old data, the *TrAdaBoost* technique is used to allow the distribution of higher-weight training samples to converge toward that of the target domain data. Finally, half of the classifiers remaining after this process is used to create the majority votes for obtaining the classification result.

In order to assess its performance, the proposed *TrCbrBoost* was applied to the land use classification of the middle part of the Pearl River Delta in China. For this purpose, 11 features were derived from the object-based analysis. The labeled data of source domains and the probability maps, which were generated by *CBR*, were imported into the boosting learning model. The aim of boosting is to reduce the distribution divergence between the source domain (training) data and the target domain data. Thus, the distribution of the selected training data set can approach that of the target domain data with each new iteration. The results reported in this work indicate that the divergence between these two distributions is reduced by more than 20%. Finally, the *TrCbrBoost* classification performance was compared to that of three models widely used in the field. The experimental results indicated that the proposed model has significant advantages when the target domain labeled data are not available.

The experiments conducted in this study revealed that almost half of the features in the data sets used were unstable. However, even in such cases, there were sufficient overlapping parts that could provide the crucial information for the classification. This fact is important for the use of our proposed knowledge transfer method. However, it should be noted that this method would fail if most of the features have high divergence across domains (no overlapping of the feature distributions). Another limitation of this method is that the *TrCbrBoost* has low computing efficiency because it is based on an ensemble method. Rectifying this issue is the subject of our future work.

Acknowledgement

This study was supported by the National Natural Science Foundation of China (Grant No. 41371376).

References

Aamodt, A., Plaza, E., 1994. Case-based reasoning: foundational issues, methodological variations, and system approaches. AI Commun. 7, 39–59.
Awrangjeb, M., Ravanbakhsh, M., Fraser, C.S., 2010. Automatic detection of residential buildings using LIDAR data and multispectral imagery. ISPRS J. Photogram. Remote Sens. 65, 457–467.
Awrangjeb, M., Zhang, C., Fraser, C.S., 2012. Building detection in complex scenes thorough effective separation of buildings from trees. Photogram. Eng. Rem. Sens. 78, 729–745.
Bennett, K.P., Demiriz, A., 1999. Semi-supervised support vector machines. Adv. Neural Inf. Process. Syst., 368–374.
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. Photogram. Remote Sens. 58, 239–258.
Chang, C.C., Lin, C.J., 2011. LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. (TIST) 2 (3), 27.
Chapelle, O., Weston, J., Schölkopf, B., 2002. Cluster kernels for semi-supervised learning. Adv. Neural Inf. Process. Syst., 585–592.
Dai, W., Yang, Q., Xue, G.R., Yu, Y., 2007. Boosting for transfer learning. In: In: Proceedings of the 24th International Conference on Machine Learning. ACM, pp. 193–200.
Dasarathy, B.V., 1990. Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques, vol. 1. IEEE Computer Society Press, Los Alamitos.
Du, Y., Zhou, C., Shao, Q., 2002. Theoretic and application research of geo-case based reasoning. Acta Geogr. Sin. – Chin. Ed. 57, 151–158.
Du, Y., Liang, F., Sun, Y., 2012. Integrating spatial relations into case-based reasoning to solve geographic problems. Knowl.-Based Syst. 33, 111–123.
Gong, P., Marceau, D.J., Howarth, P.J., 1992. A comparison of spatial feature extraction algorithms for land-use classification with SPOT HRV data. Remote Sens. Environ. 40, 137–151.
Hegarat-Mascle, L., Bloch, I., Vidal-Madjar, D., 1997. Application of Dempster-Shafer evidence theory to unsupervised classification in multisource remote sensing. IEEE Trans. Geosci. Remote Sens. 35, 1018–1031.
Jansen, L.J.M., Gregorio, A.D., 2002. Parametric land cover and land-use classifications as tools for environmental change detection. Agric. Ecosyst. Environ. 91, 89–100.
Keller, J.M., Gray, M.R., Givens, J.A., 1985. A fuzzy k-nearest neighbor algorithm. IEEE Trans. Syst. Man Cybern. 4, 580–585.
Kullback, S., Leibler, R.A., 1951. On information and sufficiency. Ann. Math. Stat. 22, 79–86.
Li, X., Yeh, A.G.O., 2000. Modelling sustainable urban development by the integration of constrained cellular automata and GIS. Int. J. Geogr. Inf. Sci. 14, 131–152.
Li, X., Yeh, A.G.O., 2002. Urban simulation using principal components analysis and cellular automata for land-use planning. Photogram. Eng. Remote Sens. 68, 341–352.
Li, X., Yeh, A., 2004. Multitemporal SAR images for monitoring cultivation systems using case-based reasoning. Remote Sens. Environ. 90, 524–534.
Li, X., Lao, C., Liu, Y., Liu, X., Chen, Y., Li, S., Ai, B., He, Z., 2013. Early warning of illegal development for protected areas by integrating cellular automata with neural networks. J. Environ. Manage. 130, 106–116.
Li, C., Wang, J., Wang, L., Hu, L., Gong, P., 2014. Comparison of classification algorithms and training sample sizes in urban land classification with landsat thematic mapper imagery. Remote Sens. 6, 964–983.
Lu, D., Weng, Q., 2007. A survey of image classification methods and techniques for improving classification performance. Int. J. Remote Sens. 28, 823–870.
Marceau, D.J., Howarth, P.J., Dubois, J.M.M., Gratton, D.J., 1990. Evaluation of the grey-level co-occurrence matrix method for land-cover classification using SPOT imagery. IEEE Trans. Geosci. Remote Sens. 28, 513–519.
Matasci, G., Volpi, M., Tuia, D., Kanevski, M., 2011. Transfer component analysis for domain adaptation in image classification. In: SPIE Remote Sensing. International Society for Optics and Photonics, 81800F-81800F-9.
Pan, S.J., Yang, Q., 2010. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22, 1345–1359.
Pan, S.J., Tsang, I.W., Kwok, J.T., Yang, Q., 2009. Domain adaptation via transfer component analysis. IEEE Trans. Neural Netw., 1–12.
Rajan, S., Ghosh, J., Crawford, M.M., 2008. An active learning approach to hyperspectral data classification. IEEE Trans. Geosci. Remote Sens. 46, 1231–1242.
Robinove, C.J., 1981. The logic of multispectral classification and mapping of land. Remote Sens. Environ. 11, 231–244.
Rosenstein, M.T., Marx, Z., Kaelbling, L.P., Dietterich, T.G., 2005. To transfer or not to transfer. In: NIPS 2005 Workshop on Transfer Learning, p. 898.
Schapire, R.E., 1999. A brief introduction to boosting. In: Proc. 16th Int. Joint Conf. Artificial Intell. Lawrence Erlbaum Associates Ltd., pp. 1401–1406.
Shrestha, D.P., Zinck, J.A., 2001. Land use classification in mountainous areas: integration of image processing, digital elevation data and field knowledge (application to Nepal). Int. J. Appl. Earth Obs. Geoinf. 3, 78–85.
Tobler, W.R., 1965. Computation of the correspondence of geographical patterns. Pap. Reg. Sci. 15, 131–139.
Torrey, L., Shavlik, J., 2009. Transfer learning. Handbook of Research on Machine Learning Applications, vol. 3. IGI Global, pp. 17–35.
Weng, Q., 2012. Remote sensing of impervious surfaces in the urban areas: requirements, methods, and trends. Remote Sens. Environ. 117, 34–49.
Wicks, T., Smith, G., Curran, P., 2002. Polygon-based aggregation of remotely sensed data for regional ecological analyses. Int. J. Appl. Earth Obs. Geoinf. 4, 161–173.
Xia, L., 1996. Technical note. A method to improve classification with shape information. Int. J. Remote Sens. 17, 1473–1481.
Zhang, Q., Wang, J., Gong, P., Shi, P., 2003. Study of urban spatial patterns from SPOT panchromatic imagery using textural analysis. Int. J. Remote Sens. 24, 4137–4160.