Object-Based Fusion of Multitemporal Multiangle ENVISAT ASAR and HJ-1B Multispectral Data for Urban Land-Cover Mapping

Yifang Ban, *Member, IEEE*, and Alexander Jacob, *Student Member, IEEE*

***Abstract*—The objectives of this research are to develop robust methods for segmentation of multitemporal synthetic aperture radar (SAR) and optical data and to investigate the fusion of multitemporal ENVISAT advanced synthetic aperture radar (ASAR) and Chinese HJ-1B multispectral data for detailed urban land–cover mapping. Eight-date multiangle ENVISAT ASAR images and one-date HJ-1B charge-coupled device image acquired over Beijing in 2009 are selected for this research. The edge-aware region growing and merging (EARGM) algorithm is developed for segmentation of SAR and optical data. Edge detection using a Sobel filter is applied on SAR and optical data individually, and a majority voting approach is used to integrate all edge images. The edges are then used in a segmentation process to ensure that segments do not grow over edges. The segmentation is influenced by minimum and maximum segment sizes as well as the two homogeneity criteria, namely, a measure of color and a measure of texture. The classification is performed using support vector machines. The results show that our EARGM algorithm produces better segmentation than eCognition, particularly for built-up classes and linear features. The best classification result (80%) is achieved using the fusion of eight-date ENVISAT ASAR and HJ-1B data. This represents 5%, 11%, and 14% improvements over eCognition, HJ-1B, and ASAR classifications, respectively. The second best classification is achieved using fusion of four-date ENVISAT ASAR and HJ-1B data (78%). The result indicates that fewer multitemporal SAR images can achieve similar classification accuracy if multitemporal multiangle dual-look-direction SAR data are carefully selected.**

***Index Terms*—Edge-aware region growing and merging (EARGM), ENVISAT advanced synthetic aperture radar (SAR) (ASAR), fusion, HJ-1B, multitemporal, segmentation, urban land cover.**

I. Introduction

In 2011, humanity passed a significant milestone—now, the planet has seven billion inhabitants with more than half living in cities [1]. Between now and 2030, the world is expected to add an additional two billion urban dwellers, or 62% of the estimated global population of 8.1 billion will live in cities. Most of the urban demographic transformation in the coming decades will occur in Asia and Africa. In China, for example, it is estimated that 400 million will move to urban areas by 2030, and the total urban population will be at 1.46 billion or 65%. Although only a small percentage of global land cover, urban areas significantly alter climate, biogeochemistry, and hydrology at local, regional, and global scales [2], [3]. Therefore, mapping urban land use/land cover in a timely and accurate manner is of critical importance for sustainable development.

Most of the efforts for urban mapping in China are based on optical remote sensing, e.g., [4]–[7]. Due to frequent cloud cover, smog, haze, and sand storms, however, optical data may not be available during critical monitoring cycle. With its all-weather capability and its unique information content, synthetic aperture radar (SAR) is an attractive data source for urban land-cover mapping. Single-date and multitemporal SAR data have been increasingly used in urban applications for mapping urban extent, land-cover classification, and change detection, e.g., [8]–[34]. In terms of SAR data used, various multitemporal spaceborne SAR data in single polarization, dual polarization, and full polarimetry have been evaluated for urban applications, and they include the European Remote Sensing Satellite 1/2, ENVISAT advanced SAR (ASAR), RADARSAT-1/2 C-band SAR, the Japanese Earth Resources Satellite 1 and Advanced Land Observation Satellite PalSAR L-band SAR, TerraSAR-X and COnstellation of small Satellites for the Mediterranean basin Observation (COSMO)/SkyMed X-band SAR, and multifrequency Spaceborne Imaging Radar-C/X-band SAR. In terms of SAR processing and feature extraction methods, texture measures, e.g., [12]–[14], [18], and [19], contextual information, e.g., [28], and classification algorithms such as statistical, e.g., [15], neural network, e.g., [15], [17], [19], and [30], support vector machine (SVM), e.g., [20], [29], [31], and [39], knowledge-based, and rule-based approaches, e.g., [8], [24], [29], and [30], have been investigated for urban mapping with varying degree of success. In addition to amplitude information, coherence information has also been explored for urban mapping, e.g., [14] and [17]. In terms of multitemporal multiangle analysis, several studies [15], [17], [29] compared multidate SAR combinations and demonstrated that the classification results could be improved by additional data. Aside from an earlier study by the author [19], only one study investigated multitemporal multi-viewing-angle SAR data for urban road network extraction using simulated COSMO/SkyMed SAR images [11]. The study found that the advantage of multiviewing analysis is usually related to the shadowing and layover effects that reduce the usefulness of single-view images. No studies, however,

Manuscript received March 4, 2012; revised July 13, 2012; accepted September 30, 2012. Date of publication February 12, 2013; date of current version March 21, 2013. This work was supported by the Swedish National Space Board.

The authors are with the Division of Geoinformatics, KTH Royal Institute of Technology, 10044 Stockholm, Sweden (e-mail: yifang@kth.se; aljacob@kth.se).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2012.2236560

investigated multitemporal multi-incidence-angle dual-orbit SAR data for urban analysis. Man-made structures have large dihedral/trihedral surfaces that result in high backscatter at appropriate look angles. The incidence angle, on the other hand, affects the detectability of settlements through its control of range resolution. Small incident angles produce poorer range resolution than larger angles. Higher resolution allows detection of smaller settlements and mapping of small complex land-cover units, while larger incident angles generate longer radar shadows than smaller angles. Radar shadows conceal or obfuscate land cover [9]. Therefore, it is desirable to evaluate the effect of temporal incidence angle and look direction of spaceborne SAR data for urban land-cover mapping.

Furthermore, the fusion of SAR and optical data has been proven advantageous by making use of the complementary nature of the data acquired by different sensors, e.g., [23] and [35]–[42]. Both SAR and optical data have their own merits and limitations, thus the fusion of SAR and optical data can overcome the deficiencies associated with a single sensor. Various studies have been conducted for urban land-cover mapping using fusion of SAR and optical data [23], [41]–[47]. However, most of these urban studies mapped only one or few urban classes. Detailed land-cover mapping in complex urban environments is a challenge for several reasons, as identified in [23] and [29]. Therefore, there is a strong need to explore SAR and optical data fusion for detailed urban land-cover mapping.

Most of the aforementioned fusion studies used a pixel-based classification. Object-based methods, on the other hand, have been increasingly adopted in urban mapping using high-resolution optical data [48], SAR data, e.g., [27] and [29], and fusion of SAR and optical data, e.g., [23] and [46], since more information such as object features and spatial relationships could be explored in the analysis, thus improving the classification accuracy. The existing automatic image segmentation techniques can be classified into four approaches: 1) thresholding techniques; 2) boundary-based methods; 3) region-based methods; and 4) hybrid techniques [49], [50]. The pros and cons of the techniques are discussed in [48] and [49]. The most commonly used segmentation method is region growing and merging in eCognition [48]. Boundary-based methods using edge detection are also explored for image segmentation [51]–[53]. Several studies found that integrating edge and region detection can produce better segmentation of images [54]–[58]. However, these studies used relatively simple images such as photographs of faces, flowers, and buildings [49], [54]–[58], not the complex spaceborne SAR and optical data in challenging urban environments. Therefore, further research is needed to develop robust methods for segmentation of SAR and optical data that integrates edge detection and region growing and merging.

The availability of the European Space Agency (ESA) ENVISAT multitemporal multiangle ASAR data and Chinese Earth Observations satellite data, such as BJ-1 and HJ-1A/1B, provides excellent opportunities for data fusion research. The objectives of this research are to develop robust method for segmentation of SAR and optical data and to investigate the potential of fusion of multitemporal ENVISAT ASAR and HJ-1B optical images for detailed land-cover mapping in rapidly changing complex urban environments.

TABLE I
ENVISAT ASAR IMAGE CHARACTERISTICS

Acquisition Date	Polarization	Beam Mode	Incidence Angle (°)	Orbit
20090507	HH/VV	IS6	39.1 – 42.8	Ascending
20090517	HH/VV	IS2	19.2 - 26.7	Ascending
20090527	HH/VV	IS2	19.2 - 26.7	Descending
20090608	HH/VV	IS4	30.6 - 36.2	Ascending
20090611	HH/VV	IS6	39.1 – 42.8	Ascending
20090804	VV	IS7	42.5 – 45.2	Ascending
20090908	VV	IS7	42.5 – 45.2	Ascending
20091013	VV	IS7	42.5 – 45.2	Ascending

II. STUDY AREA AND DATA DESCRIPTION

The study area is Beijing, the capital of China and one of the cities highly representative for rapid urbanization. The major land-cover classes include high-density built-up area, low-density built-up area, roads, airport, forest, low vegetation (including parks and crops), golf course, grass/pasture, bare fields, and water.

Eight-date ENVISAT ASAR images with multiple incidence angle, alternating polarization (HH/VV), and single polarization (VV) at 12.5-m pixel spacing (30-m resolution) were acquired from May to October 2009. All SAR images were in ascending orbit except one. The data set contains four ASAR beam modes with incidence angles ranging from 20° to 45° that may result in effects of shadow, layover, and foreshortening in urban areas, depending on building heights. These effects, however, are relatively small due to the medium resolution. For high-resolution SAR data, the effect will be much stronger. Table I shows the characteristics of the ASAR images used in this research. One scene of HJ-1B charge-coupled device (CCD) image acquired on May 12, 2009, was selected for this research since no images were acquired during the peak vegetation season when confusion between urban classes and bare fields can be reduced. The early season image also presents an opportunity for research into the role of SAR in improving land-cover classification of optical data acquired in nonoptical season as this is often the case in many parts of the world due to cloud cover. HJ-1B stands for Disaster and Environment Monitoring and Forecast Small Satellite Constellation B, which is one of the small Chinese Earth observation satellites. Launched on September 6, 2008, HJ-1B payloads include a CCD camera that acquires images in blue, green, red, and near-infrared parts of the spectrum at 30-m resolution and an infrared camera [59].

To evaluate the efficient combination of ASAR and HJ-1B data, four-date ENVISAT ASAR images, i.e., from May 7, May 27, June 8, and August 4, were selected to represent four different incidence angles (IS2, IS4, IS6, and IS7) and two look directions.

III. METHODOLOGY

In this research, an image segmentation method based on edge detection and region growing and merging algorithm was developed for segmentation of ENVISAT SAR and

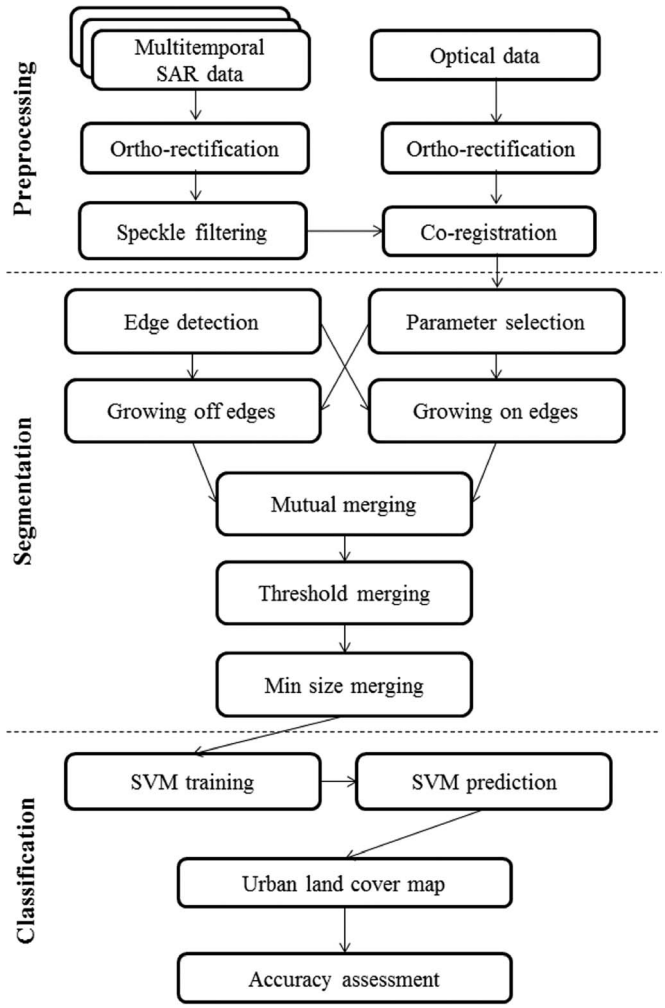

Fig. 1. Overview of the Methodology.

HJ-1B multispectral data. The analysis can be divided into three phases, i.e., image preprocessing, edge detection and image segmentation, and classification. Large parts of the processing chain have been implemented in our own software package, a 64-b JAVA application. An overview of the methodology is illustrated in Fig. 1. The details of the methodology are discussed in the following sections.

A. Image Preprocessing

Since the data sets consist of one-date multispectral image and eight-date SAR images from multiple incidence angles and two look directions, it is necessary to accurately coregister all these images together. First, all ASAR images were orthorectified based on the satellite orbital model and the Shuttle Radar Topography Mission digital elevation model with rmss less than 1 pixel (or 12.5 m) and coregistered using the Next ESA SAR Toolbox 4B. Then, the HJ-1B image was orthorectified and coregistered to the orthorectified SAR data set using the Geomatica OrthoEngine of PCI Geomatics, resulting in the final image database of 17 channels. Then, the images were subset to cover the sixth-ring road around Beijing.

To reduce speckle, a multitemporal speckle filter was applied to all SAR images. As [20] and [30] demonstrated the superiority of using the compressed 8-b data rather than the raw 16-b data for SAR image segmentation, the SAR images were scaled down to 8 b using a 95% clipped linear scaling function. For better edge detection and segmentation, the images are enhanced using histogram equalization.

B. Image Segmentation Using EARGM

1) Edge Detection: Edge-based segmentation techniques use the assumption that pixel values change rapidly at the boundary between two regions. Commonly used edge detection techniques include simple ones such as the Sobel or Roberts operators, or more complex ones such as the Canny operator [49]. In this research, edge detection is achieved by applying a 5×5 Sobel filter to the ASAR data and HJ-1B images individually. To extract the most evident edges, the resulting gradient images are then thresholded with a 99% cutoff, i.e., only the ones with 99% of the maximum gradient in the image are extracted as edges. Remaining small clusters of edge pixels are filtered out using a combined convolution and threshold operation. The threshold is selected based on empirical testing. For all filtering operations, duplication is applied to solve problems at the outer edges of the whole scene. A majority voting approach is applied on all the edge images of SAR and on that of all the optical bands, respectively, to create two binary decision images (0 for no edge and 1 for edge). Then, the two edge images are joined by a logical OR operator to create the final decision image of edge or no edge. All operations performed during edge detection are implemented using JAVA Advanced Imaging.

2) EARGM: Region-based segmentations rely on the assumption that adjacent pixels in the same region have similar visual features such as gray level, color value, or texture. The commonly used approach is region growing and merging, and the performance of this approach largely depends on the selected homogeneity criterion [49]. In this research, the segmentation is divided into four main steps (Fig. 2) and can be influenced by three parameters. The parameters are minimum and maximum segment sizes as well as a weighting between the two homogeneity criteria used. The two homogeneity criteria are a measure of color (1) as the average change of mean through a possible combination of two pixels or segments and a measure of texture (2) as the average change of standard deviation through a possible merging of two segments

$$\text{Mean\_Measure} = \frac{\sum_n \frac{\mu_b - \mu_a}{\mu_b}}{n}, \quad \begin{array}{l}\mu_b = \text{mean\_before} \\ \mu_a = \text{mean\_after} \\ n = \text{number\_of\_bands}\end{array} \tag{1}$$

$$\text{SD\_Measure} = \frac{\sum_n \frac{\sigma_b - \sigma_a}{\sigma_b}}{n}, \quad \begin{array}{l}\sigma_b = \text{SD\_before} \\ \sigma_a = \text{SD\_after} \\ n = \text{number\_of\_bands.}\end{array} \tag{2}$$

The actual segmentation is started by a region growing operation which is split into growing along edges and growing offside of edges. To decide with which neighbor pixels or segments to merge, the color criterion alone is used. The order of segment seeds is determined by the average entropy of all bands of a 5×5 neighborhood around a pixel. This initial growing is limited to a quarter of the minimum segment size. Once the

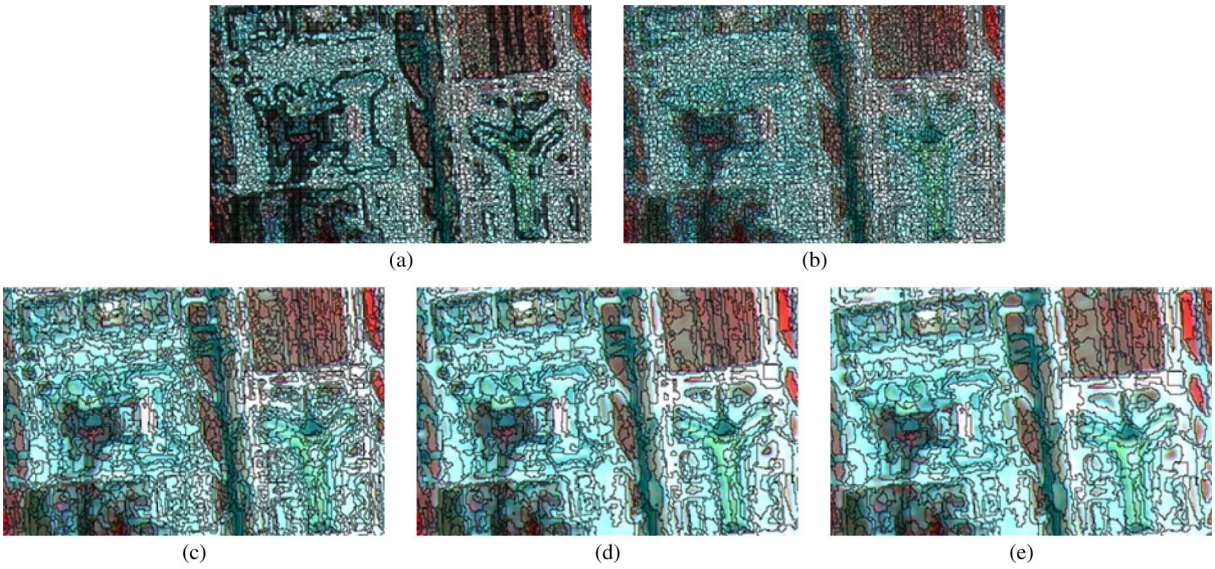

Fig. 2. Segmentation process. (a) Grow off edges. (b) Grow on edges. (c) Mutual merging. (d) Threshold merging. (e) Minimum size merging.

growing is completed, the first merging step is started. In contrast to the growing phase which was a one-sided test, the first merging step is a mutual test, i.e., only if two segments are each other's best choices determined by both the color and the texture criterion, they are actually merged. The mutual merging operation is finished when either the minimum segment size is reached or no more mutual best partners can be found.

The next step is the adaptive threshold merging operation. The threshold is defined as a third of the average of the weighted sum of the two merging criteria over all possible segment merges at this stage of the segmentation process. We chose the threshold to be based on the global homogeneity so that it is adaptive to the data set and can be determined automatically during processing. If the scene is very homogeneous, the threshold will be smaller, and if the scene is very heterogeneous, the threshold will be bigger. In this study, the threshold was determined empirically by testing on different data sets at 30-m resolution, such as ENVISAT ASAR, HJ-1B, Landsat Thematic Mapper, and Beijing-1 data. From the testing, we found that one-third of the global homogeneity works best for ENVISAT ASAR and HJ-1B individually and their fusion. The threshold merging is iteratively applied while increasing the maximum segment size and decreasing the threshold at each iteration. An iteration is completed when either all segments reached the current maximum size or no pair of segments is below the current threshold criterion after a possible merge. The last iteration is performed when the maximum segment size determined by the input parameter is reached.

As the last step in segmentation, all those segments which are below the minimum size are merged to their best fitting neighboring segments. The minimum segment size was set to 40 pixels, and the maximum was set to 4000. The weighting between the homogeneity criteria is set to 0.5 each.

Of all parameters in the segmentation process, the minimum segment size has the biggest influence on the outcome of the segmentation since it defines the smallest possible mapping unit. The maximum segment size only has an influence on the segmentation when it is not selected very big in relation to the minimum segment size. In this research, a very big maximum segment size was chosen, and hence, the influence on the results is rather small. The threshold has more effect on the results than the maximum segment size. This is mainly due to high heterogeneity of urban scenes. In more homogeneous vegetation/agricultural scenes, probably, the maximum segment size would have a stronger effect.

For comparison, segmentation of ASAR and HJ-1B data is also performed in eCognition. Multiresolution segmentation in eCognition is a bottom-up region merging technique, merging smaller objects to form bigger ones by taking criteria of homogeneity in color and shape into account. The size of the object is adjusted by the scale parameter, which determines the maximum allowable heterogeneity of the objects. Image objects are generated in a network of hierarchical levels and are logically linked. Therefore, each object knows its neighbors and sub- and superobjects [60]. In this research, the scale was set to 21 to achieve approximately the same segment size and the total number of segments (total segments: 149 583) as from our segmentation (total segments: 145 895). All images are equally weighted, the mean color index is 0.9, the shape index is 0.1, and the compactness is set to 0.5.

C. Image Classification

1) SVM: Past studies showed that SVM was effective for classification of multitemporal or multisource data, e.g., [29] and [47], since there is no requirement of the statistical model for the data to be classified. The superiority of SVM to other classifiers in the object-based classification of SAR and optical data has also been reported by various studies, e.g., [20] and [39]. Therefore, SVM is selected for postsegmentation classification and is implemented using the JAVA version of LibSVM in this research.

SVM searches for the optimal hyperplane to separate the training vectors of two classes into two subspaces through mapping the input vectors into high-dimension space. Therefore, SVM is originally a binary classifier. In this research, the

one-against-one approach is adopted. Currently, there are two types of approaches for multiclass SVM. One is by constructing and combining several binary classifiers, while the other is by directly considering all data in one optimization formulation. The formulation to solve multiclass SVM problems in one step has variables proportional to the number of classes. Therefore, for multiclass SVM methods, either several binary classifiers have to be constructed or a larger optimization problem is needed. Hence, in general, it is computationally more expensive to solve a multiclass problem than a binary problem with the same number of data [61]. For classifying N classes, $N(N-1)/2$ binary SVM classifiers are created according to the total $N(N-1)/2$ different class pairs. Each classifier is trained for a unique class pair by the training data of the two classes. In the classification process, a "Max Wins" voting strategy is applied for the final decision. In this strategy, each classifier will vote the class which is identified by this classifier for an input data set. The class that has the most votes from all the classifiers will be labeled as the class for this input data set.

There are several kernel functions used to map the input vectors into the high-dimension space. In our study, radial basis function (RBF) is selected as the mapping kernel function

$$k(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right), \qquad \gamma > 0. \tag{3}$$

When using SVM for classification with the RBF kernel, optimization of two parameters is necessary to improve the predictive accuracy: the penalty value C and kernel parameter γ. The parameter C controls the tradeoff between errors of the SVM on training data and margin maximization. In our study, the best parameters were selected from a limited searching grid through a cross-validation process [29].

To classify the objects from multitemporal SAR, optical data, and fusion of SAR and optical data, the mean and the standard deviation for each object of that data set were calculated as the object's feature input into SVM. For multidate data, a multidate string of such means and standard deviations was obtained for each object. Therefore, the elements of the object's feature vector input into the SVM consisted of mean and standard deviation string, brightness, and maximum difference indicators. The brightness $\overline{C}(v)$ is defined as

$$\overline{C}(v) = \frac{1}{w^B} \sum_{k=1}^{K} w_k^B \overline{C}_k(v) \tag{4}$$

where w_k^B is the brightness weight of the kth parameter layer in the multitemporal data set; in this study, they were all set as 1. $\overline{C}_k(v)$ is the mean of all pixels in object v in the kth parameter layer. K is the number of the total parameter layers in the multidate stack. Moreover

$$w^B = \sum_{k=1}^{K} w_k^B. \tag{5}$$

In addition, the maximum difference is given as

$$\frac{\max\limits_{i,j \in K_B} \left|\overline{C}_i(v) - \overline{C}_j(v)\right|}{\overline{C}(v)} \tag{6}$$

where $K_B = \{k \in K | w_k^B = 1\}$, and in this study, it is equal to K.

In this study, ten urban land-cover classes were classified using SVM. In practice, in the SVM classification, high-density built-up areas are divided into four subclasses, while low-density built-up areas were divided into two subclasses. To achieve better classification results, these subclasses were classified separately and then merged to the superclass that they belong. In SVM classification, the data are mapped to a hyperspace which is further separated by hyperplanes according to the clustering status of the training data. If one class has many obvious clusters, it is better to divide the class into different subclasses. Otherwise, there will be more conflicts with other classes in the hyperspace, which will increase the frequency of error. While using several subclasses, most of the conflicts could be constrained within the superclass domain. Moreover, the accuracy of the subclasses could be improved due to better description of the clusters.

To calibrate the SVM classifier, 20 representative training objects for each of the 14 classes were selected. For the purpose of comparisons, the training segments for all five processing runs were selected in roughly the same areas. The training samples are then used to train the SVM.

2) Accuracy Assessment: To validate the SVM classification results, a set of testing areas, at least 2000 pixels for every class, were randomly selected and evenly distributed over the whole study area. Confusion matrices including producer's accuracy, user's accuracy, overall accuracy, and Kappa coefficient were computed.

3) Significance Test: To evaluate the significance of the differences in the classification accuracy, McNemar's test was employed by previous studies, e.g., [28]. In this study, McNemar's test was used to evaluate the significance of the differences between various classification results.

IV. Results and Discussion

The classification results of ASAR and HJ-1B alone, the fusion of four-date ASAR and HJ-1B data, and the fusion of eight-date ASAR and HJ-1B data are presented in Fig. 3. The comparison of overall classification accuracies and kappa coefficients is presented in Table II. Among all results, significant differences were found using McNemar's test, specifically between the overall accuracy and kappa for classification using our segmentation and that of eCognition.

Three representative areas were selected to compare various segmentation and classification results (Fig. 3). The left column shows Beijing International Airport, the center column the center of Beijing including the Forbidden City, and the right column the Summer Palace and its surrounding area. From the center column, we can see that the Tiananmen Square and the moat around the Forbidden City are better extracted in segmentation and classification using our algorithm (fifth and seventh rows) than eCognition. The horizontal (east–west) streets are much better detected than the vertical (north–south) streets. This is mainly due to the look direction of the SAR sensor that vertical streets are often blocked by buildings and trees. This effect cannot be observed in optical data alone.

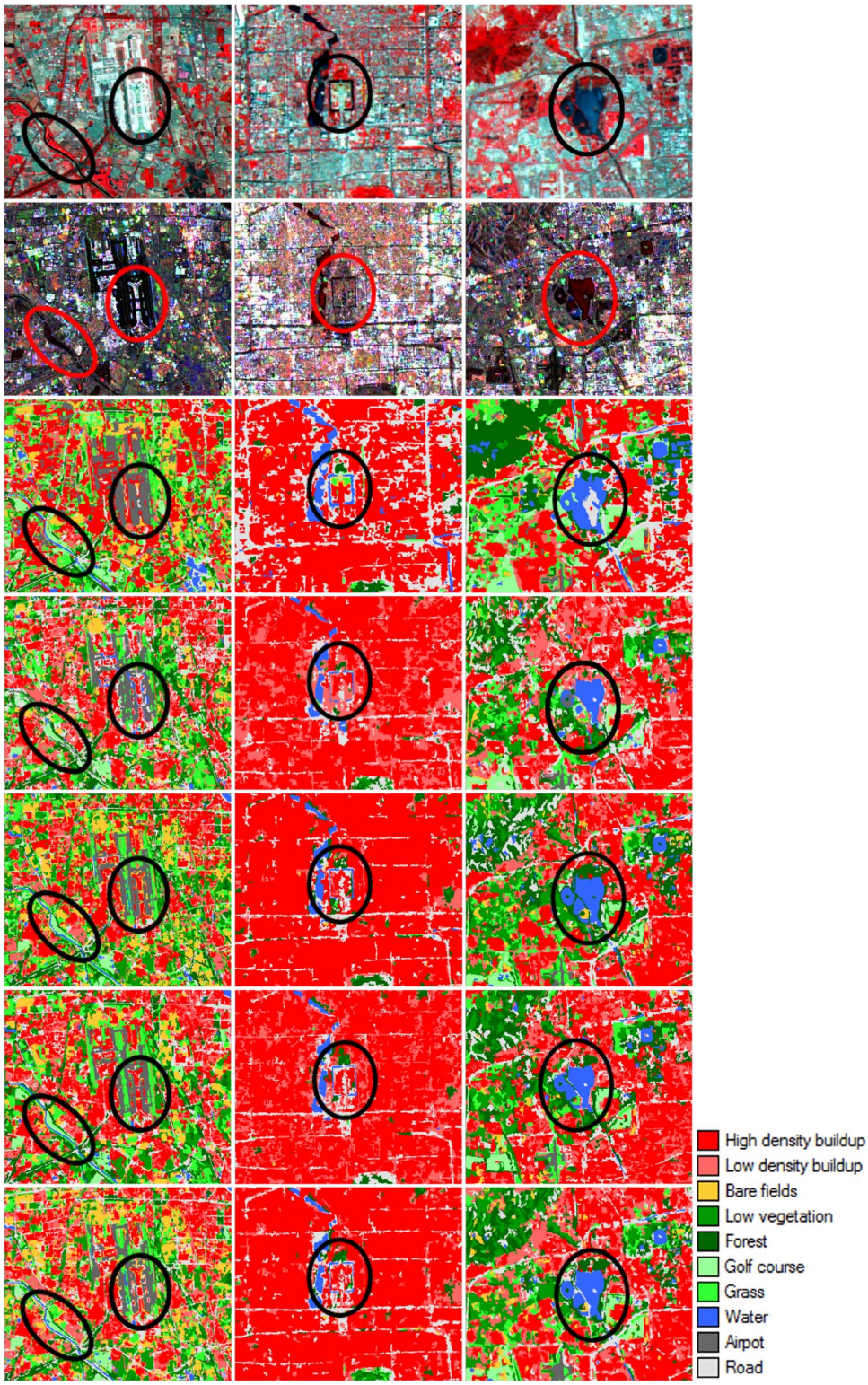

Fig. 3. Images: (Row 1) ASAR. (Row 2) HJ-1B. Classifications: (Row 3) ASAR. (Row 4) HJ-1B. (Row 5) Four-date ASAR using EARGM. (Row 6) Eight-date ASAR and HJ-1B using eCognition. (Row 7) Eight-date ASAR and HJ-1B using EARGM.

The optical data, however, should have been able to assist in improving the detection of vertical roads in the fusion results if heavier weight is given to optical data, even though it has only four channels out of 17. In comparison to the eCognition result (sixth row), the linear features such as the moat around the Forbidden City and the roads are detected better by our segmentation. In addition, high-density built-up areas were misclassified as low-density built-up areas in the city center in eCognition classification (row 6).

The area around the Summer Palace was selected as it contains lakes and several delicate features of bridges and islands. The bridges and islands were better classified by our

TABLE II
COMPARISON OF CLASSIFICATION ACCURACIES

	Fusion of all SAR & HJ-1	Fusion of 4-date SAR & HJ-1B	Fusion of all SAR & HJ-1B	ENVISAT ASAR	HJ-1B CCD
	EARGM	**EARGM**	**eCognition**	**EARGM**	**EARGM**
High-density builtup	90,1	85,6	81	85,2	82,4
Low-density builtup	85,4	78,5	64,7	72,6	38,7
Bare field	80,7	67,2	31	47,2	59,9
Low vegetation	60,2	45,1	56,2	38,7	31,8
Forest	70,5	91,3	88,4	79,9	86,2
Golf course	64,8	68,9	69,5	64,5	68
Grass/Pasture	78,4	78,8	98,7	69,8	67,4
Water	91,5	98,2	97,6	42,6	98.9
Airport	81,3	89,2	87,9	75	99,4
Major roads	82,7	68,6	71,2	59,2	51,6
Overall accuracy	**79,91**	**78,02**	**74,48**	**65,83**	**68,29**
Kappa Coefficient	**0,77**	**0,75**	**0,71**	**0,62**	**0,64**

segmentation than eCognition. The highway in the north is also very well extracted by our method but is almost invisible due to misclassification in the eCognition result. The optical result detects the highway quite well but does not show the bridges and islands in the Summer Palace. The results from four-date SAR and HJ-1B fusion (fifth row) have quite good segments in this area, but the bridge is misclassified as forest due to confusion. The area in the left column is interesting due to its heterogeneity; many different land-cover types are located here in a small area. It clearly shows the advantages (e.g., very good classification of river and golf course in HJ-1B versus very good separation of buildings and airport runways) and limitations (e.g., confusion between buildings and airport runways in HJ-1B versus confusion between river and golf course) of using optical data (third row) and SAR (fourth row) alone. Airport/built-up and river/golf course are well classified in the fusion results. These examples clearly demonstrate the synergistic effects of the SAR and optical data.

In terms of quality of the segmentations, several observations can be made when looking at the classification results presented in Fig. 3. First, the segments derived from HJ-1B data (row 3) often tend to grow over borders of different classes, particularly among roads and other urban classes, thus causing confusion among them. This is mainly due to the spectral similarity between roads and build-up areas. The results from ASAR (row 4) clearly have some advantages. The ASAR segments in general are better defined in built-up areas mainly attributed to the backscattering difference between roads (low backscatter if no layover) and buildings (high backscatter). When it comes to the actual classification, however, SAR data alone are not sufficient due to confusion among classes with low backscatter, such as water, grass/pasture, golf course, road, and airport. The segments from eCognition (row 6) are better than SAR and optical data alone but still have major problems with finding clear borders between areas of different land-cover types since eCognition only uses color (0.9) and shape (0.1) in segmentation. With the edge-aware region growing and merging (EARGM) algorithm, the quality of the segments is much improved, as shown in Fig. 3, row 5, from fusion of four-date ASAR and HJ-1B data and in Fig. 3, row 7, from fusion of eight-date ASAR and HJ-1B data. Two major advantages can be observed when comparing eCognition's result and the results using our own algorithm. The first one is that segments do not tend to grow over clear borders, and the second is that linear features such as major roads (center column) and rivers (left column: encircled area) are much better defined and can be well classified in the medium-resolution imagery.

The quality of segmentations is directly linked to the quality of classifications, as shown in Table II. The low classification accuracy of HJ-1B data is mainly caused by confusion between high- and low-density built-up areas, roads, and bare fields. On the other hand, the poor accuracy for classification of SAR alone was caused in part by low-backscatter classes, as discussed earlier, and in part by changes in vegetation from May to October. The best classification accuracy is achieved using fusion of eight-date ASAR and HJ-1B data. This represents over 5% improvement over eCognition results from the same data set. This also represents a 14% and an 11% improvement over SAR and HJ-1B data alone. The results further confirm the benefits of synergy of SAR and optical data. The most interesting finding is the closeness in the overall accuracy and kappa between fusion of four-date ASAR/HJ-1B data and that of eight-date ASAR/HJ-1B data. The results indicate that the multitemporal SAR data acquired in multiangle and dual-look directions are advantageous for urban land-cover mapping. The additional four-date SAR data with similar incidence angles do not add significant value to the classification.

Further comparison of the classification accuracies for fusion of 8-date SAR and HJ-1B data using our algorithm and eCognition (Table II) revealed that the EARGM method produced much better classification accuracies for high- and low-density built-up areas, major roads, and bare fields than eCognition. On the other hand, eCognition produced higher accuracies for several vegetation classes such as forest, grass/pasture, and golf course. This may imply that shape/compactness in eCognition segmentation could be helpful to define vegetation classes. Therefore, it is desirable to further improve our algorithm by adding shape/compactness parameters.

V. Conclusion

This paper has investigated the edge-aware region growing and merging algorithm for effective segmentation of SAR and optical data. The potential of the fusion of multitemporal multiangle ENVISAT ASAR and HJ-1B CCD data has also been explored for detailed urban land-cover mapping. The results show that the performance of EARGM is superior to that of eCognition in defining urban segments and linear features. The results also confirm that fusion of SAR and optical data provides complementary information, thus yielding higher classification accuracy than SAR or optical data alone. Furthermore, this research demonstrated that fusion of four-date SAR data and optical data can achieve similar classification accuracy as fusion of eight-date SAR data and optical data if multiangle dual-look-direction SAR data with suitable temporal compositions are selected.

Acknowledgment

The authors would like to thank Xin Niu of KTH Geoinformatics for his assistance in this research. The research is part of the project "Satellite Monitoring of Urbanization for Sustainable Urban Development" within the European Space Agency and the Chinese Ministry of Science and Technology's Dragon II program.

References

[1] UN, New York, 2011. [Online]. Available: http://www.un.org/apps/news/story.asp?NewsID=40257

[2] A. Schneider, M. A. Fried, and D. Potere, "A new map of global urban extent from MODIS satellite data," *Environ. Res. Lett.*, vol. 4, no. 4, pp. 044003-1–044003-11, Oct.–Dec. 2009.

[3] K. C. Seto, M. Fragkias, B. Güneralp, and M. K. Reilly, "A meta-analysis of global urban land expansion," *PLoS ONE*, vol. 6, no. 8, p. e23 777, Aug. 2011.

[4] C. Y. Ji, Q. Liu, D. Sun, S. Wang, P. Lin, and X. Li, "Monitoring urban expansion with remote sensing in China," *Int. J. Remote Sens.*, vol. 22, no. 8, pp. 1441–1455, Jan. 2001.

[5] A. Schneider, K. C. Seto, and D. R. Webster, "Urban growth in Chengdu, Western China: Application of remote sensing to assess planning and policy outcomes," *Environ. Plan. B, Plan. Design*, vol. 32, no. 3, pp. 323–345, Jan. 2005.

[6] K. C. Seto and M. Fragkias, "Quantifying spatiotemporal patterns of urban land-use change in four cities of China with time series landscape metrics," *Landscape Ecol.*, vol. 20, no. 7, pp. 871–888, Nov. 2005.

[7] Q. Zhang, Y. Ban, and J. Liu, "The trajectories of urban land and industrial land Shanghai over the past 30 years," in *Proc. Urban Remote Sens. Joint Event*, Shanghai, China, 2009, pp. 1–7.

[8] F. M. Henderson and Z.-G. Xia, "SAR applications in human settlement detection, population estimation and urban land use pattern analysis: A status report," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 1, pp. 79–85, Jan. 1997.

[9] Z.-G. Xia and F. M. Henderson, "Understanding the relationships between radar response patterns and the bio- and geophysical parameters of urban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 1, pp. 93–101, Jan. 1997.

[10] F. Dell'Acqua, P. Gamba, and G. Lisini, "Improvements to urban area characterization using multitemporal and multiangle SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1996–2004, Sep. 2003.

[11] F. Dell'Acqua and P. Gamba, "Texture-based characterization of urban environments on satellite SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 1, pp. 153–159, Jan. 2003.

[12] R. J. Dekker, "Texture analysis and classification of ERS SAR images for map updating of urban areas in The Netherlands," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1950–1958, Sep. 2003.

[13] M. E. Engdahl and J. M. Hyyppa, "Land-cover classification using multitemporal ERS-1/2 InSAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 7, pp. 1620–1628, Jul. 2003.

[14] T. M. Pellizzeri, P. Gamba, P. Lombardo, and F. Dell'Acqua, "Multitemporal/multiband SAR classification of urban areas using spatial analysis: Statistical versus neural kernel-based approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 10, pp. 2338–2353, Oct. 2003.

[15] P. Lombardo, M. Sciotti, T. M. Pellizzeri, and M. Meloni, "Optimum model-based segmentation techniques for multifrequency polarimetric SAR images of urban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1959–1975, Sep. 2003.

[16] M. C. Dobson, L. E. Pierce, and F. T. Ulaby, "Knowledge-based land-cover classification using ERS-1/JERS-1 SAR composites," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 1, pp. 83–99, Jan. 1996.

[17] L. Bruzzone, M. Marconcini, U. Wegmuller, and A. Wiesmann, "An advanced system for the automatic classification of multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 6, pp. 1321–1334, Jun. 2004.

[18] Y. Ban and Q. Wu, "RADARSAT SAR data for landuse/land-cover classification in the rural-urban fringe of the greater Toronto area," in *Proc. 8th AGILE Conf. Geograph. Inf. Sci.*, 2005, pp. 1–8.

[19] Y. Ban and A. I. Kazi, "ENVISAT ASAR for land cover mapping and change detection in the rural-urban fringe of the greater Toronto area," presented at the Proc. 5th Int. Symp. Retrieval Bio-Geophys. Param. SAR Data for Land Appl., Bari, Italy, Sep. 24–28, 2007.

[20] H. Hu and Y. Ban, "Urban land use/land cover mapping with high-resolution SAR imagery by integrating support vector machines into object-based analysis," in *Proc. SPIE Conf. Remote Sens. Environ. Monitoring, GIS Appl., Geol. VIII*, 2008, vol. 7110, pp. 71100K1-1–71100K1-8.

[21] M. Stasolla and P. Gamba, "Spatial indexes for the extraction of formal and informal human settlements from high-resolution SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 1, no. 2, pp. 98–106, Jun. 2008.

[22] L. Ferro-Famil and M. Lavalle, "Detection and analysis of urban areas using ALOS PALSAR polarimetric data," in *Proc. IEEE IGARSS*, Jul. 12–17, 2009, vol. 5, pp. V-142–V-145.

[23] Y. Ban, H. Hu, and I. M. Rangel, "Fusion of Quickbird MS and RADARSAT SAR for urban land-cover mapping: Object-based and knowledge-based approach," *Int. J. Remote Sens.*, vol. 31, no. 6, pp. 1391–1410, Feb. 2010.

[24] T. Esch, M. Thiel, A. Schenk, A. Roth, A. Muller, and S. Dech, "Delineation of urban footprints from TerraSAR-X data by analyzing speckle characteristics and intensity information," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 2, pp. 905–916, Feb. 2010.

[25] P. Gamba, M. Aldrighi, and M. Stasolla, "Robust extraction of urban area extents in HR and VHR SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 1, pp. 27–34, Mar. 2011.

[26] T. Esch, A. Schenk, T. Ullmann, M. Thiel, A. Roth, and S. Dech, "Characterization of land cover types in TerraSAR-X images by combined analysis of speckle statistics and intensity information," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 1911–1925, Jun. 2011.

[27] H. Hu and Y. Ban, "Multitemporal RADARSAT-2 ultra-fine beam SAR data for urban land cover classification," *Can. J. Remote Sens.*, vol. 38, no. 1, pp. 1–11, Feb. 2012.

[28] X. Niu and Y. Ban, "An adaptive contextual SEM algorithm for urban land cover mapping using multitemporal high-resolution polarimetric SAR data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 4, pp. 1129–1139, Aug. 2012.

[29] X. Niu and Y. Ban, "Multi-temporal RADARSAT-2 polarimetric SAR data for urban land-cover classification using an object-based support vector machine and a rule-based approach," *Int. J. Remote Sens.*, vol. 34, no. 1, pp. 1–26, Jan. 2012.

[30] H. Hu and Y. Ban, "Urban land-use mapping and change detection with RADARSAT fine-beam SAR data using neural network and rule-based classifiers," in *Proc. XXI Congr. ISPRS*, Beijing, China, 2008, pp. 1549–1554.

[31] L. Zhang, B. Zou, J. Zhang, and Y. Zhang, "Classification of polarimetric SAR image based on support vector machine using multiple-component scattering model and texture features," *EURASIP J. Adv. Signal Process.*, vol. 2010, pp. 1–9, Jan. 2010, DOI: 10.1155/2010/960831.

[32] L. Bruzzone and D. F. Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1171–1182, May 2000.

[33] P. Gamba, F. Dell'Acqua, and G. Lisini, "Change detection of multitemporal SAR data in urban areas combining feature-based and pixel-

based techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2820–2827, Oct. 2006.
[34] Y. Ban and O. A. Yousif, "Multitemporal spaceborne SAR data for urban change detection in China," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 5, pp. 1087–1094, Aug. 2012.
[35] Y. Ban, "Multitemporal ERS-1 SAR and Landsat TM data for agricultural crop classification: Comparison and synergy," *Can. J. Remote Sens.*, vol. 29, no. 4, pp. 518–526, Aug. 2003.
[36] C. Pohl and J. L. Van Genderen, "Multisensor image fusion in remote sensing: Concepts, methods and applications," *Int. J. Remote Sens.*, vol. 19, no. 5, pp. 823–854, May 1998.
[37] P. Gamba and B. Houshmand, "An efficient neural classification chain of SAR and optical urban images," *Int. J. Remote Sens.*, vol. 22, no. 8, pp. 1535–1553, May 2001.
[38] L. Gomez-Chova, D. Fernández-Prieto, J. Calpe, E. Soria, J. Vila, and G. Camps-Valls, "Urban monitoring using multi-temporal SAR and multi-spectral data," *Pattern Recognit. Lett.*, vol. 27, no. 4, pp. 234–243, Mar. 2006.
[39] B. Waske and J. A. Benediktsson, "Fusion of support vector machines for classification of multisensor data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 3858–3866, Dec. 2008.
[40] Z. Zhu, C. Woodcock, J. Rogan, and J. Kellndorfer, "Assessment of spectral, polarimetric, temporal, and spatial dimensions for urban and peri-urban land cover classification using Landsat and SAR data," *Remote Sens. Environ.*, vol. 117, pp. 72–82, Feb. 2011.
[41] D. Amarsaikhan, M. Ganzorig, and H. Blotevogel, "The integrated use of optical and InSAR data for urban land-cover mapping," *Int. J. Remote Sens.*, vol. 28, no. 6, pp. 1161–1171, Mar. 2007.
[42] D. Amarsaikhan, H. H. Blotevogel, J. L. van Genderen, M. Ganzorig, R. Gantuya, and B. Nergui, "Fusing high-resolution SAR and optical imagery for improved urban land cover study and classification," *Int. J. Image Data Fusion*, vol. 1, no. 1, pp. 83–97, Mar. 2010.
[43] C. Corbane, J.-F. Faure, N. Baghdadi, N. Villaneuve, and M. Petit, "Rapid urban mapping using SAR/optical imagery synergy," *Sensors*, vol. 8, no. 11, pp. 7125–7143, Nov. 2008.
[44] P. Griffiths, P. Hostert, O. Gruebner, and S. van der Linden, "Mapping megacity growth with multi-sensor data," *Remote Sens. Environ.*, vol. 114, no. 2, pp. 426–439, Feb. 2010.
[45] F. Pacifici, F. Del Prate, W. J. Emery, P. Gamba, and J. Chanussot, "Urban mapping using coarse SAR and optical data: Out come of the 2007 GRSS data fusion contest," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 331–335, Jul. 2008.
[46] T. T. Vu and Y. Ban, "Object-based fusion of ENVISAT ASAR and HJ-1 multispectral images for urban land cover mapping," in *Proc. ESA Living Planet Symp.*, H. Lacoste-Francis, Ed., Bergen, Norway, 2010.
[47] B. Waske and V. D. S. Linden, "Classifying multilevel imagery from SAR and optical sensors by decision fusion," *IEEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1457–1466, May 2008.
[48] T. Blaschke, "Object based image analysis for remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 65, no. 1, pp. 2–16, Jan. 2010.
[49] J. Fan, D. K. Y. Yau, A. K. Elmagarmid, and W. G. Aref, "Automatic image segmentation by integrating color-edge extraction and seeded region growing," *IEEE Trans. Image Process.*, vol. 10, no. 10, pp. 1454–1466, Oct. 2001.
[50] E. A. Carvalho, D. M. Ushizima, F. N. S. Medeiros, C. I. O. Martins, R. C. P. Marques, and I. N. S. Oliveira, "SAR imagery segmentation by statistical region growing and hierarchical merging," *Digit. Signal Process.*, vol. 20, no. 5, pp. 1365–1378, Sep. 2009.
[51] B. Sumengen and B. S. Manjunath, "Multi-scale edge detection and image segmentation," in *Proc. EUSIPCO*, Sep. 2005, pp. 4–7.
[52] S. S. Al-amri, N. V. Kalyankar, and S. D. Khamitkar, "Image segmentation by using edge detection," *Int. J. Comput. Sci. Eng.*, vol. 1, no. 3, pp. 804–807, May 2010.
[53] N. Senthilkumaran1 and R. Rajesh, "Edge detection techniques for image segmentation—A survey of soft computing approaches," *Int. J. Recent Trends Eng.*, vol. 1, no. 2, pp. 250–254, May 2009.
[54] X. Yu, J. Yla-Jaaski, O. Huttunen, T. Vehkomaki, O. Sipila, and T. Katila, "Image segmentation combining region growing and edge detection," in *Proc. 11th IAPR Int. Conf. Pattern Recognit., Vol.III. Conf. C, Image, Speech Signal Anal.*, 1992, pp. 481–484.
[55] P. Yu, A. K. Qin, and D. A. Clausi, "Unsupervised polarimetric SAR image segmentation and classification using region growing with edge penalty," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1302–1317, Apr. 2012.
[56] Q. Yu and D. A. Clausi, "IRGS: Image segmentation using edge penalties and region growing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 12, pp. 2126–2139, Dec. 2008.
[57] A. K. Qin and D. A. Clausi, "Multivariate image segmentation using semantic region growing with adaptive edge penalty," *IEEE Trans. Image Process.*, vol. 19, no. 8, pp. 2157–2170, Aug. 2010.
[58] M. Tabb and N. Ahuja, "Multiscale image segmentation by integrated edge and region detection," *IEEE Trans. Image Process.*, vol. 6, no. 5, pp. 642–655, May 1997.
[59] *CEOS EO Handbook—Mission Summary—HJ-1B*, CEOS, Washington, DC, 2011. [Online]. Available: http://database.eohandbook.com/database/missionsummary.aspx?missionID=464
[60] *eCognition Professional User Guide*, eCognition, Munich, Germany, 2011.
[61] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multi-class support vector machines," *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 415–425, Mar. 2002.

Yifang Ban (M'12) received the B.Sc. degree (with honors) in computer cartography and M.Sc. degree in remote sensing from Nanjing University, Nanjing, China, in 1984 and 1987, respectively, and the Ph.D. degree in remote sensing from the University of Waterloo, Waterloo, ON, Canada, in 1996. She received various scholarships and awards in China and Canada.

She was a Geographical Information System Analyst with PDM Information Technology Inc., Toronto, ON (1991–1992), and a Remote Sensing Consultant at Watts, Griffis and Mcquat Ltd. (1995–1996). She was a Senior Research Engineer with KTH Royal Institute of Technology, Stockholm, Sweden, during 1996–1997 and an Assistant Professor with Stockholm University, Stockholm, during 1998–2000. Since 2004, she has been the Chair Professor of geoinformatics with KTH Royal Institute of Technology. Before joining KTH, she was a tenured Associate Professor with York University, Toronto. She is a Reviewer for a number of major international journals. Her current research interests include synthetic aperture radar image analysis and classification, image segmentation, multisensor data fusion, urban land-cover mapping, and change detection.

Prof. Ban is the Chair of the International Society of Photogrammetry and Remote Sensing Inter-Commission II/IV/VIII Working Group on Global Land Cover Mapping and Services. She is also the Swedish Representative to EARSeL and the Cochair of the EARSeL Special Interest Group "Temporal Analysis of Satellite Images."

Alexander Jacob (S'13) received the M.Sc. degree in geodesy and geoinformatics from KTH Royal Institute of Technology, Stockholm, Sweden, in 2011, where he is currently working toward the Ph.D. degree in geoinformatics.

Before he joined KTH, he worked as a Server Site Developer in the Internet startup scene of Berlin, where he learned programming on a professional level. His research interests include image processing, segmentation, multisensor data fusion, and software development.