Object Recognition Based on the Context Aware Decision-Level Fusion in Multiviews Imagery

Fatemeh Tabib Mahmoudi, *Fellow, IEEE*, Farhad Samadzadegan, and Peter Reinartz, Jr., *Member, IEEE*

***Abstract*—Spectral similarities and spatial adjacencies between various kinds of objects, shadow, and occluded areas behind high-rise objects as well as the complex relationships between various object types lead to the difficulties and ambiguities in object recognition in urban areas. Using a knowledge base containing the contextual information together with the multiviews imagery may improve the object recognition results in such a situation. The proposed object recognition strategy in this paper has two main stages: single view and multiviews processes. In the single view process, defining region's properties for each of the segmented regions, the object-based image analysis (OBIA) is performed independently on the individual views. In the second stage, the classified objects of all views are fused together through a decision-level fusion based on the scene contextual information in order to refine the classification results. Sensory information, analyzing visibility maps, height, and the structural characteristics of the multiviews classified objects define the scene contextual information. Evaluation of the capabilities of the proposed context aware object recognition methodology is performed on two datasets: 1) multiangular Worldview-2 satellite images over Rio de Janeiro in Brazil and 2) multiviews digital modular camera (DMC) aerial images over a complex urban area in Germany. The obtained results represent that using the contextual information together with a decision-level fusion of multiviews, the object recognition difficulties and ambiguities are decreased and the overall accuracy and the kappa are gradually improved for both of the WorldView-2 and the DMC datasets.**

***Index Terms*—Contextual information, decision-level fusion, object recognition, visibility analysis.**

I. Introduction

RECENT YEARS' advances in airborne and spaceborne sensor technology and digital imaging techniques lead to the very-high resolution (VHR) remotely sensed data, those provide geo-information for automatic recognition of objects in complex urban areas. Increasing the spectral heterogeneity at VHR data leads to more within-class variances, less interclass variances, and the inadequacy of the traditional pixel-based classification approaches [1]–[6]. Already many researchers have investigated the potential of the object-based image analysis approaches for dealing with VHR imagery and the complexities in urban areas [2], [3], [5], [7]–[9].

Manuscript received May 22, 2014; revised September 28, 2014; accepted September 30, 2014.

F. Tabib Mahmoudi and F. Samadzadegan are with the Department of Geomatics, Faculty of Engineering, University of Tehran, Tehran 11365-4563, Iran (e-mail: fmahmoudi@ut.ac.ir; samadz@ut.ac.ir).

P. Reinartz is with the Department of Photogrammetry and Image Analysis, Remote Sensing Technology Institute, German Aerospace Centre (DLR), Oberpfaffenhofen 82234 Weßling, Germany (e-mail: peter.reinartz@dlr.de).

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/JSTARS.2014.2362103

As it is depicted in the previous researches in the field of OBIA, the accuracy of the object recognition results in complex urban areas directly depends on the segmentation and knowledge-based classification processes [3], [5], [9]–[11]. Moreover, depending on the viewing angle of the sensor, some parts of urban objects may be occluded by their adjacent high-rise objects such as buildings or trees [12], [13]. Therefore, using only single shot imagery, it is very difficult to obtain valuable object recognition results. The fusion of the information coming from multiviews aerial or satellite imagery is valuable for filling up occluded areas and obtaining reliable object recognition results. Fusion of multiviews imagery together with the contextual information may be considered as a solution to enhance completeness and accuracy of the object recognition results in complex urban areas.

A. Object Recognition Based on Multisource Information Fusion

Data fusion appears as an effective way for a synergistic combination of information from various sources in order to provide a better understanding of a given scene [14]–[17]. Many researchers have investigated the potential of performing various pixel, feature, and decision-level fusion algorithms for integrating multisource remotely sensed data [6], [17]–[20]. In [20], multiangle imaging spectroradiometer (MISR) data collected in four bands and at nine view angles are fused in pixel level with Landsat data in Shenzhen, China. This fused data were used to demonstrate the view-angle effects on the spectral response and discrimination of the urban land cover types.

Ran *et al.* [19] present a decision-level fusion method to produce a higher accuracy land cover map by combining multisource local data based on the Dempster–Shafer evidence theory. In decision-level fusion, single source images are processed independently and their decision outcomes are combined using weights of significance. As an advantage, decision-level fusion does not depend on the type and source of information such as images, maps, and databases. In [19], the primary objective of the land cover mapping based on the decision-level fusion is to facilitate the extraction of biogeophysical information from land cover for use in regional and global modeling studies. The results of the fusion validation analyses show a great improvement in accuracy in comparison with other land cover maps. In [6], a multilevel classification system is used for integrating the pixel-based structural features and the object-based shape features based on a decision-level probability fusion approach. In the first level, the multispectral and the structural features based

on the morphological transformation are separately fed into two support vector machines, respectively. The spectral–structural pixel-based decision fusion is carried out by considering the weighted probabilistic outputs of the two SVMs. At the second level, the object-based decision fusion is implemented by considering the probabilistic outputs of level I within the boundary of each object obtained from performing mean shift segmentation. The results showed that the extension of pixel-level to object-level decision fusion improved the accuracies of classification by 0.8% and 8.3% for the hyperspectral and QuickBird datasets, respectively.

B. Object Recognition Based on Contextual Information

Context can be defined as any information that is not directly produced by the appearance of an object. Context can be obtained from the nearby image data and neighboring objects and comprises any kind of relations between the semantic entities present in an image. Therefore, context can provide additional information to disambiguate appearance inputs in the object recognition tasks [21]–[25].

According to different levels of the utilized information, context can be further categorized concerning various points of view. In some researches, this categorization concerned several interaction levels; pixel interactions based on the notion that neighboring pixels tend to have similar labels, region interactions between image patches/segments, and object interactions between various objects represented in the scene [22], [23]. Context also can be categorized into local and global. Local context concerns relations derived from the area that surrounds the object to be detected. Global context includes information about the overall spatial layout of the image and concerns objects' occurrences and co-occurrences in the scene [22]. Hermosilla *et al.* [26] aim to define and analyze context-based descriptive features for classifying land-use in urban environments using three different object aggregation levels: object-based, internal context, and external context. Object-based features are composed of image, geometrical, and three-dimensional (3-D) features. Internal context features describe an object with respect to the subobjects contained within it. External context features characterize each object by considering the common properties of its adjacent objects. Their results of the classification tests show that the internal and external context features suitably complement the image-derived and 3-D features, improving the classification accuracy values especially between those classes with similar image and 3-D feature patterns.

Guo *et al.* [23] represent an object-based classification procedure for high-resolution images by exploiting different levels of the contextual information. The contexts used in each stage were the object's inner context (i.e., the gray constraints of different pixels in an object), the object's neighbor context (i.e., the characteristic constraint of different objects adjacent to the object of interest), and the object's scene context (i.e., the spatial homogeneity of the label distribution for different image objects and the consistency of feature distributions for different object classes in the whole scene), respectively. Their experimental results, which are based on a complex urban area, confirm that classification using high-level context yields a greater improvement than that based on the results provided by the low-level context.

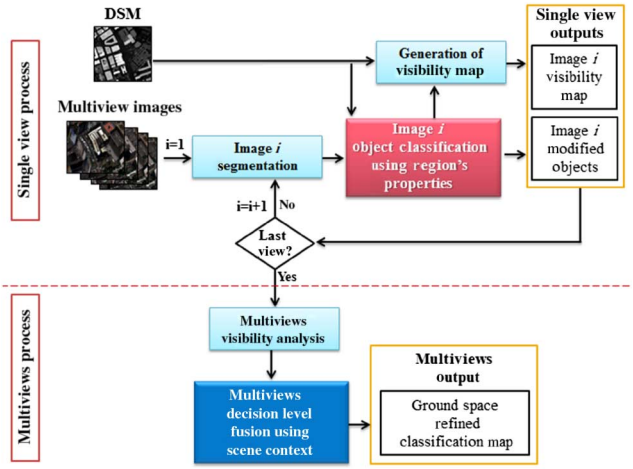

Fig. 1. General structure of the proposed object recognition strategy.

The novelty of our proposed object recognition methodology concerns the combination of the contextual information and decision-level fusion strategy in order to solve the difficulties related to generating a complete classification map. The following items introduce some of the most important aspects in the proposed method:

1) decision-level fusion of the results of the multiviews object-based image analysis together with the digital surface model as scene space in order to enhance the completeness and accuracy of the classification map;
2) considering various aspects of each image such as sensory information, occluded areas, size and shape of the object regions, and topological relationships between objects for assigning weights to each of the object classes;
3) estimating the real object type in the shadow and occluded areas by integrating the calculated weights of all predefined object classes and using topological relationships between adjacent object regions.

II. Context Aware Object Recognition

In this paper a context aware object recognition strategy composed of multiple steps is proposed for solving object recognition difficulties in complex urban areas, based on multiviews VHR remotely sensed imagery and digital surface model. As depicted in Fig. 1, the object recognition method is composed of two main stages: single view process and multiviews process.

In the first stage, the properties of each segmented region are utilized for object classification on the individual images. Per segment spectral and textural characteristics together with the structural features based on the size, shape, and height of a segmented region generate region's properties. The second stage of the proposed methodology performs decision-level fusion on the multiviews classified regions based on the scene context in order to reduce the ambiguities and uncertainties in the generated classification map.

A. Single View Process

Object-based image analysis requires generating segmented regions as the classification units. In this research, multiresolution segmentation technique is applied to the content of each of the individual images in order to segment it into regions. The multiresolution segmentation algorithm starts with single image objects of one pixel and repeatedly merges a pair of image objects into larger ones. The merging decision is based on the local homogeneity criterion, describing the similarity between adjacent image objects [27]. After performing segmentation, a knowledge-based classification process should be performed on each of the segmented regions. Therefore, it is necessary to gather proper knowledge composed of per segment spectral and textural features and structural characteristics of each segmented region in order to provide region's properties.

1) Region's Properties: Per segment spectral and textural characteristics together with the structural features of each segmented region can provide region's properties that are the main tool for performing knowledge-based classification [23], [26].

1) *Spectral features*: The ratios between the reflectance values of every possible combination of spectral bands are called index ratios; those can be used to establish the spectral characteristics of a region. According to the spectral capabilities of the remotely sensed data, in this paper, normalized difference indices (NDI) and simple ratios (SR) are used for proper definition of spectral features based on the mean values of the spectral bands (Bandl, Bandk) over the regions

$$\mathrm{NDI}_{k,l} = (\mathrm{Mean}_{\mathrm{Band}_k} - \mathrm{Mean}_{\mathrm{Band}_l}) / (\mathrm{Mean}_{\mathrm{Band}_k} + \mathrm{Mean}_{\mathrm{Band}_l}) \quad (1)$$

$$\mathrm{SR}_{k,l} = \mathrm{Mean}_{\mathrm{Band}_k} / \mathrm{Mean}_{\mathrm{Band}_l}. \quad (2)$$

2) *Textural features*: Textural features can be measured based on the gray value relationships between pixels over the entire preidentified segmented region. In this paper, referring to the complexities of the object recognition in urban areas, more than one feature is used for identifying the textural characteristics of the objects. Table I represents these features with some basic mathematics for each of them.

3) *Structural features*: Calculating structural features based on the spatial characteristics and heights of segmented regions provide another part of the region's properties for using in the object classification process. In this paper, 2-D structural features such as area, rectangularity, elongation, roundness, and solidity are used together with the mean height value of each segment as 3-D structural feature; those have high potential in recognizing objects in complex urban areas. Table II represents these features with their basic mathematics.

After generation of the above-mentioned spectral, textural, and structural features based on the image data and DSM, for generating a rich knowledge base of region's properties, an optimum feature selection is performed. The threshold values for the optimum features are determined by the combination of expert knowledge and quantitative analysis. In this phase, optimum spectral and textural feature selections are based on the capabilities of the input spectral bands and various textural features for recognition of each individual object types. Optimum structural feature selection is based on the analysis of relations between features and objects, for instance, the relation between elongation and road objects or rectangularity and building objects.

TABLE I
BASIC MATHEMATICS OF TEXTURAL FEATURES

Name	Formula
GLCM energy	$\sum_{i=0,j=0}^{N-1} \mathrm{GLCM}(i,j)^2$
GLCM entropy	$\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} \mathrm{GLCM}(i,j)\mathrm{Log}(\mathrm{GLCM}(i,j))$
GLCM contrast	$\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} \mathrm{GLCM}(i,j)\lvert i-j\rvert$
GLCM homogeneity	$\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} \frac{\mathrm{GLCM}(i,j)}{1+\lvert i-j\rvert}$
Semi-Variogram	$\frac{1}{2n_{(h)}}\sum_{i=1}^{n(h)}\left((I(x_i)-I(x_i+h))^2\right)$

GLCM (i, j), value of the pixel (i, j) in GLCM space; N, segment size as local window; n(h), number of pixels in a segmented region in the distance lag h from the central pixel of the region; xi, central pixel of the segmented region; I(xi), value of the image at pixel xi.
GLCM: In gray-level co-occurrence matrix, according to the definition of a local neighboring window (segmented region) and proper orientation selection, the relationships between the gray values of the pixels transform to the co-occurrence matrix space. Sum of the pixel values in GLCM space for each segment after performing the statistical operations on them is assigned to the segment as textural attributes.

TABLE II
BASIC MATHEMATICS OF 2-D STRUCTURAL FEATURES

Name	Formula
Rectangularity	$\mathrm{Rectangularity} = \mathrm{Area}/(\mathrm{MajorLength} * \mathrm{MinorLength})$
Elongation	$\mathrm{Elongation} = \mathrm{MajorLength}/\mathrm{MinorLength}$
Solidity	$\mathrm{Solidity} = \mathrm{Area}/\mathrm{ConvexArea}$
Roundness	$\mathrm{Roundness} = \frac{4\times\mathrm{Area}}{\pi\times\mathrm{Length}^2}$

Rectangularity indicates how well a shape is described by a rectangle having major and minor lengths. Elongation indicates the ratio of the major axis of the bounding rectangle to the minor axis of it. Solidity compares the area of the region to the area of the smallest convex polygon that can contain the region. Roundness indicates the ratio of region's area to the length of its bounding rectangle.

The object classification can be performed by encapsulating the knowledge base into a rule set and the definition of a strategy for object recognition. The proposed strategy is a multiprocess classification model that is based on the spectral, textural, and structural reasoning, respectively. Table III shows the structure of reasoning rules in different steps of the recognition strategy.

Despite the high potential of the strategies in the object-based image analysis, the object classification results based on the

TABLE III
STRUCTURE OF REASONING RULES IN THE PROPOSED OBJECT CLASSIFICATION SCHEME

Spectral reasoning rules	If NDI(bandi,bandj)s< TNDI And SR(bandi,bandj)s< TSR Then Spectral Candidate Object s ==Class K1
Textural reasoning rules	If Spectral Candidate Object s ==Class K1 And Entropy s <= TEntropy And Homogeneity s <= Thomogeneity Then Textural Candidate Object s ==Class K2
Structural reasoning rules	If Textural Candidate Object s == Class K2 And Area of Region s >= TArea And Elongation of Region s >= TElongate OR Height of Region s >= THeight Then Final Candidate Object s == Class K3

s, indicator of a segmented region; T_{NDI}, predefined threshold for NDI(bandi,bandj); T_{SR}, predefined threshold for SR(bandi,bandj); $T_{Entropy}$, predefined threshold for entropy; $T_{homogeneity}$, predefined threshold for homogeneity; T_{Area}, predefined threshold for area of the region; $T_{Elongation}$, predefined threshold for elongation of the region; T_{Height}, predefined threshold for height of the region.
Object classes K1, K2, and K3 can be the same or not.

region's properties still cannot deal with the uncertainties and ambiguities related to the shadow and occlusion in the complex urban areas. Using higher levels of the contextual information related to the objects' occurrences in the whole scene and considering the classification accuracies in multiviews may increase the reliability of the classification map. Therefore, as the final process in the first stage of the algorithm, the visibility map should be generated for each of the individual images using the digital surface model in the ground space. Generation of the visibility maps is based on detecting the areas occluded by the high-rise natural or man-made structures in each of the individual views. As depicted in (3), shown at the bottom of the page (the rules are the simplifications of the real set of rules), considering the recognized 3-D urban objects such as buildings and trees in the results of the object-based image analysis, one can detect occluded areas by analyzing height values and off-nadir angles, as shown at the bottom of the page.

In which, Threshold3D Object Class j is the predefined minimum height of 3-D objects in class j and threshold view angle is the largest reliable off-nadir angle for images considering the line of sight from sensor to the image. If the prerecognized 3-D object i has the reasonable height in 3-D object class j, heights of its surrounding 3-D objects should be considered together with the viewing angle of the sensor in order to reconstruct the line of sight from the sensor to the object and estimate the state of visibility. Individual visibility maps will be used for further analysis in the second stage of the proposed method.

B. Multiviews Process

High-rise objects together with the viewing angle of the sensor make some uncertainties in the object classification results which cannot be solved using only single shot imagery. Therefore, in the proposed object recognition methodology, another processing stage is defined based on the context aware decision-level fusion of the object-based image analysis on multiviews. This stage is composed of two main operations: preanalysis on the object classification results for all images and the decision-level fusion of the multiviews.

1) Preanalysis on Classified Regions: The preanalysis is based on generating total visibility map from the summation of all of the individual visibility maps as the main tool for performing visibility analysis. Visibility analysis determines the number of views the pixel (x, y) of the ground space is visible in all of them. Performing visibility analysis, one can categorize all of the ground space pixels into three groups: visible in all images, visible in some images, and visible in none of the (occluded in all) images. This categorization is useful for the definition of the scene context in the proposed decision-level fusion algorithm.

2) Decision-Level Fusion: For the decision-level fusion of the classified regions in multiviews, results of the visibility analysis are utilized through a higher level context aware strategy. The first step of the decision-level fusion is the back projection from each ground space pixel to its preidentified visible images (all images for the group of visible in none) based on the results of performing visibility analysis. These back projections find the classified regions in multiviews that pixels belong to.

As scene context of an object can be defined in terms of its co-occurrences with other objects and its occurrences in the whole scene, information regarding sensor's look angle, distance from occluded areas, heights, and areas of the object regions that pixels belong to are the fundamentals for the scene context definition. In this research, scene contextual information is utilized for weighting all object classes in order to assign them to the ground space pixels. The decision-level fusion strategy takes place on the level of the predefined object classes based on the object classification results. Therefore, if there are n various recognizable object classes in the multiviews, also n different weights should be calculated for assigning the

$$\begin{cases} \text{If View} = \text{View}_{\text{K}} \\ \bullet\ \text{If Object}_{\text{i}} \text{ belongs to 3D Object Class}_{\text{j}} \text{ AND Height}_{\text{i}} < \text{Threshold}_{\text{3D Object Class j}} \\ \quad \text{AND Off} - \text{nadir Angle}_{\text{K}} > \text{Threshold}_{\text{View Angle}} \text{ Then Visibility Map} = 0 \\ \bullet\ \text{If Object}_{\text{i}} \text{ belongs to 3D Object Class}_{\text{j}} \text{ AND Height}_{\text{i}} \geq \text{Threshold}_{\text{3D Object Class j}} \\ \quad \text{AND Height}_{\text{i}} \geq \text{Height}_{\text{Surrounding Objects}} \text{ Then Visibility Map} = 1 \\ \bullet\ \text{If Object}_{\text{i}} \text{ belongs to 3D Object Class}_{\text{j}} \text{ AND Height}_{\text{i}} \geq \text{Threshold}_{\text{3D Object Class j}} \\ \quad \text{AND Height}_{\text{i}} < \text{Height}_{\text{Surrounding Objects}} \text{ AND Off} - \text{nadir Angle}_{\text{K}} > \text{Threshold}_{\text{View Angle}} \\ \quad \text{Then Visibility Map} = 0 \end{cases} \tag{3}$$

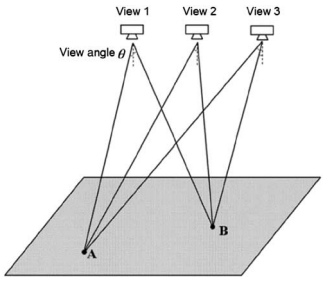

Fig. 2. Illustration of the angle-based rule [28].

object classes to each of the ground space pixels. Following items represent the components for defining scene contextual information:

1) *Sensor weight (W Sensor)*: This weight depends on the off-nadir viewing angle (θ) for each of the multiview images. According to the angle-based rule for composing multiview images, the nearest to nadir view has the most reliability because of taking the values of nonoccluded regions as vertically as possible [28]. Object classes in the nearest to nadir view have the largest sensor weights for assigning to the ground space pixels (Fig. 2).
2) *Weight of occlusion (W Occlusion)*: It depends on the neighborhood relations between each object region and the occluded areas in the multiviews visibility maps [29]. Having less common borders with occluded areas leads to the less uncertainty and larger weight of occlusion for each object region.
3) *Weight of area (W Area)*: It depends on the area of the object regions. If the ground space pixel (x, y) belongs to a large object region, the object class of the large region has more reliability and the larger weight of area assigns to it. The reason of this rule is the uncertainties of small regions. For instance, if a small road region is completely contained by the building object class, it can be caused as an error by the spectral similarities between road and building roofs.
4) *Weight of topological relationships (W Topology)*: For an occluded region in all views, it depends on its neighboring relationship with the nearest visible object regions. Using the weight of topological relationship, the class label of the neighboring visible object region with the smallest height difference assigns to the occluded region in all views. The simplification of the real developed rules for topological relationships is depicted in (4), shown at the bottom of the page.

In which, $\mathrm{Threshold}_{\Delta \mathrm{H}}$ is the predefined maximum height difference between neighboring regions and $\mathrm{Threshold}_{\mathrm{Area}}$ is the predefined minimum area for large regions. As it is depicted in (4), if there is a small amount of height differences between neighboring object regions ($\mathrm{W\ Topology} = 1$), the class label of the neighbor visible region with the smallest height difference should be assigned to the occluded one.

For the ground space pixel (x, y), which is categorized in the groups of visible in all or visible in some images, calculating the weight of scene contextual information for assigning object class i to this pixel is based on the summation of sensor, occlusion, and area weights in the visible views, those classified pixel (x, y) into the object class i (5). For the ground space pixel (x, y), which is visible in none of the images, the weight of topological relationships and the sensor weight in all views are utilized for calculating the weight of scene context for each of the object classes in order to assign the neighboring visible object class i to this occluded pixel (6)

$$\mathrm{W(Scene\ Context)}^{(\mathrm{x,y})}_{\mathrm{Class\ i}_{\mathrm{i}\in[1,\mathrm{n}]}} = \sum_{\mathrm{k}\in \mathrm{visibleviews}} \left(\mathrm{W}_{\mathrm{Sensor_k}} + \mathrm{W}_{\mathrm{Occlusion_k}} + \mathrm{W}_{\mathrm{Area_k}}\right)_{(\mathrm{x,y})\in \mathrm{class\ i}} \tag{5}$$

$$\mathrm{W(Scene\ Context)}^{(\mathrm{x,y})}_{\mathrm{Class\ i}_{\mathrm{i}\in[1,\mathrm{n}]}} = \sum_{\mathrm{k}\in \mathrm{Allviews}} \mathrm{W}_{\mathrm{Topology_k}} * (\mathrm{W}_{\mathrm{Sensor_k}}). \tag{6}$$

In addition, classification accuracies of various object types in each view also affect on the decision-level fusion results [6]. Classification weights of the object classes are determined based on the user and producer accuracies in the results for each of the multiviews object-based image analysis

$$\mathrm{W(Classification)}^{\mathrm{K}}_{\mathrm{Class\ i}} = \frac{2\mathrm{A}^{\mathrm{k}}{}_{\mathrm{U(Class\ i)}} \times \mathrm{A}^{\mathrm{k}}{}_{\mathrm{P(Class\ i)}}}{\mathrm{A}^{\mathrm{k}}{}_{\mathrm{U(Class\ i)}} + \mathrm{A}^{\mathrm{k}}{}_{\mathrm{P(Class\ i)}}} \tag{7}$$

where $\mathrm{W(Classification)}^{\mathrm{K}}_{\mathrm{Class\ i}}$ is the classification weight for object class i in view k and $\mathrm{A}^{\mathrm{k}}_{\mathrm{U(Class\ i)}}$ and $\mathrm{A}^{\mathrm{k}}_{\mathrm{P(Class\ i)}}$ are the user and producer accuracies for the object class i in view k. For calculating the producer accuracy, the total number of correct classified pixels in an object class is divided by the total

$$\begin{cases} \mathrm{If\ Region_K = Occluded\ in\ All\ Views} \\ \bullet\ \mathrm{If\ Region_K\ Compeletely\ Contained\ By\ Visible\ Regions} \\ \quad \mathrm{AND\ Minimum(\Delta Height(Region_K,\ Neighboring\ Visible\ Regions)) \le Threshold_{\Delta H}} \\ \quad \mathrm{Then\ W_{Topology} = 1(Assign\ Label\ of\ Neighbor\ Visible\ Region\ with\ Minimum(\Delta Height))} \\ \bullet\ \mathrm{If\ Region_K\ Partialy\ Contained\ By\ Visible\ Regions} \\ \quad \mathrm{AND\ Minimum(\Delta Height(Region_K,\ Neighbouring\ Visible\ Regions)) \le Threshold_{\Delta H}} \\ \quad \mathrm{AND\ Area_{Region_K} > Threshold_{Area}} \\ \quad \mathrm{Then\ W_{Topology} = 1(Assign\ Label\ of\ Neighbor\ Visible\ Region\ with\ Minimum(\Delta Height))} \\ \bullet\ \mathrm{Else\ W_{Topology} = 0} \end{cases} \tag{4}$$

number of reference pixels of that object class. If the number of correct classified pixels in an object class is divided by the total number of classified pixels in that class, the measure of commission error or user accuracy is obtained. Reference maps in each datasets (Fig. 6) are used for calculating the user and producer accuracies in each of the individual object-based image analysis results.

Therefore, the total weight calculation for each of the object classes is based on the weights of scene context and classification in multiviews. The object class with the largest total weight should be selected as the winner class label for assigning to the ground space pixel (x, y)

$$\text{Total Weight}_{\text{Class i}}^{(x,y)} = \sum_{\substack{k=\text{Visible Views} \\ (x,y)\in\text{class i}}} \text{W(Classification)}_{\text{Class i}}^{K} \times \text{W(Scene Context)}_{\text{Class i}}^{(x,y)} \quad (8)$$

$$\text{Winner Class}_{(x,y)} = \text{Max(Total Weight}_{\text{Class i}}^{(x,y)}) \quad i \in 1, 2, \ldots, n. \quad (9)$$

If the winner class is shadow, the structural and height-based relations are used in order to determine the true object types instead of the shadow area. According to the rules depicted in (10), shown at the bottom of the page, those are simplifications of the real developed rules, determining the true object type in a shadow region also depends on the class labels of its neighboring visible regions with the small amount of height differences.

According to (10), for each of the shadow regions, considering its area, the visible neighbor object class with the longest neighboring border, with respect to the predefined threshold ($\text{Threshold}_{\text{Border}}$), should be used for analyzing the height and 2-D structural features such as rectangularity and roundness. If the height difference between the investigating shadow region and its longest visible neighboring object class is below a predefined threshold ($\text{Threshold}_{\Delta\text{H}}$), the class label of this neighboring object is assigned to the shadow region.

III. Experiments and Results

A. Dataset

The potential of the proposed methodology is evaluated for automatic object recognition based on two multiviews datasets. The first dataset is the multiangular pan-sharpened WorldView-2 satellite imagery over Rio de Janeiro (Brazil) which was collected in January 2010 with half a meter spatial resolution and 8 spectral bands and within a 3-min time frame with satellite elevation angles of 44.7° and 56.0° in the forward direction, and 59.8° and 44.6° in the backward direction. The multiangular sequence contains the downtown area of the city, including a number of large and high buildings, commercial and industrial structures, and a mixture of community parks and private housing. Moreover, using multiangular WorldView-2 images, the DSM with a grid width of 50 cm is generated from multiple pairs of panchromatic stereo images—in epipolar geometry—using the Semi-Global Matching (SGM) algorithm [30].

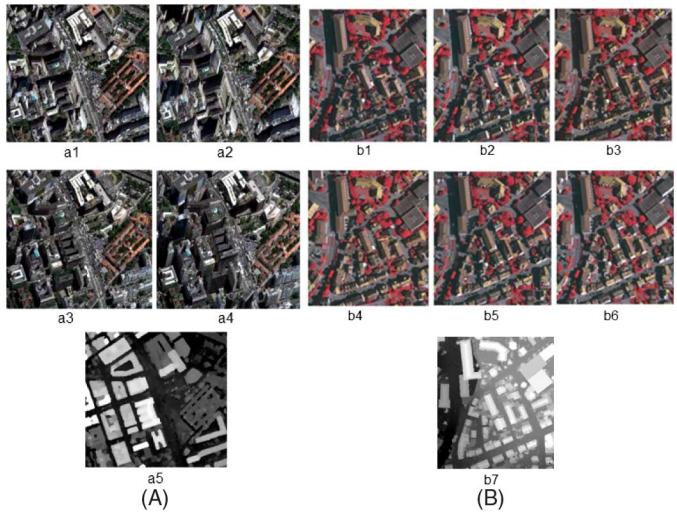

Fig. 3. (A) WorldView-2 dataset. (a1–a4) Four multiangular WorldView-2 satellite images. (a5) Generated digital surface model from matching. (B) The DMC dataset. (b1–b6) Six multiviews DMC aerial images. (b7) Lidar DSM.

The other dataset contains multiviews digital aerial imagery over Vaihingen in Germany. The aerial images were acquired using an Intergraph/ZI DMC on 24 July and 6 August, 2008 [31]. As depicted in Fig. 3, a DSM that is interpolated from the airborne laser scanner (ALS) point cloud with a grid width of 25 cm, using only the points corresponding to the last pulse, is used together with the six overlapping pan-sharpened color infrared aerial images with a ground sampling distance of 8 cm and a radiometric resolution of 11 bits. The sample area is situated in the center of the city of Vaihingen. It is characterized

$$\begin{aligned}
&\text{If } \text{Region}_{\text{i}} = \text{Shadow} \\
&\bullet\ \text{If } \text{Area}_{\text{Shadow i}} < \text{Threshold}_{\text{Area}} \text{ AND Neighboring Border}(\text{Shadow}_{\text{i}}, \text{Object}_{\text{j}}) > \text{Threshold}_{\text{Border}} \\
&\quad \text{AND } \Delta\text{Height}(\text{Object}_{\text{j}}, \text{Shadow}_{\text{i}}) < \text{Threshold}_{\Delta\text{H}} \\
&\quad \text{Then True Object Class of Shadow}_{\text{i}} = \text{Class Label}_{\text{j}} \\
&\bullet\ \text{If } \text{Area}_{\text{Shadow i}} \geq \text{Threshold}_{\text{Area}} \text{ AND Neighboring Border}(\text{Shadow}_{\text{i}}, \text{Object}_{\text{j}}) > \text{Threshold}_{\text{Border}} \\
&\quad \text{AND } \Delta\text{Height}(\text{Object}_{\text{j}}, \text{Shadow}_{\text{i}}) < \text{Threshold}_{\Delta\text{H}} \\
&\quad \text{AND Rectangularity}_{\text{Shadow i}} < \text{Threshold}_{\text{Rectangularity}} \text{ AND Roundness}_{\text{Shadow i}} < \text{Threshold}_{\text{Roundness}} \\
&\quad \text{Then True Object Class of Shadow}_{\text{i}} = \text{Class Label}_{\text{j}}.
\end{aligned} \quad (10)$$

TABLE IV
OPTIMUM SELECTED FEATURES AND THEIR THRESHOLDS

Optimum features		Thresholds based on object classes				
		Building	Road	Tree	Grass	Shadow
WV-2						
Spectral	NDI(RE, CB)	>-0.38 & <-.06			>=-0.58 & <=-0.3	
	NDI(R, RE)	<0.4				
	SR(G, R)			<1.3		<1.14
	SR(NIR1,Y)					>=1.6
	SR(NIR2, RE)		>=1.25 &<=1.4	>2.8		
Textural	GLCM Hom	>0.45	>=0.27 & <=0.4	<0.3		<0.2
	GLCM Ent	>=4.8 & <=5.5	>=5.49 & <=6			
	GLCM Con	>=50& <=190	>=30& <=44			
Structural	Rect	>0.4	<=0.4			
	Round	<2	>=2& <=4.5	<2	<2	

Optimum features		Thresholds based on object classes				
		Building	Road	Tree	Grass	Shadow
DMC						
Spectral	NDI(IR, R)			>=0.14	>=0.14 & <=0.4	
	SR(G, IR)				>=0.99	
	SR(R, IR)				>=1.1& <=2	
Textural	NDI(IR, R)	>=0.99	<0.99			
	GLCM Hom	>0.45	>=0.27 & <=0.4	<0.3		<0.2
	GLCM Ent	>=5.3 & <=5.4	>=4.8& <=6.55			
	GLCM Con	>=50& <=190	>=35& <=44			
Structural	Rect	>=0.62	<0.62			<0.62
	Round	<2	>=2			

by dense development consisting of historic buildings having rather complex shapes, but also has some trees.

B. Obtained Results

In the stage of performing single view process, the multiresolution segmentation algorithm is applied to the content of each of the individual images using eCognition software. For the WorldView-2 dataset, the values 90, 0.2, and 0.1 and for the DMC dataset, the values 150, 0.2, and 0.5 are used for the scale parameter, compactness, and shape parameters, respectively. These parameters were determined based on the trial and error in order to provide suitable size of the segments for recognizing various object types in both datasets. Then, all of the spectral, textural, and structural features mentioned in Section II-A1 are measured on all of the image regions on the whole datasets. Optimum feature selection and threshold setting for each object class is performed semiautomatically by an expert operator using the quantitative and visual analysis on each of the features in feature view of eCognition software. Table IV depicts the optimum-selected features and their estimated thresholds based on all images in both datasets, for the generation of the knowledge base and performing object-level classification on the segmented regions (see Section II-A1 for descriptors).

As depicted in Table IV, building, road, tree, grassland, and shadow area are the preidentified object classes based on the visual inspections. Despite shadow being not a real object class, detecting true objects instead of shadow areas based on the spectral responses is a difficult task dealing with VHR imagery. Therefore, in processing based on single view, shadow is recognized as a separate object class, and in a later step, we are going to recover shadow areas based on the decision-level fusion of topological relationships in multiviews processes.

TABLE V
VISIBILITY ANALYSIS ON THE TOTAL VISIBILITY MAP

Dataset	Visible in all views (%)	Visible in some views (%)	Occluded in all views (%)
WV-2	62.42	33.98	3.52
DMC	66.35	31.85	1.65

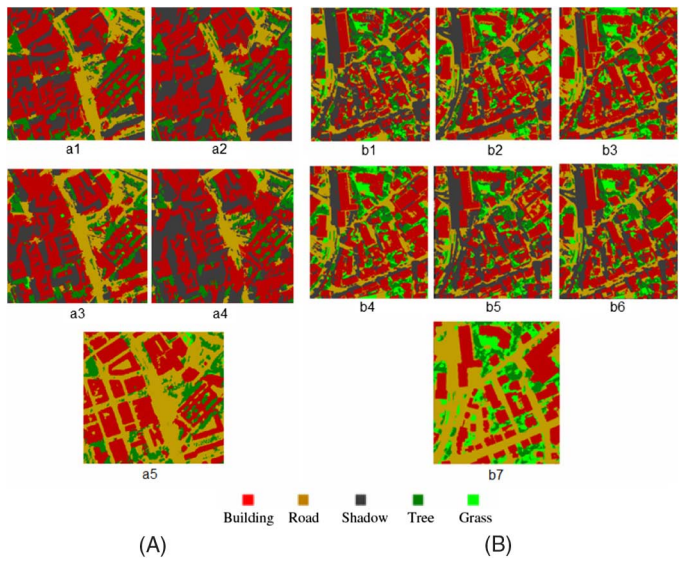

Fig. 4. (A) Results on the WorldView-2 dataset. (a1–a4) Object-based image analysis on individual WV-2 satellite images. (a5) Final decision-level fusion object recognition results. (B) Results on the DMC dataset. (b1–b6) Object-based image analysis on individual DMC aerial images. (b7) Final decision-level fusion object recognition results.

According to the different characteristics of the ground space (which is represented by the DSM) and image spaces (individual object-based image analysis results), performing projection is necessary for all processing steps that need the

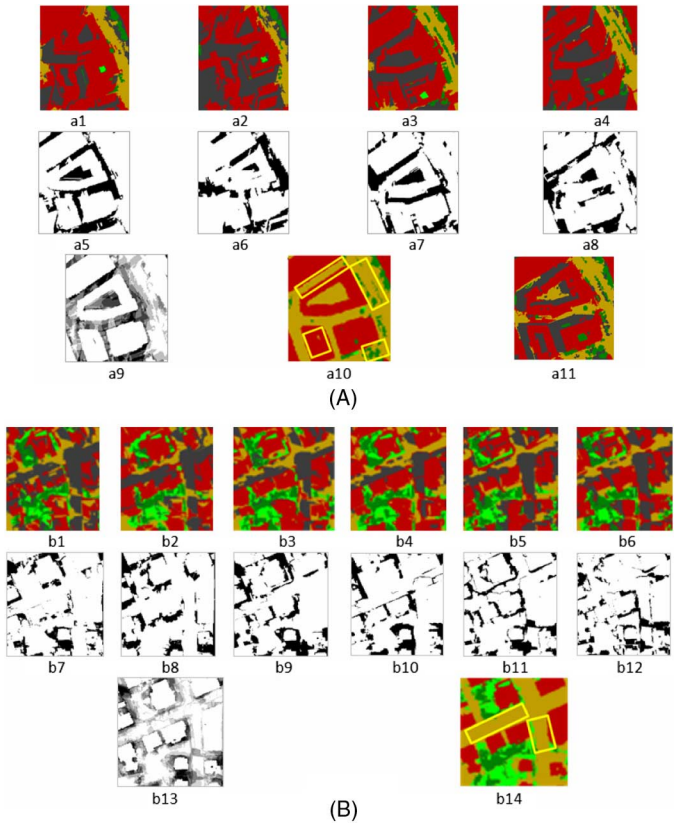

Fig. 5. (A) Results on the sample patch of the WorldView-2 dataset. (a1–a4) OBIA on individual WV-2 satellite images with off-nadir angles 30.2°, 45.4°, 34.00°, and 45.3°, respectively. (a5–a8) Visibility maps on individual WV-2 satellite images. (a9) Total visibility map. (a10) Final decision-level fusion object recognition result. (a11) OBIA on the most nadir view with 8.6° off-nadir angle. (B) Results on the sample patch of the DMC dataset with off-nadir angles 50.52°, 50.43°, 11.15°, 14.63°, 42.00°, and 43.37°, respectively. (b1–b6) OBIA on individual DMC aerial images. (b7–b12) Visibility maps on individual DMC aerial images. (b13) Total visibility map. (b14) Final decision-level fusion object recognition result (yellow boxes highlight improvements).

transformation between the ground space and each of the image spaces (such as OBIA structural reasoning, decision-level fusion algorithm, and later for the accuracy assessment). In the WorldView-2 dataset the projection is performed using the rational polynomial coefficients (RPC) and in the DMC dataset, the photogrammetric linearity equation is used based on the provided exterior orientation parameters for each of the images. As it is depicted in Table V, before decision-level fusion of multiviews, by performing visibility analysis on the total visibility map, ground space pixels are categorized into three groups, visible in all images, visible in some images, and occluded in all images. After performing decision-level fusion on the OBIAs based on the proposed context aware strategy, analysis shows that decision-level fusion removes majority of the shadow regions from object recognition results and detects road regions occluded by the high-rise buildings or trees especially in areas that are visible in some images.

Fig. 4 depicts the visual comparison between the object recognition results on each of the individual images and decision-level fusion of them in both datasets.

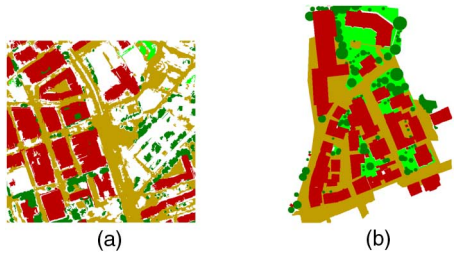

Fig. 6. References for quantitative evaluation of the results on (a) WorldView-2 dataset and (b) DMC dataset.

TABLE VI
ACCURACY ASSESSMENT OF THE DECISION-LEVEL FUSION OBJECT RECOGNITION RESULTS

Dataset		Building	Road	Tree	Grass
WV-2	True positive	360 873	410 833	84 951	3573
	False positive	250 080	244 580	123 810	790
	False negative	39 017	29 987	46 119	6972
	Completeness	0.90	0.93	0.65	0.34
	Correctness	0.59	0.63	0.41	0.82
	Quality	0.56	0.60	0.33	0.31
	Overall accuracy (%)		87		
	Kappa		0.75		
DMC	True positive	89 992	83 822	21 795	17 956
	False positive	72 980	132 810	34 087	10 797
	False negative	10 283	9329	9045	9923
	Completeness	0.90	0.90	0.71	0.64
	Correctness	0.55	0.39	0.39	0.62
	Quality	0.52	0.37	0.34	0.46
	Overall accuracy (%)		85		
	Kappa		0.78		

TABLE VII
ACCURACY ASSESSMENT OF THE OBJECT RECOGNITION RESULTS OF THE INDIVIDUAL VIEWS

	WorldView-2 dataset				DMC dataset					
	View1	View2	View3	View4	View1	View2	View3	View4	View5	View6
Off-nadir angle	30.2°	45.4°	34.0°	45.3°	50.5°	50.4°	11.1°	14.6°	42.0°	43.4°
Overall accuracy(%)	72.45	55.73	71.87	62.61	56.63	56.50	61.93	54.85	61.47	67.26
Kappa	0.56	0.32	0.51	0.42	0.43	0.42	0.48	0.41	0.46	0.55

For providing more visual details, Fig. 5 illustrates the results of multiviews object-based image analysis and decision-level fusion of them on two selected patches. Some of the classification improvements are highlighted by yellow boxes in the results of performing decision-level fusion on both datasets. For example, in the WorldView-2 dataset, the upper left and upper right yellow boxes highlight some recognized road regions that are occluded or shadowed by high-rise buildings in the individual object recognition results. Moreover, in the DMC dataset, yellow boxes highlight improvements in recognizing grass and road regions after solving shadow and occlusion. As it is depicted in Fig. 5, the object-based image analysis of the most nadir WorldView-2 image is provided in Fig. 5(a11) for

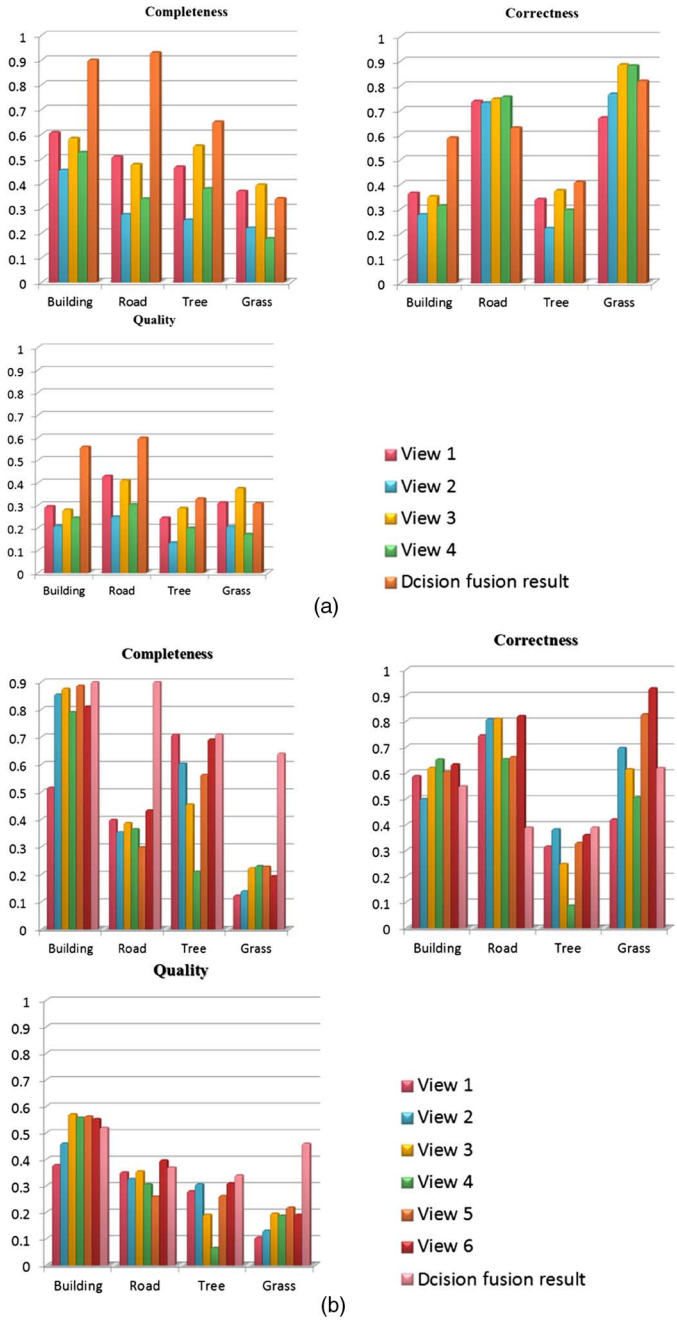

Fig. 7. Comparing the completeness, correctness, and quality of the object recognition results for individual views and decision fusion of them in (a) WorldView-2 dataset and (b) DMC dataset.

visual comparison with the final decision-level fusion of the four oblique images.

For the quantitative evaluation of the results, the following references are used for each dataset. On the WorldView-2 dataset, some segmented regions of the predefined object classes are manually selected by an expert operator based on the digital surface model generated from multiangular WorldView-2 images and the most nadir view [Fig. 6(a)]. On the DMC dataset, the reference classification map is generated by photogrammetric plotting for multiviews DMC images [Fig. 6(b)] [31].

A confusion matrix is produced for comparing the reference areas with their corresponding results from different steps of the object recognition methodology. As depicted in Table VI, the comparison is based on the numbers of correctly detected pixels (true positive), wrongly detected pixels (false positive), and the not correctly recognized pixels (false negative), determined after performing the object recognition algorithm. Moreover, using the confusion matrix and the quantitative values for each object class, completeness, correctness, and quality criteria together with the overall accuracy and Kappa value are determined from the results [32]

$$\text{Completeness} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalseNegative}} \tag{11}$$

$$\text{Correctness} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalsePositive}} \tag{12}$$

$$\text{Quality} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalsePositive} + \text{FalseNegative}}. \tag{13}$$

In order to perform more quantitative analysis on the results, the object-based image analysis of each of the individual views is projected to the ground space for comparing with the reference regions. Then, the overall accuracies and kappa of all individual object-based image analysis are compared with their decision-level fusion results. Comparing Tables VI and VII with each other shows that using the proposed context aware algorithm for decision-level fusion of multiviews object recognition results increases the amount of overall accuracy and kappa values in the classification results.

Fig. 7 illustrates the improvements of the completeness, correctness, and quality of the classification for each of the individual images and decision-level fusion of them.

IV. Discussion and Conclusion

A context aware strategy is proposed for decision-level fusion of the object-based image analysis on multiviews VHR imagery and digital surface model. According to the various off-nadir angles of the sensors, high-rise 3-D objects such as buildings may cause occlusion and shadow areas in the remotely sensed imagery. In such a situation, true class labels of some parts of the object regions cannot be detected. Therefore, large numbers of false positive and false negative pixels decrease the classification accuracies. Using the proposed decision-level fusion of the multiviews based on the developed context aware strategy enhances the completeness and accuracy of the object recognition results. Comparing Tables VI and VII illustrate that the minimum values of improvements in the overall accuracy and kappa using the decision-level fusion object recognition algorithm are about 15 and 0.19 for the WorldView-2 dataset and 18 and 0.24 for the DMC dataset, respectively. Comparing the results from both datasets reveals that increasing the number of individual views may improve the accuracies of their decision-level fusion. In this evaluation, we used the number of 4 WorldView-2 satellite images and the number of 6 DMC aerial images. Moreover,

the accuracy of the decision-level fusion results also depends on the precision of the utilized digital surface model. In the WorldView-2 dataset, we used DSM with 0.5-m spatial resolution which is generated based on the matching and SGM algorithm. On the other hand, in the DMC dataset, we used the more precise Lidar DSM with 25 cm spatial resolution. Therefore, more improvements in the classification accuracies from the decision-level fusion of DMC aerial images depend on using more precise DSM together with more number of individual images. According to Fig. 7, despite improving the completeness and quality of the classification in most object types in both datasets, correctness values of the decision-level fusion results have not improved in all object types for DMC dataset. This situation relates to the dependencies between the classification accuracies of each of the individual images and the decision-level fusion of them. As it is depicted in Table VII, strong spectral capabilities of WorldView-2 satellite images led to the better accuracies of object recognition results in each of the individual images. However, the DMC aerial images with only three spectral bands cannot obtain the accuracies as well as WorldView-2 images. Moreover, the most nadir views in the DMC dataset do not have the best classification accuracies. Therefore, considering large weights of scene context for the most nadir views with weak classification accuracies led to decrease the correctness of the decision-level fusion results in the DMC dataset.

This method still needs further modifications in the field of defining the contextual information such as neighborhood definition for each of the regions and using artificial intelligence techniques such as multiagent system for decreasing the classification errors related to the lack of segmentation capabilities. Moreover, the potential of the proposed method can be further evaluated for remotely sensed data with higher spectral capabilities such as high resolution hyperspectral data with more object classes. For performing the proposed decision-level fusion method on the object-based image analysis of other remotely sensed datasets, all aspects of the method is transferable but only the spectral features may need to modify according to the spectral capabilities of new datasets.

ACKNOWLEDGMENT

The authors would like to thank IEEE GRSS Data Fusion Technical Committee and the German Society for Photogrammetry, Remote Sensing and Geoinformation (DGPF) for providing the WorldView-2 and Vaihingen datasets, respectively, used in this paper [27], http://www.ifp.uni-stuttgart.de/dgpf/DKEP-Allg.html.

REFERENCES

[1] A. K. Shackelford and C. H. Davis, "A hierarchical fuzzy classification approach for high-resolution multispectral data over urban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1920–1932, Sep. 2003.

[2] T. Blaschke, "Object based image analysis for remote sensing," *ISPRS Int. J. Photogramm. Remote Sens.*, vol. 65, no. 1, pp. 2–16, 2010.

[3] W. Zhou and A. Troy, "An object-oriented approach for analyzing and characterizing urban landscape at the parcel level," *Int. J. Remote Sens.*, vol. 29, no. 11, pp. 3119–3135, Jun. 2008.

[4] D. C. Duro, S. E. Franklin, and M. G. Dubé, "A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery," *Remote Sens. Environ.*, vol. 118, no. 2012, pp. 259–272, 2012.

[5] S. W. Myint, P. Gober, A. Brazel, S. Grossman-Clarke, and Q. Weng, "Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery," *Remote Sens. Environ.*, vol. 115, no. 5, pp. 1145–1161, 2011.

[6] X. Huang and L. Zhang, "A multilevel decision fusion approach for urban mapping using very high resolution multi/hyperspectral imagery," *Int. J. Remote Sens.*, vol. 33, no. 11, pp. 3354–3372, 2012.

[7] A. Peets and Y. Etzion, "Automated recognition of urban objects and their morphological attributes using GIS," in *Proc. ISPRS Arch. Vol. XXXVIII, Part. 4-8-2-W9 Core Spat. Databases–Updating Maintenance Serv. Theory Pract.*, Haifa, Israel, 2010, pp. 58–63.

[8] A. Jacquin, L. Misakova, and M. Gay, "A hybrid object-based classification approach for mapping urban sprawl in periurban environment," *Landscape Urban Plann.*, vol. 84, no. 2008, pp. 152–165, 2008.

[9] A. S. Laliberte, D. M. Browning, and A. Rango, "A comparison of three feature selection methods for object-based classification of sub-decimeter resolution UltraCam-L imagery," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 15, no. 2012, pp. 70–78, 2012.

[10] E. Ivits, B. Koch, T. Blaschke, M. Jochum, and P. Adler, "Landscape structure assessment with image grey-values and object-based classification at three spatial resolutions," *Int. J. Remote Sens.*, vol. 26, no. 14, pp. 2975–2993, 2005.

[11] R. V. Platt and L. Rapoza, "An evaluation of an object-oriented paradigm for land use/land cover classification," *Prof. Geogr.*, vol. 60, no. 1, pp. 87–100, 2008.

[12] A. F. Habib, E. Kim, and C. Kim, "New methodologies for true orthophoto generation," *Photogramm. Eng. Remote Sens.*, vol. 73, no. 1, pp. 25–36, 2007.

[13] K. I. Bang, A. F. Habib, C. Kim, and S. Shin, "Comprehensive analysis of alternative methodologies for true orthophoto generation from high resolution satellite and aerial imagery," in *Proc. Annu. Conf. Amer. Soc. Photogramm. Remote Sens.*, Tampa, FL, USA, 2007.

[14] C. Pohl and J. L. Van Jenderen, "Review article multisensor image fusion in remote sensing: Concepts, methods and applications," *Int. J. Remote Sens.*, vol. 19, no. 5, pp. 823–854, 1998.

[15] J. Esteban, A. Starr, R. Willetts, P. Hannah, and P. Bryanston-Cross, "A review of data fusion models and architectures: Towards engineering guidelines," *J. Neural Comput. Appl.*, vol. 14, no. 4, pp. 273–281, 2004.

[16] T. Stathaki, *Image Fusion, Algorithms and Applications*. New York, NY, USA: Academic Press, 2008.

[17] A. H. S. Solberg, "Signal and image processing for remote sensing," in *Data Fusion for Remote Sensing Applications*, 2nd ed., C. H. Chen, Ed. Boca Raton, FL, USA: CRC Press, 2012, ch. 23, pp. 463–484.

[18] Z. Ye, S. Prasad, W. Li, M. He, and J. E. Fowler, "Classification based on 3D DWT and decision fusion for hyper spectral image analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 173–177, Jan. 2014.

[19] Y. H. Ran, X. Li, L. Lu, and Z. Y. Li, "Large-scale land cover mapping with the integration of multisource information based on the Dempster–Shafer theory," *Int. J. Geogr. Inf. Sci.*, vol. 26, no. 1, pp. 169–191, 2012.

[20] B. Huang, H. Zhang, and L. Yu, "Improving landsat ETM+ urban area mapping via spatial and angular fusion with MISR multi-angle observations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 1, pp. 101–109, Feb. 2012.

[21] L. Wolf and S. Bileschi, "A critical view of context," *Int. J. Comput. Vis.*, vol. 69, no. 2, pp. 251–261, 2006.

[22] C. Galleguillos and S. Belongie, "Context based object categorization: A critical survey," *Comput. Vis. Image Understand.*, vol. 114, no. 6, pp. 712–722, 2010.

[23] J. Guo, H. Zhou, and Ch. Zhu, "Cascaded classification of high resolution remote sensing images using multiple contexts," *Inf. Sci.*, vol. 221, no. 2013, pp. 84–97, 2013.

[24] J. J. Lim, P. Arbeláez, C. Gu, and J. Malik, "Context by region ancestry," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, 2009, pp. 1978–1985.

[25] C. Galleguillos, B. McFee, S. Belongie, and G. Lanckriet, "From region similarity to category discovery," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR'11)*, 2011, pp. 2665–2672.

[26] T. Hermosilla, L. A. Ruiz, J. A. Recio, and M. Cambra-Lez, "Assessing contextual descriptive features for plot-based classification of urban areas," *Landsc. Urban Plann.*, vol. 106, no. 2012, pp. 124–137, 2012.

[27] M. Baatz and A. Schape, "Multi-resolution segmentation: An optimization approach for high quality multi-scale image segmentation," in *Proc. Angew. Geogr. Inf. XII Beitragezum AGIT-Symp.*, Salzburg, Austria, 2000, pp. 12–23.
[28] Y. Sheng, P. Gong, and G. S. Biging, "True orthoimage production for forested areas from large-scale aerial photographs," *Photogramm. Eng. Remote Sens.*, vol. 69, no. 3, pp. 25–36, 2003.
[29] M. Neilsen, "True orthophoto generation," MM thesis, Informatics and Mathematical Modeling, Technical Univ. Denmark, 2004.
[30] H. Hirschmüller, "Stereo processing by semi global matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, Feb. 2008.
[31] M. Cramer, "The DGPF test on digital aerial camera evaluation—Overview and test design," *Photogramm. Fernerkundung Geoinf.*, vol. 2, no. 2010, pp. 73–82, 2010.
[32] F. Tabib Mahmoudi, F. Samadzadegan, and P. Reinartz, "Object oriented image analysis based on multi-agent recognition system," *Comput. Geosci.*, vol. 54, pp. 219–230, Apr. 2013.

Fatemeh Tabib Mahmoudi received the Ph.D. degree in photogrammetry from the Department of Geomatics Engineering, Faculty of Engineering, University of Tehran, Tehran, Iran.

Her research interests include most aspects of computer vision specifically object recognition, classification, and mapping in urban areas.

Farhad Samadzadegan received the Ph.D. degree in photogrammetry from the Department of Geomatics Engineering, Faculty of Engineering, University of Tehran, Tehran, Iran.

He has been the Head of Geomatics Department, University of Tehran for four years.

Peter Reinartz (M'09) received the Ph.D. degree in civil engineering from the University of Hannover, Hanover, Germany.

He is the Head of the Photogrammetry and Image Analysis Department, Remote Sensing Technology Institute (IMF), German Aerospace Centre (DLR), Weßling, Germany.