CrossMark

ORIGINAL PAPER

Automatic building extraction from very high-resolution image and LiDAR data with SVM algorithm

Fevzi Karsli[1] · Mustafa Dihkan[1] · Hayrettin Acar[1] · Ali Ozturk[2]

Abstract Recent developments in the field of remote sensing have introduced new sensor technologies in usage of LiDAR, SAR, and high-resolution optical data. Classification performance is expected to increase through combining these various data sources. The purpose of this study is to develop a new approach for automatic extraction of buildings in urbanized and sub-urbanized areas. For this purpose, multi-feature extraction process including the spatial, spectral, and textural features were conducted on the very high spatial resolution multispectral aerial images and the LiDAR data set. SVM algorithm was trained by using this multi-feature data, and the classification was performed. After the classification of building and non-building, objects were extracted with high accuracy for the test areas. As a result, it has been proven that multi-features derived from combination of optical and LiDAR data can be successfully applied to solve the problem of automatic detection of buildings by using the proposed approach.

Keywords Automatic building detection · LiDAR · SVM · Classification · Multi-feature extraction

✉ Fevzi Karsli
  fkarsli@ktu.edu.tr

[1] Department of Geomatics, Faculty of Engineering, Karadeniz Technical University, 61080 Trabzon, Turkey

[2] Trabzon Vocational High School, Karadeniz Technical University, 61300 Trabzon, Turkey

Introduction

Extraction of geographic objects such as buildings and roads from remotely sensed data is very important as it can be used for city planning, cartographic mapping, land valuation, disaster management, and many other spatial information-dependent applications (Chen et al. 2009). Fast and reliable extraction of spatial information on the buildings and roads is a vital step in production phase process of 3D city models, because this type of data is needed for building, updating, and maintaining of a related geographic information system (GIS) database. Remote sensing technologies have been intensively used to produce this type of data in large urbanized regions and metropolises. Nowadays, in order to increase the reliability, decrease processing time, and minimize the human factor, many studies have been focused on automatization of processing remotely sensed data to obtain the required geographical features with high accuracy. There are many reference studies on how to extract the buildings and similar objects from remote sensing dataset, available in literature (Mayer 1999; Cheng et al. 2013; Gilani et al. 2016; Zarea and Mohammadzadeh 2016; Zhao et al. 2016) (http://www.dgpf.de/neu/WWW-Projekt-Seite/DKEP-Allg.html).

Building extraction is a demanding computer vision process due to occlusion, poor contrast, shadow effect, and inconvenient image perspective problems (Sohn and Dowman 2007; Chen et al. 2009; Awrangjeb et al. 2010). There are mainly three different types of solutions including sensor type and source suggested by remote sensing community to overcome this problem (Lee et al. 2008). The first one is based on algorithm which solely uses optical sensors' data. Today's technical achievements in optical sensor technologies have made very high spatial and spectral resolution data widely available. Therefore, a great amount of information can be collected on building extraction process from supplied digital

Springer

image data. Data with dense information content from optical sensors can sometimes create confusion due to similar spectral characteristics of the objects which may decrease the separability of the buildings from other manmade objects. The success of automatic detection algorithms for buildings has been limited by spectral mixing problem when only optical digital images are used (Sohn and Dowman 2007). The second approach that can be used for extraction of manmade objects is based on Light Detection and Ranging (LiDAR) technology. Within this frame, point cloud produced by LiDAR has been used in the specially designed algorithms instead of optical sensor images to solve automatic building extraction problems (Weidner 1996; Wang and Schenk 2000; Vosselman 1999; Maas and Vosselman 1999). The purpose of such algorithms is to isolate the solution of building extraction problem from optical sensors defects. Meanwhile, LiDAR data can offer the elevation information for available objects with high accuracy and reliability with some limitations as LiDAR data has relatively low spatial accuracy in horizontal plane which makes extremely difficult to extract building corners (Sampath and Shan 2007; Demir et al. 2009). Furthermore, there are some reciprocal advantages and disadvantages for both mentioned solutions above in classification of buildings. Therefore, in order to overcome the classification problem, a final hybrid solution that incorporates LiDAR and optical sensor data has been proposed by remote sensing community. This third approach seeks for solution by fusing optical sensor data with LiDAR data to exploit reliable and accurate elevation information from LiDAR as well as texture and buildings edges from optical digital images. However, how to integrate LiDAR and optical data has become a tough popular research topic in feature extraction and image fusion literature (Rottensteiner et al. 2005; Zhang 2010).

Multi-dimensional datasets are constructed by combining multi-sensor data. Classification algorithm performance has a great importance in classifying multi-dimensional datasets. Non-parametric algorithms generally perform better than parametric classification algorithms due to some inherent statistical assumptions tied to parametric methods in classification of such datasets (Pal and Mather 2005). The well-known problem in parametric methods is the assumption that spectral signature of classes is in normal distribution, which may negatively affect the overall performance of the method in case the assumption is null (Kavzoglu and Reis 2008). However, despite the fact that non-parametric methods require so many initial parameters selection, they can yield very high classification accuracies when appropriate parameters are selected. SVM has positively affected the classification accuracy as it has high statistical learning capacity based on structural risk minimization (SRM) (Vapnik 1995) and high generalization potential on multi-dimensional spaces (Pal and Mather 2005; Kavzoglu and Colkesen 2009).

The proposed approach enables automatic building extraction without human interference. For this purpose, LiDAR and multispectral digital aerial images were combined to extract the textural, spatial, and spectral features. In order to eliminate edge location error caused by discrete sampling distance of LiDAR sensor, 26 dimensional feature set for the buildings and other objects were created. Classification process was completed by applying SVM on multi-feature data set produced with automatically selected training data using image processing techniques. Thus, successful and reliable results were obtained by applying the proposed technique. Moreover, due to the automatization capability of the algorithm, it can be safely used in different regions without losing its performance.

Study area and data

In order to test the proposed approach, two different regions characterized as urban and semi-urban with dimensions of 326 × 351 and 350 × 347 m were selected as test area (Fig. 1). The elevation of study area is changing between 207 and 333 m. The areas are intentionally selected because of special topographic character with moderate slope and varying morphology. In the study area, it is almost impossible to automatically extract buildings when only optical sensor data is used because of the fact that available buildings are in different geometrical structure, at various spectral characteristics of their roof blocks, and under neighboring effects of some impervious surfaces such as asphalt and concrete. Another important characteristic of the study area is the adjacency position of high trees and buildings in semi-urban regions. There is no dominating single geometric shape for the roof types; moreover, many roof types such as flat, gabled and shed can be observed in the area. Therefore, it is also impossible to automatically detect the buildings based on only LiDAR data. Consequently, due to mentioned problems above, an alternative approach can be applied by combining LiDAR and optical sensor data to solve automatic building extraction problem.

The LiDAR data were collected and recorded as a *.las file format. It was captured by ALS50 Leica Geosystems mounted on an aircraft with a point density of eight points per square meter and geo-referenced in UTM system. This data were acquired using ALS50 at an average flying height of 500 m. The average LiDAR point spacing is roughly 30–35 cm depending on the number of overlapping strips (Cramer 2010). A
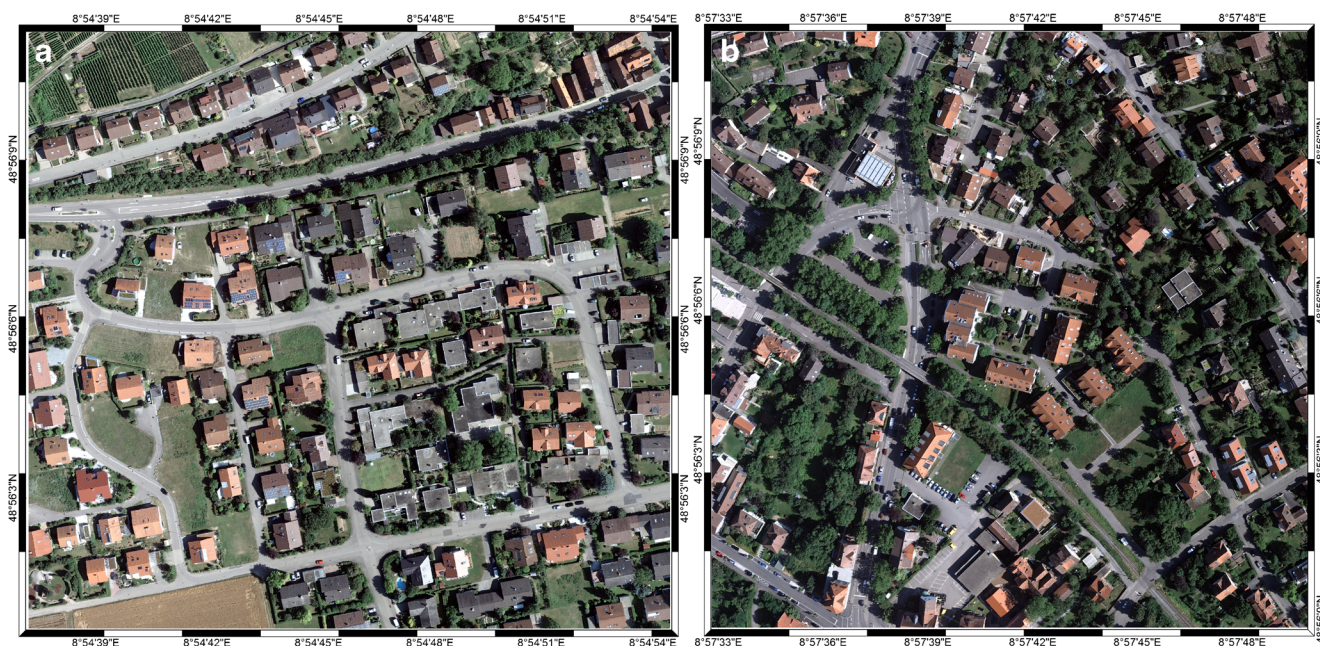
Fig. 1 Study area. **a** Test area I. **b** Test area II

simultaneous acquisition of very high-resolution multi-spectral image with LiDAR data was performed with DMC digital camera produced by Intergraph Z/I Imaging Company. An image data was acquired on August 2012 at 7.54 a.m. local time. This data with *.tiff file format has 8 cm GSD and 16-bit radiometric resolution.

Methodology

As a new hybrid approach in this contribution, very high-resolution multispectral digital aerial image was used in combination with LiDAR point cloud data for automatic extraction of buildings. In this new approach, multisource data sets were used to detect textural and spectral features of the target objects. Then, a self-training semi-supervised classification approach was performed on prepared multi-feature dataset. Processing steps of the proposed algorithm is shown as a flow chart in Fig. 2.

LiDAR data processing

LiDAR point cloud was used for generating DTM (digital terrain model) and DSM (digital surface model) to obtain elevation values for buildings and some other objects. Accordingly, raw LiDAR data was classified and filtered to represent bare-earth surface as one class and a surface with elevated objects as a second class. A progressive TIN (triangulated irregular network) densification algorithm by Axelsson (2000) was used for filtering of the raw LiDAR data in order to get DTM

information. After separating ground point cloud, the data was converted to a raster format by defining row, column, and pixel dimensions. At this stage, pixel size was determined as 10 cm by considering LiDAR data point density for DTM (Fig. 3b, e) generation (Isenburg et al. 2006a, b). Similarly, TIN model was created from all point cloud by considering first returns and eliminating aliasing effect of the surface objects, then a 10 cm pixel size raster DSM (Fig. 3a, d) was produced (Isenburg et al. 2006a, b). In both processes, elevation information was assigned as gray value. As a next step, a normalized DSM (nDSM) was produced by subtracting DTM from DSM (Fig. 3c, f). Finally, absolute elevations for buildings and non-ground objects were determined in a certain elevation plane and used in all forthcoming image processing and classification stages. Meanwhile, a slope image (Fig. 4a, c) was created from nDSM in order to be used in classification process.

Pre-processing and geo-referencing optical image to LiDAR data

Pre-processing and referencing stage is a necessity to pool all the available data in a single reference system and datum. In this study, data in raster format from different sources was resampled to the same pixel size to carry out necessary spatial analysis on aerial image, DTM, DSM, and nDSM. This processing step is important in order to perform pairwise comparison of all data sets. However, due to irregular distribution of LiDAR point cloud, it is highly difficult to select control points
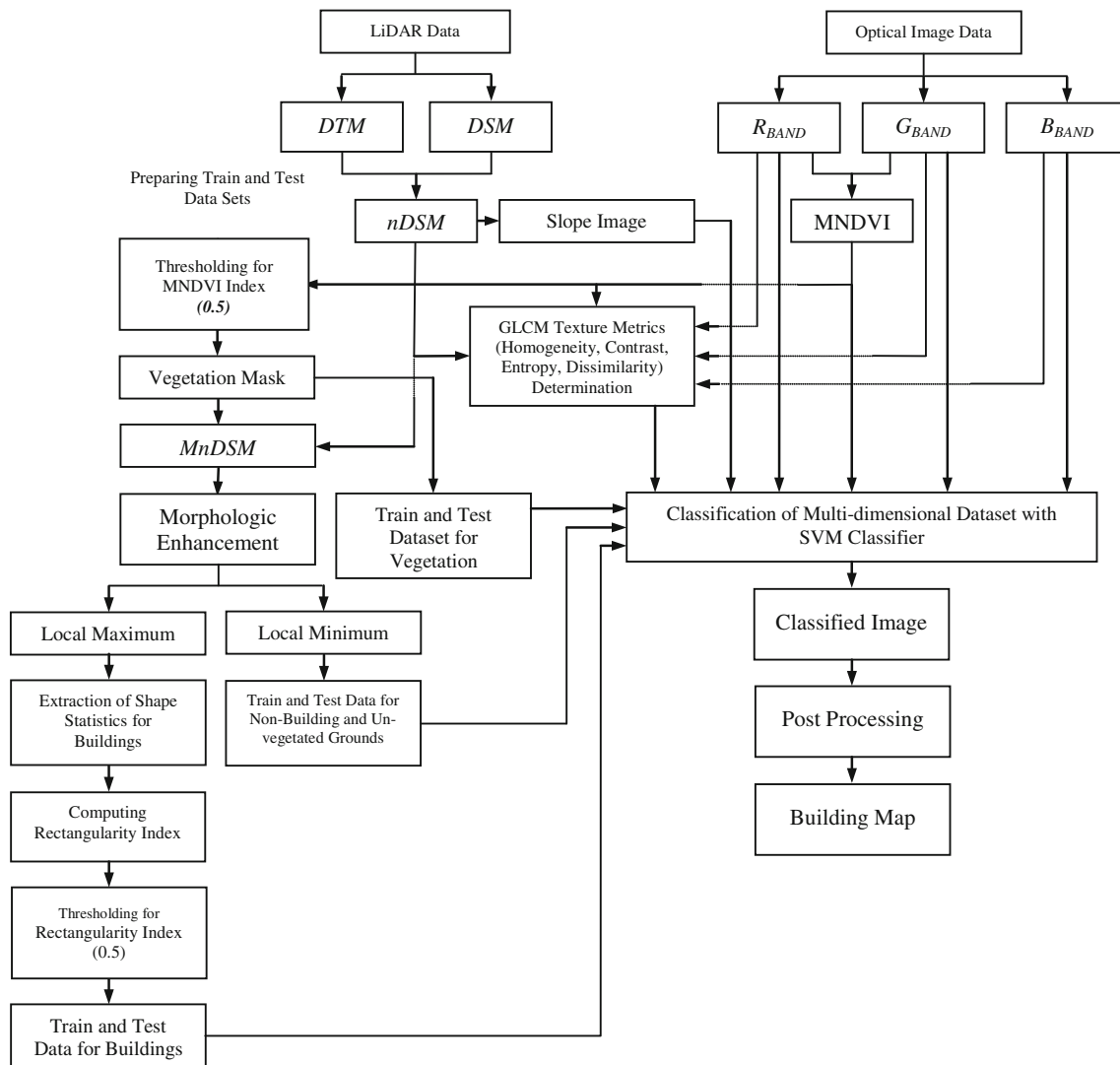
Fig. 2 Flowchart of proposed methodology

on test area. Therefore, an intensity image was produced by taking account all returns to select required control points, and then multispectral image of the same area was registered with intensity image, and projective transformation was implemented by using a sufficient number of control points. There was no need for orthorectification process as the image data was already delivered in orthorectified form. In order to combine and match image and LiDAR data or raster based elevation models, they must have the same geometric resolution. After referencing stage, bicubic re-sampling method was applied to resample the image data from 8 to 10 cm pixel size.

LiDAR and image datasets each have specific advantages and disadvantages in terms of horizontal and vertical positioning accuracy. Optic or image data produced by photogrammetric technique can provide broad 2D information such as high-resolution texture and color

information as well as with accurate horizontal accuracy. In contrast with photogrammetric technology, LiDAR can quickly acquire dense and accurate height data of the areas by emitting and receiving laser pulses. The height changes are more appropriate for detecting building boundaries more than the spectral and texture changes. However, the horizontal accuracy of LiDAR data is poor because of laser pulse discontinuousness. Compared with photogrammetric imagery, LiDAR provides more accurate height information but less accurate boundary lines. Considering the complementary advantages of LIDAR data and high-resolution optic image data, the fusion of two data sources is regarded as a promising procedure to detect the building boundaries (Awrangjeb et al. 2010; Li et al. 2013). Therefore, horizontal spatial information from high-resolution optic image and vertical spatial information from LIDAR data were included to 26 dimensional multi-feature dataset in
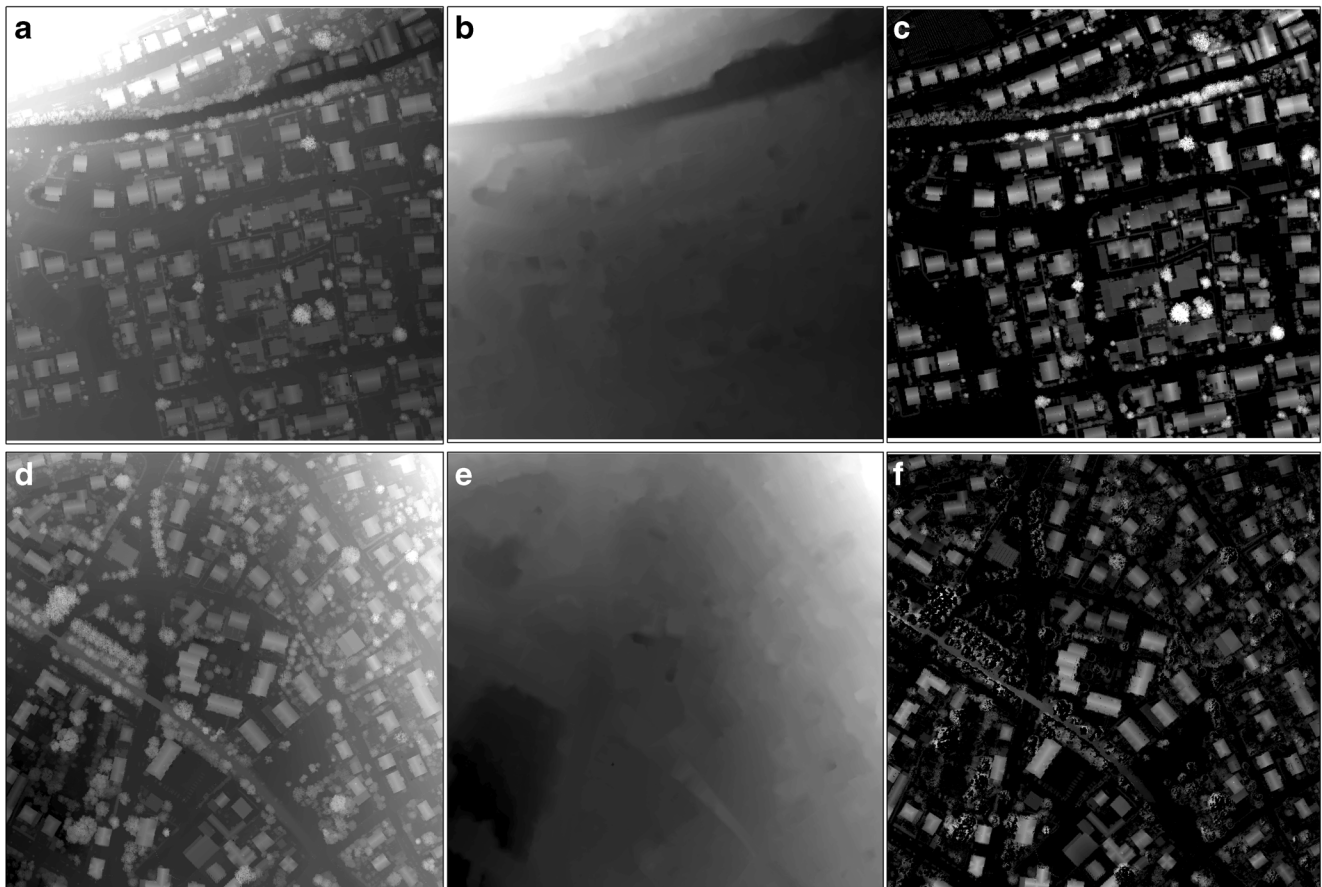
Fig. 3 **a, d** DSM produced from raw LiDAR dataset. **b, e** DTM created by Progressive TIN Densification algorithm. **c, f** nDSM image. **a–c** For Test Area I. **d–f** For Test Area II

this study. Due to stability of the horizontal accuracy of the image data, all feature dataset were compounded with 10 cm raster format resolution.

Feature extraction and automatically prepared training data set

At this step, feature extraction processes were applied on different data sources. Training data set was automatically produced for self-training of the SVM algorithm to obtain different classes (vegetation, buildings, and other objects) that would be used in machine learning and classification stage.

Extraction of MNDVI features

Normalized difference vegetation index (NDVI) has been proved very successful in separating green cover with chlorophyll content from other land use/land cover classes when multispectral image data is used (Rouse et al. 1973). The digital camera used in this study has an ability to acquire data in very wide intervals of visible part of the electromagnetic spectrum. The operation range of the instrument can reach up to

675 nm in visible range. Therefore, there is a need to alter narrow satellite band based NDVI index, thus a modified type of NDVI so called modified normalized difference vegetation index (MNDVI) is proposed in this study. It is expected that detection of vegetation class would be improved by extending the spectral interval and using wide band intervals of the available sensor. The calculation for MNDVI is given in Eq. 1.

$$\mathrm{MNDVI} = \frac{\mathrm{Red}_{(590-675\mathrm{nm})} - \mathrm{Green}_{(500-650nm)}}{\mathrm{Red}_{(590-675\mathrm{nm})} + \mathrm{Green}_{(500-650\mathrm{nm})}} \quad (1)$$

In this study, training data set was automatically produced using MNDVI algorithm. The reason for using MNDVI was not direct extraction of the vegetation class by masking the multi-dimensional data set but to decide on training pixel locations representing vegetation with high probability by means of automatically defined threshold value. Moreover, MNDVI features contributed to the classification process by enhancing the separability of the vegetated areas.

MNDVI image produced by applying Eq. 1 was processed by applying a threshold value to obtain the vegetation mask (Fig. 4b, d). The threshold value was decided by taking into account descriptive statistics of MNDVI image. It is basically

Fig. 4 **a**, **c** Slope image and **b**, **d** MNDVI image for Test Area I and II, respectively
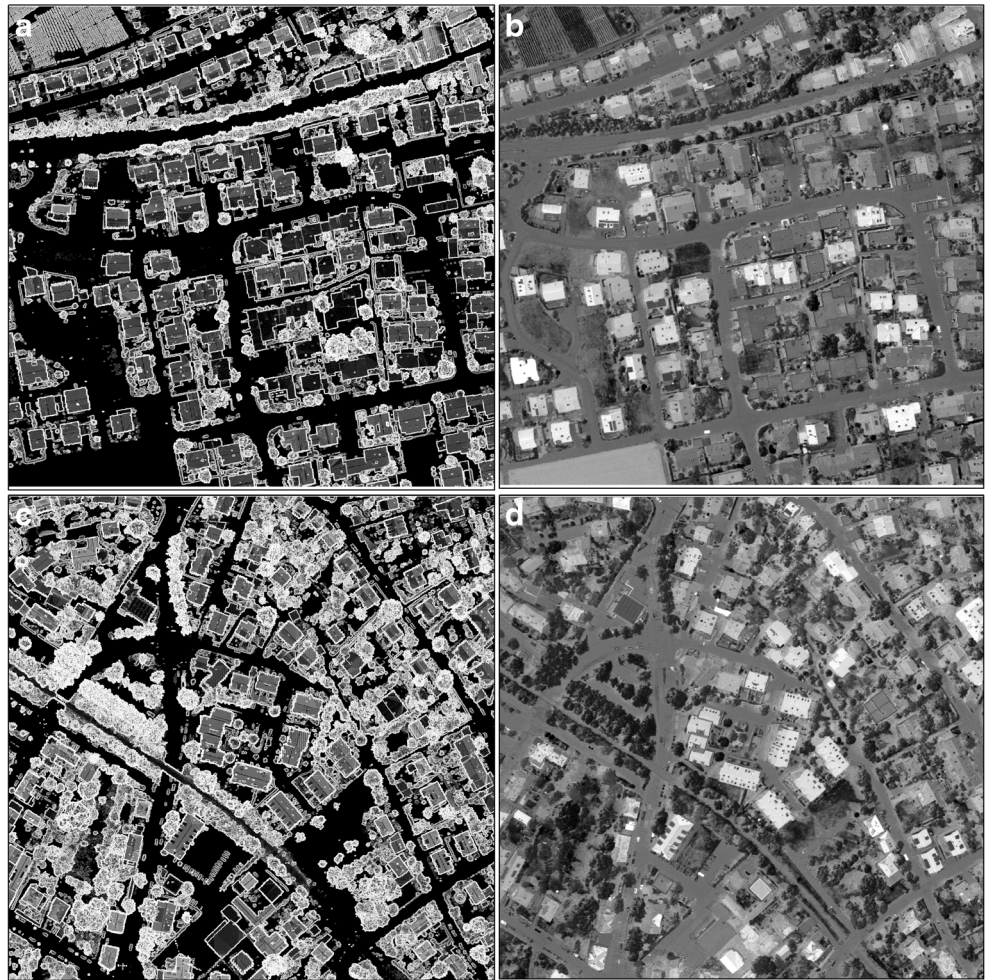
derived by subtraction of standard deviation (std) from mean (*m*) value of the MNDVI image pixel values. Pixel (600 pixels) locations used for training were simultaneously determined from the same vegetation mask image with stratified random sampling method. Location information of these pixels was stored to be used during the classification process as train data. Furthermore, test data set with appropriate spatial distribution was manually selected and recorded.

Producing training data set for buildings and other objects

In order to perform a successful extraction of building surfaces, in addition to vegetation class, two more classes (buildings and other objects) other than vegetation were decided. Buildings and other objects' training data set which is required for self-training of the SVM algorithm were automatically produced in this step. Similar to procedure applied for training of the vegetation class, some pre-processing operations were carried out for defining pixel locations of buildings and other objects with high probability. Pixel locations were only used during the process of creating training data set.

For this purpose, buildings and some ground regions were extracted from nDSM image using with image processing techniques. In order to decide whether those regions are buildings or not, a multi-step decision making strategy was followed by evaluating shape-related parameters. Previously prepared vegetation mask was used to minimize possible errors in determination building pixels on nDSM image to obtain masked nDSM (MnDSM) image. The resultant MnDSM image contains elevation information of buildings and some other built-up objects. Some morphologic operations were carried out on MnDSM image to exclude very tall non-building objects that may affect building pixel selection. A rectangular shape morphologic structuring element (70 × 70 pixels) which corresponds to smallest building dimension on original image was defined. Morphologic enhancement was achieved by applying the structuring element and reconstruction algorithm-based image opening and closing operations, then regional maxima of enhanced MnDSM image was computed (Vincent 1993). Those regions represent the buildings on the image. Moreover,

shape properties of the objects were also incorporated to ensure that available regions are most likely buildings. For this purpose, a rectangularity index (R_i) that is widely used in the literature was applied to estimate rectangularity of an object (Rosin 2003), where (A_i) is an area of a building and (MBR_i) is minimum bounding rectangle (Eq. 2).

$$R_i = \frac{A_i}{MBR_i} \qquad (2)$$

The index value is changing between 0 and 1 where "1" represents an exact rectangularity. In order to increase the reliability of finding building regions, an additional threshold (0.5) was applied using rectangularity parameter. The threshold value does not limit the overall automatization performance and applicability of the algorithm to some other regions due to rectangular geometry of the building objects. After thresholding operation, remaining pixels are most likely buildings. These regions were only used for training data set; final building extraction process was carried out in following classification step. From pixels inside of those regions, stratified random sampling methodology was used to automatically select 600 pixels and their locations for training of the classifier. In order to be used for accuracy assessment procedure, pixels belonging to building class with appropriate spatial distribution were manually selected and recorded as independent test data (400 pixels). As a next step, un-vegetated and non-building regions were remained to be decided. Therefore, previously processed and morphologically enhanced MnDSM images of the study area were used to find regional minima (Vincent 1993). Finally, the extracted regions were non-buildings and un-vegetated areas. Training pixels were selected from these regions similar to previous steps as explained before to be used in classification stage.

*Dataset preparation and textural feature extraction*

Automatic detection of various objects using multi-source data (optical or LiDAR) requires some additional information on spatial, spectral, and textural properties of the related objects (Aksoy et al. 2010). Different studies have reported on how to use shape and size features, spectral and textural features of multispectral aerial images (Vogtle and Steinle 2000), and textural features of elevation images produced from LiDAR data (Maas and Vosselman 1999) in building extraction problem. These features may belong to a pixel or regions created after a pre-processing stage. In this study, the applied methodology accounts for spectral information from optical image data, spatial properties of LiDAR

data as well as textural information of both datasets. Textural features can definitely improve the separability of spectrally similar objects such as buildings and adjacent regions. For rectangular objects like buildings, second order statistics (i.e., co-occurrence features) computed by local rectangular image windows are very effective in characterization of textural features. Therefore, four different textural metrics which are contrast, dissimilarity, entropy, and homogeneity were computed from GLCM (gray level co-occurrence matrix) and used for characterization of textural features (Haralick 1979). A kernel with 5×5 window size was designed for producing GLCM to extract textural features. The kernel was passed over the image to assign the calculated textural metric value to central pixel in this technique. Textural features were produced for RGB bands, nDSM, and MNDVI images; 20 different textural features were totally obtained. Finally, a 26 dimensional multi-feature dataset (3 RGB bands, 1 nDSM elevations, 1 MNDVI index, 1 slope ratio index, and 20 textural features) was prepared to be used in classification stage.

Machine learning and SVM classification

For automatic building extraction problem, three classes which are building, vegetation, and other objects were decided. Prior to classification process as explained in "Extraction of MNDVI features" and "Producing training data set for buildings and other objects" sections labeled data was used for automatic self-training of SVM classifier (600 pixels). Spectral signatures for each available class were produced by using 26 features value from training dataset. Non-parametric SVM has proved to be successful in classification of linearly inseparable multi-dimensional datasets (Vapnik 1995). However, selection of initial parameters in SVM has a vital importance as it is for many similar non-parametric classifiers. Therefore, radial based kernel (RBK) function was preferred in SVM process due to requirement of less initial parameters for solution and better performance for SVM algorithm (Mathur and Foody 2008; Kavzoglu and Colkesen 2009). Radial based kernel parameters were computed from training data by applying grid search method and assigned as $C = 3000$ and $\gamma = 0.07$. The SVM algorithm was automatically applied with these parameters and finalized by assigning every pixel to three different classes. After classification, two classes (building and non-building) were produced by passing vegetation class into other objects (Fig. 5a, c). In accordance with the aim of the study, building surfaces should preserve the unity and homogeneity structure. Hence, small areas appeared on image as holes due to building appliances such as chimneys, solar panels,

Fig. 5 **a**, **c** Classified image and **b**, **d** Enhanced resultant image after post processing for Test Area I and II, respectively
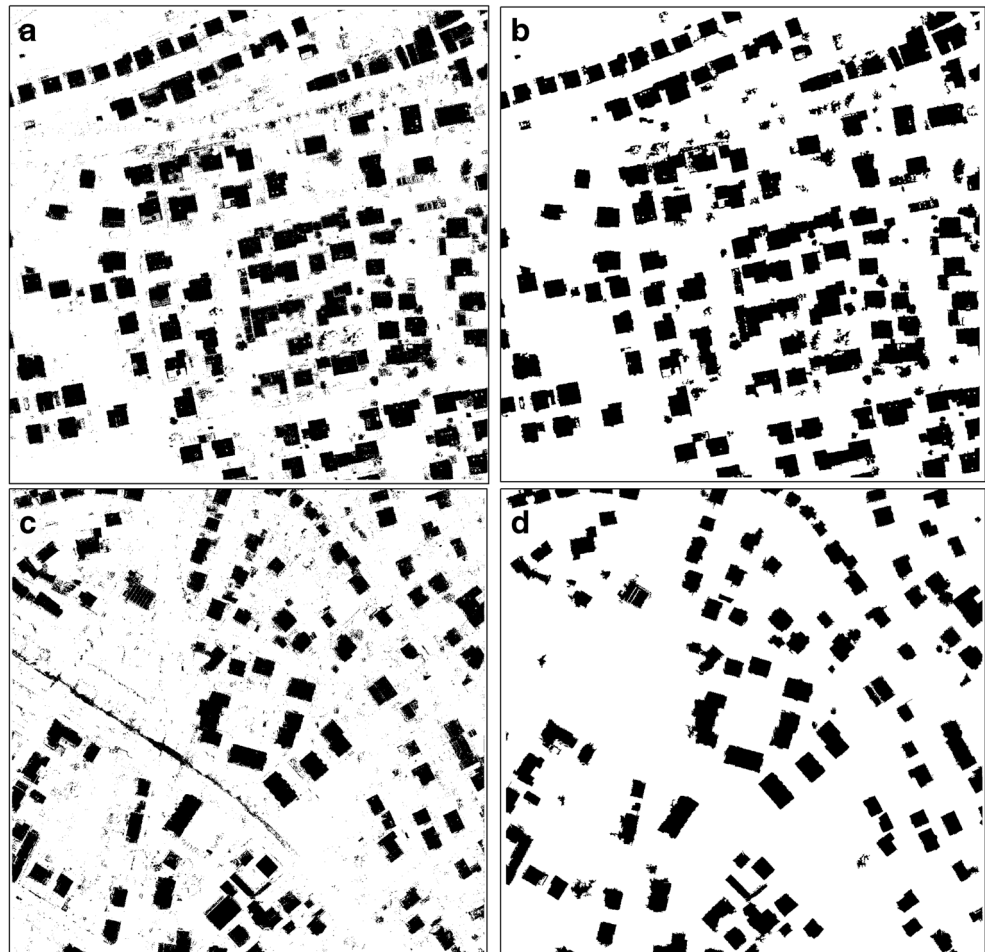
etc. were filled by using some morphologic operations. After filling operations, 8-connected neighborhood was used to eliminate erroneous parts on building edges and a final building map was produced (Fig. 5b, d).

Results

A two-step accuracy assessment methodology was employed to test obtained results of the proposed approach. As a first step, traditional pixel-based site-specific accuracy assessment technique was applied using prepared test data to get the confusion matrix. Overall accuracy (Oa), user accuracy (Ua), producer accuracy (Pa), and kappa index were computed from confusion matrix as shown in Table 1.

Moreover pixel-based accuracy metrics such as correctness ($Corr_{pix}$), completeness ($Comp_{pix}$), and quality (Q_{pix}) were computed. These are widely used metrics for accuracy assessment in two class (object and background) object recognition problem (Rutzinger et al. 2009; Awrangjeb and Fraser 2014).

Table 1 Confusion matrix and pixel-based accuracy statistics

SVM classification results (Test area I)				
Classes (percent)	B	NB	V	Pa
B	97.51	1.15	0.80	97.51
NB	3.00	96.00	1.00	93.80
V	0.25	1.50	98.75	95.99
Ua	98.08	94.75	94.33	95.75
		Oa: 95.75		Kappa: 0.94
SVM classification results (Test area II)				
Classes (percent)	B	NB	V	Pa
B	89.53	0.33	2.00	89.53
NB	8.99	99.50	1.20	99.50
V	1.48	0.17	96.80	96.80
Ua	98.38	88.42	97.38	94.58
		Oa: 94.58		Kappa: 0.92

B building, *V* vegetation, *NB* non-building

Table 2 Metrics results for each test area

	Test Area I (%)	Test Area II (%)
$Comp_{pix}$	87.1	92.0
$Corr_{pix}$	93.8	82.0
Q_{pix}	82.3	76.6

The calculation of the correctness (matching rate), completeness (detection rate), and quality metrics was carried out by applying Eqs. 3, 4, and 5, and the results for each test area were given in Table 2.

$$\text{Corr}_{\text{pix}} = \frac{\|T\text{P}\|}{\|\text{TP}\| + \|\text{FP}\|} \tag{3}$$

$$\text{Comp}_{\text{pix}} = \frac{\|\text{TP}\|}{\|\text{TP}\| + \|\text{FN}\|} \tag{4}$$

$$\text{Quality}_{\text{pix}} = \frac{\|\text{TP}\|}{\|\text{TP}\| + \|\text{FN}\| + \|\text{FP}\|} \tag{5}$$

In Eqs. 3, 4, and 5, an entity classified as an object that also corresponds to an object in the reference is classified as a true positive (TP). A false negative (FN) is an entity corresponding to an object in the reference that is classified as background, and a false positive (FP) is an entity classified as an object that does not correspond to an object in the reference. A true negative (TN) is an entity belonging to the background both in the classification and in the reference data (Rutzinger et al. 2009; Shufelt 1999).

The results in Table 2 showed that building extraction procedure was satisfactory in the proposed methodology.

Considering the first part of the accuracy assessment procedure, it was proved that overall accuracy were around 95.75 and 94.58 % for test area I and II, respectively. It can be considered very high in terms of building extraction process for the two test areas. Despite the recently increasing controversy on its usage due to reliability issues, kappa statistic which is a commonly accepted accuracy metric by remote sensing community was also around 0.92 and 0.94. Classification performance can be qualitatively confirmed from Fig. 5.

The second step of accuracy assessment involves a pattern-based approach (Foody 2008; Dihkan et al. 2013). Therefore, pixel-based accuracy assessment results were confirmed by comparing those two techniques; furthermore, the unity level of building surfaces was accurately delineated. For applying pattern-based accuracy assessment approach in the test areas, reference building map, which represents homogeneously distributed 25 and 31 building objects with different shapes, textures, and spectral characteristics, were created by manual digitizing process (Fig. 6a, c).

Polygon-based fuzzy local matching technique by Power et al. (2001) was used to show the geometric similarity level between building objects of reference and resultant building maps. Selected building polygons were compared by considering areal intersection. In this methodology, after pairwise object comparison, matching level is calculated based on fuzzy inference system by considering various characteristics of the polygons. Preliminary characteristics were area of intersection, area of disagreement, and size of polygon. Within frame of fuzzy inference system, a fuzzification process is first applied to polygon characteristic (Power et al. 2001). Then, a fuzzy rule set was applied to obtain fuzzy local matching membership. Finally, the area centroids of the defuzzification algorithm were used to assign the local matching numbers (L) for each unique polygon (Jager 1995). Two final outputs, comparison (agreement) map and global matching index, can be produced from fuzzy polygon-based approach. Spatial local matching value distribution of the target polygons is available from comparison map which determines the superiority of the pattern-based approach against traditional pixel-based site-specific techniques. Very high similarity is visually confirmed by comparison of reference and building maps (Fig. 6b, d). Moreover, unity level of the target buildings as well as missing or erroneous parts can be both quantitatively and qualitatively observed from comparison map. Local matching value 0.7 and over is generally considered successful in the literature (Power et al. 2001; White 2006). However, it becomes very difficult to decide about a single definite threshold value in terms of accuracy assessment. Meanwhile, global fuzzy matching index (g) that it is between "0 and 1" is also available to show global similarity level of the two maps (Eq. 6).

$$g = \left(\sum_i L_i A_i \right) / \left(\sum_i A_i \right) \tag{6}$$

Where L_i is a local matching value for polygon i and A_i is an area of polygon i. Global matching index (g) values were computed as 0.795 and 0.743 for the test areas I and II in this study. This value shows the success of proposed building extraction process. Differences in g indices are relatively low when compared to traditional accuracy metrics (Oa, Pa, Ua, Kappa, etc.) as the traditional metrics do not include a fuzzy-system-based evaluation capability (Foody 2002). Map Comparison Kit by RIKS (2005) was used for pattern-based accuracy assessment in this study.

Discussion

Different from threshold-based building extraction methods, in this study, building extraction problem is evaluated by combining classical thresholding and automatically trained SVM classifier (Sohn and Dowman 2007; Cheng et al. 2008). The proposed methodology can considerably improve the automatization level of the process. In building extraction problem, different techniques follow various accuracy evaluation
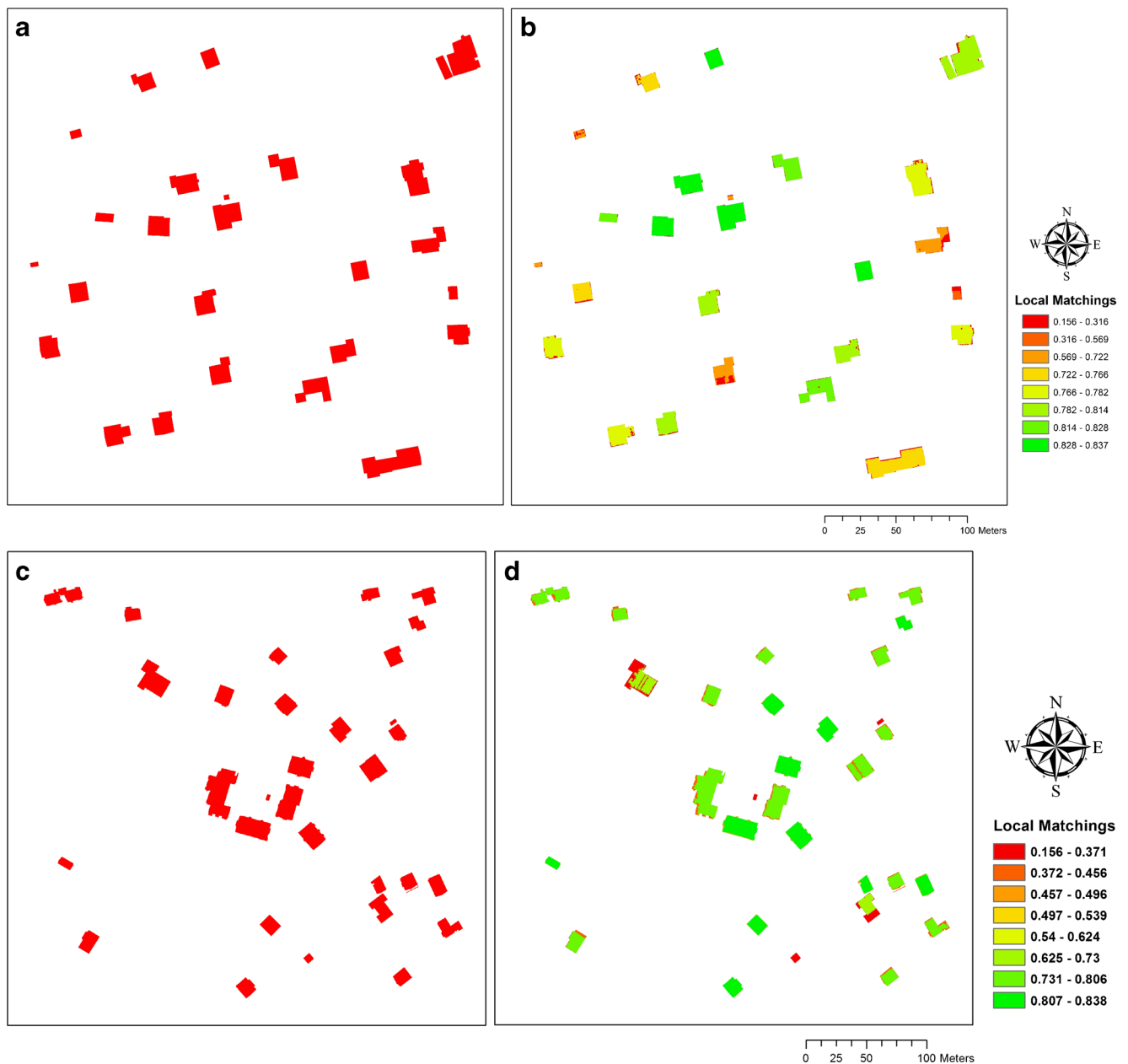
Fig. 6 **a**, **c** A reference map for Pattern Based Accuracy Assessment and **b**, **d** Comparison map between reference and building maps for Test Area I and II, respectively

processes which makes it difficult to compare obtained results in the available literature.

There are many studies available in the literature on building extraction problem. For example, Sohn and Dowman (2007) used LiDAR data for automatic extraction of buildings by using standard building models and obtained an accuracy up to 90.1 % (the correctness) and 80.5 % overall quality by applying pixel-based accuracy assessment. Unfortunately, the accuracy level was negatively affected by regions with lower point density on LiDAR data and available buildings having non-standard shape geometry. Similarly, Rottensteiner et al. (2005), Rottensteiner et al. (2007), and Rutzinger et al.

(2009) used Dempster-Shafer methodology for automatic extraction of buildings by fusing LiDAR and multispectral image data. Vu et al. (2009) applied a multi-scale mathematical morphology for extracting the building features from LiDAR and spectral data. It is argued that proposed methodology is capable of detection of diverse building shapes, and better accuracy can be reached depending on availability of higher density LiDAR data. Pixel-based completeness and correctness accuracies were around 80 and 94 %, respectively. Analyzing Table 2, for the test area I, the completeness value was around 87.1 % and correctness value was 92 %. Similarly, for the test area II, completeness and correctness values were

93.8 and 82.0 %, respectively. Therefore, it can be inferred that the algorithm performed better in test area I than test area II based on reference data.

Furthermore, many researchers have tested various methods for building detection of ALS and aerial image data (Kabolizade et al. 2010; Awrangjeb et al. 2010; Chen et al. 2012; Cheng et al. 2013; Awrangjeb et al. 2013; Li et al. 2013; Rottensteiner et al. 2014). The general accuracy level was around 90 % which is comparable to our results in this study. Thus, it is clear from current study that our proposed methodology revealed satisfactory results in terms of accuracy for building extraction.

Conclusion

Building extraction problem has been an attractive research area in remote sensing and photogrammetric applications. Even though, there have been many achievements in this field, there are still some challenges related to complex building or object morphology. In this study, multisource remotely sensed data were combined to derive textural, spatial, and spectral features of objects of interest.

It is concluded that the usage of multisource data has revealed satisfactory results. The horizontal accuracy of boundaries extracted from LIDAR data is relatively low due to laser pulse characteristics. Low edge detection accuracy in horizontal plane was improved by optical image data whilst spectral mixing of image data was minimized by LiDAR data. Moreover, promising results were obtained by applying SVM classifier in decision space on multi-feature dataset. The robustness of SVM algorithm in classification of multi-feature data set is already well known and widely used in remote sensing applications. The proposed methodology has some artifacts related to vegetation morphology and color characteristics such as very dark or brownish canopies and very tall trees adjacent to building objects. Even though that problem was minimized using MNDVI index, it can potentially be sorted out by adding NDVI index for the proposed methodology as there was no NIR band available in our data set.

References

Aksoy S, Akcay HG, Wassenaar T (2010) Automatic mapping of linear woody vegetation features in agricultural landscapes using very high-resolution imagery. IEEE Trans Geosci Remote Sens 48:511–522

Awrangjeb M, Fraser CS (2014) Automatic segmentation of raw LiDAR data for extraction of building roofs. Remote Sens 6(5):3716–3751

Awrangjeb M, Ravanbakhsh M, Fraser CS (2010) Automatic detection of residential buildings using LIDAR data and multispectral imagery. ISPRS J Photogramm Remote Sens 65:457–467

Awrangjeb M, Zhang C, Fraser CS (2013) Automatic extraction of building roofs using LIDAR data and multispectral imagery. ISPRS J Photogramm Remote Sens 83:1–18

Axelsson P (2000) DEM Generation From Laser Scanner Data Using Adaptive TIN Models. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXIII(Pt. B4/1):110–117

Chen YH, Su W, Li J, Sun ZP (2009) Hierarchical object oriented classification using very high resolution imagery and LiDAR data over urban areas. Adv Space Res 43(7):1101–1110

Chen D, Zhang L, Li J, Liu R (2012) Urban building roof segmentation from airborne LiDAR point clouds. Int J Remote Sens 33(20):6497–6515

Cheng L, Gong J, Chen X, Han P (2008) Building boundary extraction from high resolution imagery and LIDAR data. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXVII (Part B3b):693–698

Cheng L, Li T, Chen Y, Zhang W, Shan J, Liu Y, Li M (2013) Integration of LiDAR data and optical multi-view images for 3D reconstruction of building roofs. Opt Lasers Eng 51:493–502

Cramer M (2010) The DGPF-Test on Digital Airborne Camera Evaluation—Overview and Test Design, PFG, 02/2010

Demir N, Poli D, Baltsavias E (2009) Extraction of buildings using images and LIDAR data and a combination of various methods. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 38:71–76

Dihkan M, Guneroglu N, Karsli F, Guneroglu A (2013) Remote sensing of tea plantations using an SVM classifier and pattern-based accuracy assessment technique. Int J Remote Sens 34(23):8549–8565

Foody GM (2002) Status of land cover classification accuracy assessment. Remote Sens Environ 80:185–201

Foody GM (2008) Harshness in image classification accuracy assessment. Int J Remote Sens 29(11):3137–3158

Gilani SAN, Awrangjeb M, Lu G (2016) An automatic building extraction and regularisation technique using LiDAR point cloud data and Orthoimage. Remote Sens 8:258

Haralick RM (1979) Statistical and structural approaches to texture. Proc IEEE 67:786–804

Isenburg M, Liu Y, Shewchuk J and Snoeyink J (2006a) Streaming Computation of Delaunay Triangulations. Proceedings of SIGGRAPH'06, July 2006, 1049–1056

Isenburg M, Liu Y, Shewchuk J, Snoeyink J and Thirion T (2006b) Generating Raster DEM from Mass Points via TIN Streaming. In: Proceedings of Geographic Information Science, LNS 4197:186–198

Jager R (1995) Fuzzy logic in control. The Delft University of technology publishing, Netherlands, Delft, pp. 44–147, ISBN 90-9008318-9

Kabolizade M, Ebadi H, Ahmadi S (2010) An improved snake model for automatic extraction of buildings from urban aerial images and LiDAR data. Comput Environ Urban Syst 34:435–441

Kavzoglu T, Colkesen I (2009) A kernel functions analysis for support vector machines for land cover classification. Int J Appl Earth Obs Geoinf 11(5):352–359

Kavzoglu T, Reis S (2008) Performance analysis of maximum likelihood and artificial neural network classifiers for training sets with mixed pixels. GIScience and Remote Sensing 45:330–342

Lee D, Lee K, Lee S (2008) Fusion of LiDAR and imagery for reliable building extraction. Photogramm Eng Remote Sens 74(2):215–226

Li Y, Wu H, An R, Xu H, He Q, Xu J (2013) An improved building boundary extraction algorithm based on fusion of optical imagery and LiDAR data. Optik 124:5357–5362

Maas HG, Vosselman G (1999) Two algorithms for extracting building models from raw laser altimetry data. ISPRS J Photogramm Remote Sens 54(2–3):153–163

Mathur A, Foody GM (2008) Crop classification by support vector machine with intelligently selected training data for an operational application. Int J Remote Sens 29:2227–2240

Mayer H (1999) Automatic object extraction from aerial imagery—a survey focusing on buildings. Computer Vision Image Understanding 74(2):138–149

Pal M, Mather PM (2005) Support vector machines for classification in remote sensing. Int J Remote Sens 26:1007–1011

Power C, Simms A, White R (2001) Hierarchical fuzzy pattern matching for the regional comparison of land use maps. Int J Geogr Inf Sci 15(1):77–100

RIKS (2005) Map Comparison Kit: product information. Available at: http://www.riks.nl/mck/

Rosin PL (2003) Measuring shape: ellipticity, rectangularity, and triangularity. Mach Vis Appl 14:172–184

Rottensteiner F, Trinder J, Clode S, Kubik K (2005) Using the Dempster-Shafer method for the fusion of LIDAR data and multi-spectral images for building detection. Information Fusion 6(4):283–300

Rottensteiner F, Trinder J, Clode S, Kubik K (2007) Building detection by fusion of airborne laser scanner data and multi-spectral images: performance evaluation and sensitivity analysis. ISPRS J Photogramm Remote Sens 62(2):135–149

Rottensteiner F, Sohn G, Gerke M, Wegner JD, Breitkopf U, Jung J (2014) Results of the ISPRS benchmark on urban object detection and 3D building reconstruction. ISPRS J Photogrammetry and Remote Sensing 93:256–271

Rouse JW, Haas RH, Schell JA and Deering DW (1973) Monitoring Vegetation Systems in The Great Plains with ERTS. Proceedings of the Third Earth Resources Satellite-1 Symposium (Maryland: NASA SP-351) 309–317

Rutzinger M, Rottensteiner F, Pfeifer N (2009) A comparison of evaluation techniques for building extraction from airborne laser scanning. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 2(1):11–20

Sampath A, Shan J (2007) Building boundary tracing and regularization from airborne LIDAR point clouds. Photogramm Eng Remote Sens 73(7):805–812

Shufelt JA (1999) Performance evaluation and analysis of monocular building extraction from aerial imagery. Pattern Analysis and Machine Intelligence, IEEE Transactions on 21(4):311–326

Sohn G, Dowman I (2007) Data fusion of high-resolution satellite imagery and LIDAR data for automatic building extraction. ISPRS J Photogramm Remote Sens 62(1):43–63

Vapnik VN (1995) The nature of statistical learning theory. Springer-Verlag, New York

Vincent L (1993) Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. IEEE Trans Image Process 2:176–201

Vogtle T and Steinle E (2000) 3D Modelling of Buildings Using Laser Scanning and Spectral Information. In: International Archives of Photogrammetry and Remote Sensing, Amsterdam, the Netherlands, Vol. XXXIII, Part B3:927–934

Vosselman G (1999) Building reconstruction using planar faces in very high density height data. International Archives of Photogrammetry and Remote Sensing 32(2):87–92

Vu T, Yamazaki F, Matsuoka M (2009) Multi-scale solution for building extraction from LIDAR and image data. Int J Appl Earth Obs Geoinf 11(4):281–289

Wang Z, Schenk T (2000) Building extraction and reconstruction from LIDAR data. International Archives of Photogrammetry and Remote Sensing 33(B3):958–964

Weidner U (1996) An Approach to Building Extraction from Digital Surface Models. Proceedings of 18th ISPRS Congress, Vienna, Austria. B3:924–929

White R (2006) Pattern based map comparisons. J Geogr Syst 8(2):145–164

Zarea A, Mohammadzadeh A (2016) A novel building and tree detection method from LiDAR data and aerial images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 9(5):1864–1875

Zhang J (2010) Multisource remote sensing data fusion: status and trends. International Journal of Image and Data Fusion 1(1):5–24

Zhao Z, Duan Y, Zhang Y, Cao R (2016) Extracting buildings from and regularizing boundaries in airborne lidar data using connected operators. Int J Remote Sens 37:889–912