Automatic building footprint extraction from high-resolution satellite image using mathematical morphology

Nitin L. Gavankar and Sanjay Kumar Ghosh

Department of Civil Engineering, I.I.T. Roorkee, Roorkee, India

ABSTRACT
Automatic building extraction from High-Resolution Satellite (HRS) image has been an important field of research in the area of remote sensing. Different techniques related to radiometric, geometric, edge detection and object based have already been discussed and used by various researchers for building extraction. However, faithfulness of extraction is highly dependent on user intervention. This study proposes a novel morphological based automatic approach for extraction of buildings using HRS image. Moreover, using such an automatic approach, buildings can be detected having different size and shape. The proposed technique integrates morphological Top-hat filter, and K-means algorithm to extract buildings having bright and dark rooftops. Further, extracted bright and dark rooftop building segments have been combined together to obtain the final output that contains final extracted building segments. In order to eliminate false-detected buildings, different parameters like area, eccentricity, and axis ratio (major/minor axis) have been used. The suitability of the technique has been judged using different indicators such as completeness, correctness and quality.

Introduction
Automatic extraction of buildings has found its applications in various areas like land use land cover mapping, change detection, urban planning, disaster management and many other socio-economic activities. Remote sensing and GIS techniques play a major role in such applications. However, application areas of remote sensing have been considerably increased with availability of sub-metre resolution data from high-resolution earth satellites, such as, IKONOS, WorldView, and QUICKBIRD. HRS imagery facilitates to identify various urban related features, such as, road, tree, building and other natural and man-made structures, present on the earth’s surface. Since, manual extraction of buildings using satellite imagery is time consuming and costly affair, hence, automated methods for building extraction have been emerging as a time and cost effective solution with minimal user involvement.

Automatic extraction of building from satellite image has been always a difficult task for many reasons, such as, building structure and shape, which may vary, or presence of obstacles posed by surrounding objects, such as, trees, high rise buildings, etc. Further, the contrast between roof of building and surrounding region may be low, which has been an important criteria in segmentation and varying roof material reveals different spectral characteristics. Considering all these difficulties, different filters, which have been used to extract edges using satellite imagery, have been broadly classified into three groups, namely, gradient based, laplacian based and morphology based (Katiyar & Arun, 2014). Since, both gradient- and laplacian-based filters are very sensitive to noise, mathematical morphology-based technique has been used in this study to extract buildings using HRS imagery.

Building tops in urban area do not have similar shape, size and texture. However, these buildings have certain common characteristics, such as, their bright appearance and high contrast to the surrounding features. Considering these common building characteristics, the objective of this study is to extract buildings from HRS imagery using morphological Top-hat transformation with minimal human intervention. Moreover, K-means algorithm has been used to determine centroids of predefined classes and “salt and pepper” effect and impulse points from the segmented image have been removed by using median filter. In order to validate performance of proposed methodology, both object-based and pixel-based analysis has been carried out and obtained results have been illustrated using confusion matrix. The proposed methodology has been implemented using MATLAB 13 while ArcMap 10.2.1 has been used for digitization purpose in order to prepare the reference building map.

CONTACT Nitin L. Gavankar nitin.gavankar@walchandsangli.ac.in Department of Civil Engineering, I.I.T. Roorkee, Roorkee, India

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
**Literature review**

Literature reveals a great deal of applications of satellite imagery for feature extraction. A wide range of automatic and semi-automatic techniques for feature extraction have been proposed in different studies. In the following discussion, efforts of such relevant studies have been highlighted.

Lin and Nevatia (1998) have used edge detection-based technique to detect rectangular buildings with flat roofs using geometric and projective constraints from a single-intensity image. The technique further constructs a 3D model of verified buildings using 2D, 3D evidences and shadow cast by buildings. However, this technique detects only rectilinear structures, which limits its utility to the areas covering only rectangular buildings.

Mayer (1999) carried out a comprehensive survey on aerial image-based building extraction techniques. The comparative assessment of different techniques has been done on the basis of certain criteria such as, assessment and complexity, data and its complexity. Further, considering the benefits and limitations of different techniques, a combined model and strategy for object extraction using aerial imagery has been proposed. The proposed model addresses the fundamental issues in object extraction, such as, scale, context and 3D-structure, etc.

Wei, Zhao, and Song (2004), has proposed a supervised clustering and edge detection-based technique to extract large buildings from HRS QUICKBIRD panchromatic imagery having shadow evidence. However, the technique failed to extract small buildings having little or no shadow.

In order to extract small buildings, Jin & Davis (2005) used spectral information along with structural and contextual details using IKONOS satellite imagery of Columbia city, Missouri. In the proposed technique, structural and contextual details have been used to distinguish buildings from parking and other features having similar spectral information. The technique has been found efficient, which extracts 72.7% of building area with a quality percentage 58.8%. However, it is observed that integrating data from structural, contextual, and spectral results has been an extremely complicated procedure.

Advanced morphological operators, such as, Hit or Miss transformation with varying size and shape of structuring elements have been used by Lefèvre, Weber, and Sheeren (2007), to extract buildings from HRS QUICKBIRD panchromatic imagery. The accuracy of the methodology has been computed in terms of precision rate, which has been 88% with a kappa value of 63%. In order to improve the overall accuracy of building extraction procedure, Huang and Zhang (2012) introduced Morphological Shadow Index (MSI) and used along with Morphological Building Index (MBI) [Huang and Zhang (2011)] in an object-based framework. Further, Differential Morphological shadow and building operator have been used by Singh and Mehrotra (2014), to retrieve buildings using HRS GeoEye-1 imagery of the city Washington DC Mall. The Overall Accuracy (OA) of the proposed methodology has been 95.12%.

Huang, Lu, and Zhang (2014) has proposed multi-index learning (MIL) method for HRS images using set of indices, such as, MBI, MSI, and Normalized Difference Vegetation Index (NDVI) to improve classification results over urban areas. Similarly, in order to improve classification accuracy using remote sensing image, Huang et al. (2016) introduced Generalized Differential Morphological Profile (GDMP), which has been found advantageous over traditional Differential Morphological Profile (DMP) [Pesaresi and Benediktsson (2001)]. Further, Benediktsson, Pesaresi, & Amason (2003) investigated methods to pre-process the DMPs, such as, decision boundary feature extraction for neural networks, discriminant analysis feature extraction, and simple sorting feature selection in order to reduce the computational load when DMPs have been used for classification by neural networks. Singh, Maurya, Shukla, Sharma, and Gupta (2012), used NDVI-based segmentation method to separate man-made and natural features and morphological operators have been used to remove false detected regions. However, their approach has been based on the assumption that building has rectangular shape, which may not be true in modern urban scenario.

Koc San and Turker (2010) have used Hough transformation and Support Vector Machine (SVM) classification technique to extract rectangular and circular buildings from HRS panchromatic and pan-sharpened IKONOS imagery. The results have been tested on identified industrial and residential areas in the imagery using parameters, such as, Building Detection Percentage (BDP) and Quality Percentage (QP). For industrial buildings, the values for BDP and QP have been 93.45 and 79.51%, respectively, similarly residential area having rectangular and circular buildings, BDP and QP values for rectangular buildings are 95.34 and 79.05% and 78.74 and 66.81% for circular buildings. In order to improve classification accuracy, Huang and Zhang (2013) proposed Support Vector Machine (SVM) enabled approaches (C-voting, P-fusion, and OBSA), which combines spectral, structural and semantic features for the classification of HRS image. The proposed approaches have been tested at both pixel and object levels and reported that, object-based C-voting and P-fusion technique improve overall accuracy by 0.3–2.0% compared to their pixel-based version.

Dey et al. (2011) have proposed a context-based multi-level segmentation method using pan-sharpened and Multi Spectral (MS) GeoEye-1 imagery of the city Hobart, Australia. The proposed
methodology extracts homogeneous buildings using shadow-object geometry with an accuracy of 71%. Further, building extraction accuracy using pan-sharpened image was 5% greater than that of MS image. However, this methodology cannot extract buildings having little or no shadow, although, these buildings have the considerable area on the ground.

Evaluation of results of automatic building detection has been highly influenced by the evaluation method, since there is no unique definition to identify what constitutes a “correct” detection [Shufelt (1999)] and comparison itself may be carried out by different ways, such as, visual inspection [Vögtle and Steinle (2003)] or automatically [Rottensteiner, Trinder, Clode, and Kubik (2005)]. In order to assess how the evaluation results have been influenced by the evaluation method, Rutzinger, Rottensteiner, and Pfeifer (2009) has carried out a critical overview on various existing pixel-based [Shufelt (1999)] and object-based [Matikainen, Hyppä, and Hyppä (2003); Zhan, Molenaar, Tempfli, and Shi (2005); Rottensteiner et al. (2005); Shan and Lee (2005)] methods for the evaluation of building detection algorithm and proposed a comprehensive evaluation strategy for building detection.

**Study area**

The performance of proposed methodology has been tested on the IKONOS panchromatic image having 0.60-m resolution, captured in the region Santa Ana, California in the USA, Figure 1. The extents of image covers area between latitude 33.69° to 33.71° N and longitude 117.91° to 117.90° E, which is equivalent to an area of about 1.60 sq. km (1.10 km x 1.45 km). Santa Ana is situated in Southern California, adjacent to the Santa Ana River, the second largest metropolitan area in the United States. It is located at Latitude 33.75° N and Longitude 117.87° W.

**Methodology adopted**

The methodology adopted for automatic building extraction has been shown in Figure 2. The proposed technique is based on Top-hat transformation and $K$-means algorithm.

![Figure 1. Study area: Santa Ana, USA.](image-url)
**Image acquisition (I\(_{PAN}\))**

A scene including various size, shape, and varying reflectance of building tops has been selected as input captured in the region Santa Ana, California in USA by HRS IKONOS panchromatic sensor. The scene selected includes various urban features, such as, road, vegetation, shadow, and parking lots.

**Top-hat transformation (I\(_{WTH}\) and I\(_{BTH}\))**

Top-hat transformation using morphological operations has been used to isolate dark and bright regions of an image Serra (1982). Moreover, morphological operations simplify image data, preserve essential shape characteristics, and eliminate irrelevancies (Haralick, Sternberg, & Zhuang, 1987). There are two variants of Top-hat transformation, white Top-hat and black Top-hat, which have been used to extract bright and dark features of an image respectively. However, selection of appropriate size and shape of structuring element (S) plays a key role to identify different features of an image. In this study, white Top-hat and black Top-hat transformation have been applied to an input image (I\(_{PAN}\)) by selecting appropriate size (8) and shape (disk) of
structuring element using MATLAB 13 functions, and used in image enhancement process by using fundamental morphological operations as shown below

White Top-hat transformation

\[ \text{WTH} = I - I \circ S \]  

Black Top-hat transformation

\[ \text{BTH} = I \bullet S - I \]  

where, \((I \circ S)\) and \((I \bullet S)\) are morphological opening and closing operations respectively.

**Image enhancement (I\text{En})**

Image pre-processing has been frequently used to remove noise, make image corrections, and to enhance image. Image enhancement through contrast enhancement has been an important technique in which dark and bright areas present in the image are separated to make image better for human vision (Gupta & Kaur, 2014). Local contrast enhancement is a technique in which original image is added to the difference between Top-hat and bottom-hat transformed image (Ritika, Kaur, & Ritika, 2013). Following relation has been used to enhance the contrast of an image.

\[ I_{\text{En}} = I + \text{WTH} - \text{BTH} \]  

where, \(I_{\text{WTH}}\) and \(I_{\text{BTH}}\) are extracted white and black image regions using white Top-hat and black Top-hat transformations, respectively.

**White top-hat transformation (I\text{WTH})**

The increased contrast of enhanced image further facilitates differentiation between bright and dark features, which improves the feature extraction during Top-hat transformation process. In this step, white Top-hat transformation is applied by selecting appropriate size \(30\) and shape \(\text{disk}\) of structuring element on enhanced image \((I_{\text{enh}})\). Further, K-means algorithm has been used to determine threshold value from the obtained image to separate bright and dark regions of an image.

**K-means clustering algorithm (I\text{k})**

K-means algorithm (Lloyd, 1982) is an unsupervised clustering algorithm and it classifies a given set of data into \(K\) number of clusters in two separate phases. In the first phase, it randomly selects \(K\) centroids and in the second phase, it partitions each pixel to the closest centroid from the respective pixel, where, \(K = 1,2,\ldots,n\) is an user-defined value indicate, the number of predefined classes.

In a typical urban scenario, Building rooftops have different colours, size and shape. However, in panchromatic image building rooftop appears from bright to dark (i.e. different shades of grey scale from 0 to 255 (65,536 for 16 bit image)) depending upon their rooftop (colour) reflectance. Considering this important phenomenon of building rooftops in this study, the value of parameter \(K\) in K-means algorithm has been set to 3 (three) with respect to bright, dark, and intermediate building rooftops respectively. Further, K-means algorithm determines the centroids of three predefined building classes; dark \((T_d)\), intermediate \((T_i)\), and bright \((T_b)\), respectively. In the next step, centroid of intermediate class has been used as a threshold value to divide image into two separate images, dark \((I_d)\) and bright \((I_b)\), respectively. After threshold, following steps (Section 4.6-4.8) have been followed separately for both the images \((I_d\) and \(I_b)\), Figure 2.

**Median filter**

Image filtering is the process to emphasize certain features or to remove undesirable pixels (noise) from an image. However, removing noise and restoring the original image has been one of the most important issues in image processing applications. Although in literature, different filters have been used to remove noise, the standard median filter (Tukey, 1977) has been frequently used in many image processing applications due to its simplicity, edge preserving property and its robustness to impulsive noise. Moreover, median filter of size \(3 \times 3\) has been found to be effective to eliminate “salt and pepper” effect, which has been commonly observed in an image transmission (Zhu & Wang, 2012) and (Tejaswi, Rao, Nair, & Prasad, 2013). The working of nonlinear median filter is to run through a masking window across each pixel entry of an image and replace luminance value of the centre pixel of filtering window with median of luminance values of pixels containing within the window. Further, it has been observed that the median filter separates segments by eliminating thin lines connecting different segments. According to Morgan and Tempfli (2000), elimination of these thin lines is important where 8-direction connectivity is used for segmentation. In addition, median filter further avoids segment merging problem and isolates the segments created.

**Connected component labelling**

Connected component labelling (Morgan & Tempfli, 2000) has been used to identify the connected pixels (segment) and assign a unique label \((C_k)\) for each extracted segment. Moreover, Median filter has been used before connected component
labelling to reduce noise and misclassification problem. Usually segmentation has been done after removing non-interested pixels; however in this study; different parameters have been used to eliminate non-building segments (section 4.8).

**Removal of false-detected building segments**

After connected component labelling, it has been observed that the candidate building segments include non-building objects, such as, road, parking lots, and other urban features. During segmentation, due to similar reflectance properties, these objects have been misclassified in building class. However, in order to filter out these false candidate building segments, initially, all the extracted segments have been numbered \( B_n \), where \( n = 1,2,3, \ldots, n \). In the next step, parameters, such as, axis ratio, eccentricity, and segment area have been used to eliminate false-detected candidate building segments.

**Segment axis ratio (\( S_r \))**

Most of the buildings in an urban scenario are rectangular or square in shape. In order to extract these buildings, Singh and Mehrotra (2014) have used the ratio of the area of the candidate building to the area of its Minimum Enclosing Rectangle (MER) for calculating rectangular fit. However, in this study, following relation has been used to extract rectangular buildings.

\[
S_r = \frac{\text{Major axis length}}{\text{Minor axis length}}
\]  

(4)

Candidate building segment having \( S_r \) value “1” has been assigned square shape, whereas \( S_r \) value greater than “1” has been identified as a rectangular shape building. In order to obtain the threshold value for \( S_r \), considering the general shape of building in an urban scenario, values in the range 1–10 have been tested on the input image and suitable threshold value (7) has been obtained, such that, rectangular and square buildings in the region have been extracted.

**Segment eccentricity (\( S_e \))**

Eccentricity returns a scalar that specifies the eccentricity of an ellipse. The eccentricity is the ratio of the distance between the foci of the ellipse and its major axis length and its value lies in between 0 to 1. An ellipse having an eccentricity “0” (zero) is actually a circle while an ellipse having an eccentricity “1” is a line segment. In order to obtain the threshold value for eccentricity, different shapes of circle (or ellipse) have been plotted with respect to the eccentricity value in the range 0 to 1. Further, by observing these plotted circle (or ellipse), the highest elliptical shape which can be a building rooftop shape has been identified and respective eccentricity value (0.85) has been selected as respective eccentricity value. The candidate building segments having eccentricity below the threshold value have been accepted as circular buildings whereas the candidate building segments having eccentricity greater than threshold value have been eliminated as misclassified building segments.

**Segment area (\( S_{aw} \))**

The area of each candidate building segment has been the number of pixels in the corresponding connected region (Singh & Mehrotra, 2014). In general, false-detected non-building urban features, such as, road and barren land are usually either elongated or cover large area on the ground and very small features, such as, vehicles covers very small area on the ground. However, buildings in urban area usually cover considerable area on ground at the same time their area on ground is not very large. Considering general building area in the study area, appropriate threshold range (400–40,000 pixels) has been set, such that very small features, such as, vehicle and very large features, such as, road, barren land, etc. have been eliminated. Building candidates that have a value within the threshold range (Maximum and Minimum) have been accepted as true buildings. However, threshold range may vary depending on resolution of an image and residential or industrial urban scenario under consideration.

**Segment merging and segment edge detection**

After removing misclassified candidate building segments by using above parameters, the dark and bright roof-top building segments have been obtained in two separate images, Figure 2. Further, these separately extracted building segments have been combined together to get a final extracted building segment image. Moreover, small holes present in the extracted building segments have been filled up and edges of all extracted building segments have been extracted by using Boundary Tracing function in MATLAB 13. In the next step, these boundaries have been overlaid on the input image \( I_{PAN} \) to get the final output image \( I_{PAN} \).

**Quality assessment**

The quality of building extraction technique has been validated by examining quality measures, such as, Completeness, Correctness and Quality (Singh et al., 2012). Before applying these measures, the extracted candidate building segments have been categorized in to three classes as explained below.

a. **True Positive (TP)** are the buildings among the extracted buildings that are actually buildings.
b. **False Positive (FP)** are the buildings among the extracted buildings that are not buildings and falsely extracted as a building.

c. **False Negative (FN)** are the buildings that are actually the true buildings; however, they are not extracted and left as non-building area.

After categorization, the quality measures have been applied, which are defined as below.

(I) **Completeness** returns the number of true building segments extracted from the actual number of buildings in a scene. It has been calculated as

\[
\text{Completeness} = \frac{\text{Number of extracted buildings that are actually buildings}}{\text{Total number of buildings in a scene}} \approx \frac{\text{TP}}{\text{TP} + \text{FN}} \quad \text{Completeness} \in [0; 1] \quad (5)
\]

(II) **Correctness** determines the chance of being a building of each extracted candidate building segment and it has been calculated as

\[
\text{Correctness} = \frac{\text{Number of extracted buildings that are actually buildings}}{\text{Total number of extracted buildings}} \approx \frac{\text{TP}}{\text{TP} + \text{FP}} \quad \text{Correctness} \in [0; 1] \quad (6)
\]

(III) **Quality** tells us how good approach is. It has been the measure of both completeness and correctness. It can be calculated as

\[
\text{Quality} = \frac{\text{Number of extracted buildings that are actually buildings}}{\text{Number of buildings extracted in output}} \approx \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad \text{Quality} \in [0; 1] \quad (7)
\]

**Object-based evaluation**

The main objective of object-based evaluation is to identify the number of true building segments (TP), misclassified non-building segments (FP), and non-extracted true buildings (FN) by comparing extracted building segments and reference building map. The reference building map has been obtained by digitizing input satellite image (I_{PAN}) and further used for comparison. In general, during object-based evaluation, the extracted building segments which has a certain minimum overlap, typically 50–70%, with buildings in the reference building map have been accepted as TP [[Matikainen et al. (2003); Zhan et al. (2005); Rottensteiner et al. (2005)]. However, in this study, the extracted building segments which has overlap greater than 60% (Average of range 50–70) have been identified TP, while remaining extracted segments have been considered as FP. Further, the evaluation parameters completeness, correctness, and quality obtained by object-based analysis have been denoted by Completeness_{obj}, Correctness_{obj}, and Quality_{obj}, respectively.

**Pixel-based evaluation**

During object-based evaluation strategy, there has been no such unique criterion to decide whether an object has been TP or not. Further, Shufelt (1999) considered a pixel-based metrics which has been more objective than object-based metrics. During pixel-based analysis, the raster representation of the extracted results has been compared with the reference image [Rutzinger et al. (2009)]. The evaluation of results has been presented as an image, which represent the spatial distribution of the TP, FP and FN. Further, the evaluation parameters completeness, correctness, and quality obtained by pixel-based analysis have been denoted by Completeness_{pix}, Correctness_{pix}, and Quality_{pix}, respectively.

**Result and discussion**

The proposed automatic building extraction technique has been based on Top-hat transformation and K-means algorithm. The Top-hat transformation has been used to separate dark and bright building segments, while, K-means algorithm determines the threshold value, which divides the dark and bright building segments into two separate images (I_d and I_b). However, misclassification problem, which has been commonly observed in feature extraction, has been removed by using parameters, such as, axis ratio, eccentricity and segment area. Figure 3(a-c) shows the enlarged view of results of removing misclassified buildings after using parameters. Thereafter, extracted bright and dark roof-top building segments have been shown (Figure 4(a) and Figure 4(b)), respectively.

Further, these separately extracted building segments have been combined together to obtain a final extracted building image (Figure 5(a)). Here, the boundaries of all extracted building segments have been determined and overlaid on the input image (I_{PAN}) to get the final output (Figure 5(b)). Thereafter, buildings in the original input image have been digitized using ArcMap 10.2.1 (Figure 5(c)) and used further as a reference building map during evaluation of result. The evaluation of result of automatically extracted buildings has been carried out by using both, object-based and pixel-based analysis as discussed in Section 4.11 and 4.12.
Object-based evaluation

During object-based evaluation, the digitized reference building map, Figure 5(c), has been compared visually [Singh et al. (2012); Heipke, Mayer, Wiedemann, and Jamet (1997)] with extracted final output image, Figure 5(b). The results of visual comparison have been shown in Table 1 and used further in analysis and assessment of quality. False-detected non-building regions and non-extracted true buildings have been shown in Figure 6(a) and Figure 6(b) respectively.

Here, 14 non-building urban features have been misclassified in building class due to their similar reflectance and structural properties to that of true buildings, while, 12 buildings, having similar reflectance value to that of the road and other surrounding urban features has been eliminated. Further, Values of completeness, correctness and quality have been computed using above Equations (5, 6, and 7) and results obtained has been given in Table 2.

The computed values of completeness, correctness and quality together indicate the performance of the methodology adopted. High value of completeness
(0.92) ensures that all the true buildings have been extracted. Further, high value of correctness (0.90) shows that very few non-building urban features have been falsely detected as buildings. However, quality has been the measure of both completeness and correctness, which indicates total error in building extraction methodology. In this case, high value of quality (0.83) confirms that true buildings have been successfully extracted from the input image with minimum error.

**Pixel-based evaluation**

In order to critically analyse the performance of proposed methodology, pixel-based evaluation [Rutzinger et al. (2009)] has been carried out. During pixel-based analysis, the raster representation of the extracted results has been compared with the reference image and extracted pixels have been classified into three categories TPs, FPs and FNs. Figure 7 shows the pixels classified into three respective categories, TPs: Green, FPs: Red and FNs: Black, using pixel based evaluation. Further, values of completeness, correctness and quality have been computed using above Equations (5, 6, and 7) and result obtained has been given in Table 3.

The obtained values for the parameters completeness, correctness, and Quality using pixel-based analysis have been 0.86, 0.89 and 0.78, respectively, which are acceptable when compared with Rutzinger et al. (2009).
Furthermore, in order to validate performance of proposed methodology, methodology has also been tested on a scene selected from another IKONOS panchromatic image having 0.60m resolution acquired in the region Naples, Florida in USA, Figure 8(a). The region selected includes buildings having various size, shape, and reflectance, which are morphologically different than buildings in first image, Figure 1. Further, the proposed methodology shown in Figure 2 has been applied on input image Figure 8(a) and final extracted building image has been shown in Figure 8(b).

Thereafter, evaluation of obtained result has been carried out by using both, object-based and pixel-based analysis and has been shown in Tables 4, 5 and 6, respectively. Figure 9 shows the pixels classified into three respective categories, TPs: Green, FPs: Red and FNs: Black, using pixel-based evaluation technique.

The obtained values for the parameters completeness, correctness, and Quality using object-based and pixel-based analysis are acceptable when compared with Rutzinger et al. (2009).

**Conclusion**

Automatic building extraction has been an important area of research in remote sensing. It has been contributing to various applications such as land-use land cover mapping, disaster management and many other socio-economic activities. Proposed methodology successfully extracts buildings having different size and shape with minimal human intervention. Further, the problem of misclassification has been removed by using parameters, such as, axis ratio, eccentricity and segment area.

Interpretation of obtained result shows that buildings with very bright and dark roof tops have been

| Region  | Total buildings in the scene | Extracted true buildings | Not extracted true buildings | Non-building objects extracted |
|---------|----------------------------|--------------------------|----------------------------|--------------------------------|
| Naples  | 45                         | 41                       | 4                          | 5                              |

| Region  | TP (pixels) | FP (pixels) | FN (pixels) | Completeness_obj | Correctness_obj | Quality_obj |
|---------|-------------|-------------|-------------|------------------|-----------------|-------------|
| Naples  | 41          | 5           | 4           | 0.91             | 0.89            | 0.82        |

**Table 3. Result obtained by applying quality parameters using pixel-based evaluation.**

| Region  | TP (pixels) | FP (pixels) | FN (pixels) | Completeness_pix | Correctness_pix | Quality_pix |
|---------|-------------|-------------|-------------|------------------|-----------------|-------------|
| Santa Ana | 914,305    | 113,397     | 150,204     | 0.86             | 0.89            | 0.78        |

**Table 4. Result obtained by visual comparison.**

| Region  | Total buildings in the scene | Extracted true buildings | Not extracted true buildings | Non-building objects extracted |
|---------|----------------------------|--------------------------|----------------------------|--------------------------------|
| Naples  | 45                         | 41                       | 4                          | 5                              |

Figure 8. (a) Input image (b) Final extracted building image.
extracted successfully. However, few building roof tops revealing similar reflectance value to that of the road and other surrounding features in the image have been eliminated during misclassified object removal procedure. Similarly, few non-building features have been classified in building class due to their similar reflectance value and structural properties to that of the true buildings. Overall accuracy of object-based evaluation technique in terms of completeness, correctness, and quality has been 0.92, 0.90, 0.83 and 0.91, 0.89, 0.82 for the regions Santa Ana and Naples, respectively. However, overall accuracy of pixel-based evaluation technique in terms of completeness, correctness, and quality has been 0.86, 0.89, 0.78 and 0.82, 0.87, 0.74 for the regions Santa Ana and Naples, respectively, which are acceptable when compared with Rutzinger et al. (2009).

The proposed automatic building extraction methodology is simple, fast, and effective, which does not require any extra information, such as, training samples and digital elevation model and yields high accuracy. Proposed methodology can be further used for various applications, such as damage estimation by identifying damaged and undamaged buildings and calculation of density of buildings in the region.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**ORCID**

Nitin L. Gavankar http://orcid.org/0000-0002-8511-5143
Sanjay Kumar Ghosh http://orcid.org/0000-0001-7849-9313

**References**

Benediktsson, J.A., Pesaresi, M., & Amason, K. (2003). Classification and feature extraction for remote sensing images from urban areas based on morphological transformations. *IEEE Transactions on Geoscience and Remote Sensing*, 41(9), 1940–1949. doi:10.1109/TGRS.2003.814625

Dey, V., Zhang, Y., & Zhong, M., 2011, August. Building detection from pan-sharpened GeoEye-1 satellite imagery using context based multi-level image segmentation. In *Image and Data Fusion (ISIDF)*, 2011 *International Symposium on* (pp. 1–4). IEEE.

Gupta, S., & Kaur, Y. (2014). Review of different local and global contrast enhancement techniques for a digital image. *International Journal of Computer Applications*, 100(18), 18–23. doi:10.5120/17625-8384

Haralick, R.M., Sternberg, S.R., & Zhuang, X. (1987). Image analysis using mathematical morphology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (4), 532–550. doi:10.1109/TPAMI.1987.4767941

Heipke, C., Mayer, H., Wiedemann, C., & Jamet, O. (1997). Evaluation of automatic road extraction. *International Archives of Photogrammetry and Remote Sensing*, 32(3 SECT 4W2), 151–160.

Huang, X., Han, X., Zhang, L., Gong, J., Liao, W., & Benediktsson, J.A. (2016). Generalized differential morphological profiles for remote sensing image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(4), 1736–1751. doi:10.1109/JSTARS.2016.2524586

Huang, X., Lu, Q., & Zhang, L. (2014). A multi-index learning approach for classification of high-resolution remotely sensed images over urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 90, 36–48. doi:10.1016/j.isprsjprs.2014.01.008

Huang, X., & Zhang, L. (2011). A multidirectional and multiscale morphological index for automatic building extraction from multispectral GeoEye-1 imagery. *Photogrammetric Engineering & Remote Sensing*, 77(7), 721–732. doi:10.14358/PERS.77.7.721

**Table 6.** Result obtained by applying quality parameters using pixel-based evaluation.

| Region | TP (pixels) | FP (pixels) | FN (pixels) | Completeness _pix_ | Correctness _pix_ | Quality _pix_ |
|--------|------------|------------|------------|-------------------|------------------|--------------|
| Naples | 51,230     | 7425       | 11,002     | 0.82              | 0.87             | 0.74         |

**Figure 9.** Result obtained by pixel-based evaluation, Green: TPs; Red: FPs; Black: FNs.
Huang, X., & Zhang, L. (2012). Morphological building/shadow index for building extraction from high-resolution imagery over urban areas. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(1), 161–172. doi:10.1109/JSTARS.2012.2168195

Huang, X., & Zhang, L. (2013). An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 51(1), 257–272. doi:10.1109/TGRS.2012.2202912

Jin, X., & Davis, C.H. (2005). Automated building extraction from high-resolution satellite imagery in urban areas using structural, contextual, and spectral information. *EURASIP Journal on Advances in Signal Processing*, 2005(14), 1–11. doi:10.1155/ASP.2005.2196

Katiyar, S.K., & Arun, P.V., 2014. Comparative analysis of common edge detection techniques in context of object extraction. *arXiv preprint arXiv:1405.6132*

Lefèvre, S., Weber, J., & Sheeren, D. (2007, April). Automatic building extraction in VHR images using advanced morphological operators. In *Urban remote sensing joint event*, 2007 (pp. 1–5). IEEE.

Lin, C., & Nevatia, R. (1998). Building detection and description from a single intensity image. *Computer Vision and Image Understanding*, 72(2), 101–121. doi:10.1006/cviu.1998.0724

Lloyd, S.P. (1982). Least squares quantization in PCM. *Information Theory, IEEE Transactions On*, 28(2), 129–137. doi:10.1109/TIT.1982.1056489

Matikainen, L., Hyppä, J., & Hyppä, H. (2003). Automatic detection of buildings from laser scanner data for map updating. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 34(3/W13), 218–224.

Mayer, H. (1999). Automatic object extraction from aerial imagery—A survey focusing on buildings. *Computer Vision and Image Understanding*, 74(2), 138–149. doi:10.1006/cviu.1999.0750

Morgan, M., & Tempfli, K. (2000). Automatic building extraction from airborne laser scanning data. *International Archives of Photogrammetry and Remote Sensing*, 33(B3/2; PART 3), 616–623.

Pesaresi, M., & Benediktsson, J.A. (2001). A new approach for the morphological segmentation of high-resolution satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 39(2), 309–320. doi:10.1109/36.905239

Ritika, Sandeep, K, & Ritika, R. (2013). Contrast enhancement techniques for images-A visual analysis. *International Journal of Computer Applications*, 64(17), 20–25. doi:10.5120/107277-5679

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. doi:10.1016/j.inffus.2004.06.004

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. doi:10.1109/JSTARS.2009.2012488

San, D.K., & Turker, M. (2010). Building extraction from high resolution satellite images using Hough transform. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science*, 38(Part 8), 1063–1068.

Serra, J. (1982). *Image analysis and mathematical morphology*, v. 1. Academic press London, 1982. ISBN 012637242X, pp. 610.

Shan, J., & Lee, S.D. (2005). Quality of building extraction from IKONOS imagery. *Journal of Surveying Engineering*, 131(1), 27–32. doi:10.1061/(ASCE)0733-9453(2005)131:1(27)

Shufelt, J.A. (1999). Performance evaluation and analysis of monocular building extraction from aerial imagery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(4), 311–326. doi:10.1109/34.761262

Singh, D., Maurya, R., Shukla, A.S., Sharma, M.K., & Gupta, P.R., 2012, March. Building extraction from very high resolution multispectral images using NDVI based segmentation and morphological operators. In *Engineering and Systems (SCES), 2012 Students Conference on* (pp. 1–5). IEEE.

Singh, K.K., & Mehrotra, A., 2014, March. Building extraction from VHR imagery using morphological shadow and building operator. In *Computing for Sustainable Global Development (INDIACom), 2014 International Conference on* (pp. 463–467). IEEE.

Tejaswi, K.P., Rao, D.S., Nair, T., & Prasad, A.V.V., 2013. GPU accelerated automated feature extraction from satellite images. *arXiv preprint arXiv:1304.3992*

Tukey, J.W. (1977). *Tukey exploratory data analysis*. Reading, MA: Addison, Wesley.

Vögtle, T., & Steinle, E. (2003). On the quality of object classification and automated building modeling based on laserscanning data. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 34(Part 3), W13.

Wei, Y., Zhao, Z., & Song, J., 2004, September. Urban building extraction from high-resolution satellite pan-chromatic image using clustering and edge detection. In *Geoscience and Remote Sensing Symposium, 2004. IGARSS’04. Proceedings. 2004 IEEE International (Vol. 3, pp. 2008–2010). IEEE.

Zhan, Q., Molenaar, M., Tempfli, K., & Shi, W. (2005). Quality assessment for geo-spatial objects derived from remotely sensed data. *International Journal of Remote Sensing*, 26(14), 2953–2974. doi:10.1080/0143116050057764

Zhu, R., & Wang, Y. (2012). Application of improved median filter on image processing. *Journal of Computers*, 7(4), 838–841. doi:10.4304/jcp.7.4.838-841