REVIEW ARTICLE

Building detection methods from remotely sensed images

Naveen Chandra[1,]* and Himadri Vaidya[2]

[1]Wadia Institute of Himalayan Geology, 33-GMS Road, Dehradun 248 001, India
[2]Formerly Uttarakhand Technical University, Dehradun 248 007, India

With the availability of high-resolution satellite imagery, new applications have been developed for solving geospatial issues in urban regions. Building detection from remote sensing images has been an active area of research due to its broad range of applications, including city modelling, map updating and urban monitoring. The manual processing of an image is a time-consuming and laborious task. Therefore, researchers have developed methods that involve less or no human effort. At present, building detection has improved through various automated and semi-automated methods/algorithms/ techniques suggested in various studies. The objective of the present study is to review the efforts of such studies. Here, the building detection methods are categorized into six groups: (i) low-level feature-based methods, (ii) snake models, (iii) graph-based methods, (iv) shadow detection-based methods, (v) cognition-based methods and (vi) deep learning models. We hope that this study will aid the researchers working in this domain.

Keywords: Building detection, classification, geospatial issues, remote sensing images, urban areas.

REMOTE sensing technology provides the geospatial data of a large geographical region. The overall time and cost of generating the data with remote sensing approaches are lower compared to other methods[1]. Earlier, it was difficult to extract man-made and natural objects from the images acquired from satellites (Landsat) due to their resolution. However, with the emergence of very high-resolution (VHR) satellite images (QuickBird and IKONOS), the difference between objects present on the surface of the earth can be observed[2]. Remote sensing images (RSIs) are a key and valuable source of information for object detection. It has been estimated that more than 50% of the world's population lives in the suburban and urban regions. Thus accurate and reliable detection of buildings from RSIs is a prime task for various applications such as urban mapping, military intelligence, map-making, change monitoring, damage detection, estimation of population and land-use/land-cover analysis[3–12]. Human experts are unable to label the buildings in RSIs due to their complex geometrical properties (shape and size) and also since buildings may be surrounded by other objects like trees[4,9]. Remote sensing data capture a large geographical region and hence labelling each building in the image is a tedious and time-consuming process[4,9]. Also, the key interpretation elements (contrast, resolution and illumination) may not be sufficient to detect buildings from RSIs[4,9]. The present study provides a summary of several building detection techniques from RSIs over the last 30 years.

Building detection methods

In broad terms, the process of building detection from satellite images has been divided into two parts, namely object- and threshold-based. The object-based approach generates segments and characterizes them through features (shape, spectral and height). The threshold-based approach generates normalized difference vegetation index (NDVI) and digital surface model (DSM) to detect buildings.

The structure of buildings in two-dimensions (2D) and three-dimensions (3D) has a tremendous impact globally. Hence the methods/tools developed for the extraction and detection of both are different. In the past, many methods/ algorithms have been introduced for 2D building extraction. However, few articles have been published describing their limitations and capabilities. Mayer[13] provides a review of the building detection methods developed until mid-1990s. The review consists of a summary of the strategies and models of the developed methods. The details of the developed methods are given in Table 1. A survey on the type of knowledge being used for object detection from satellite images is given in Baltsavias[14]. It focuses on the issues encountered while using and upgrading the existing knowledge. It also provides a crisp review of the current trends in image analysis.

On the other hand, 3D building extraction is a different approach. It can provide the vertical as well as horizontal information of a particular area/city, which can be obtained from the stereo-mapping-based satellites. Most of the research focuses on the 2D level because access to 3D dataset is limited and expensive. A review emphasizing light detection and ranging (LIDAR)-based reconstruction approaches along with their achievements has been presented by Brenner[15]. It covers a detailed review of the semiautomatic[16–19] and automatic[20–29] reconstruction methods and their properties. Unsalan and Boyer[30] extend the review presented in Mayer[13] by providing a comparative evaluation

*For correspondence. (e-mail: naveenchandra0408@gmail.com)

Table 1. Description of previous works on building detection methods from satellite images[8,9]

Author	Data type	Image type	Article type	Year
Huertas and Nevatia	ABD	GSI	RsC	1988
Irvin and Mckeown	ABD	GSI	RsC	1989
Liow and Pavlidis	ABD	GSI	RsC	1990
Shufelt and Mckeown	ABD	GSI	RsC	1993
McGlone and Shufelt	ABD	GSI	RsC	1994
Weinder and Forstner	ABD	GSI	RsC	1995
Krishnamachari and Chellappa	ABD	GSI	RsC	1996
Baillard	ELD	GSI	RsC	1998
Zang	SBD	MSI	RsC	1999
Helmut Mayer	SBD and ABD	GSI and MSI	RvC	1999
Stassopoulou and Caelli	ABD	GSI	RsC	2000
Cord *et al.*	ELD	GSI	RsC	2001
Ruther *et al.*	ELD	GSI	RsC	2002
Lee *et al.*	SBD	GSI	RsC	2003
Benediktsson *et al.*	SBD	GSI	RsC	2004
Baltsavias	SBD and ABD	GSI and MSI	RvC	2004
J. Peng and Y. C. Liu	ABD	GSI	RsC	2005
Unsalan and Boyer	SBD and ABD	GSI and MSI	RvC and RsC	2005
Brenner	SBD and ABD	MSI and LDD	RvC and RsC	2005
Hongjian and Shiqiang	LDD	GSI	RsC	2006
Sohn and Dowman	SBD and ABD	MSI	RsC	2007
Katartzis and Sahli	ABD	MSI	RsC	2008
Karantzalos and Paragios	SBD and ABD	GSI	RsC	2009
Salman Ahmadi *et al.*	ABD	GSI	RsC	2010
Haala and Kada	ABD and ELD	LDD	RvC	2010
Cui *et al.*	ABD	MSI	RsC	2011
Tack *et al.*	SBD	MSI	RsC	2012
Mohammad Izadiand and Parvaneh Saeedi	SBD	MSI	RsC	2012
Senaras *et al.*	SBD	MSI	RsC	2013
Ali Ozgun Ok	SBD	MSI	RsC	2013
Ali Ozgun Ok *et al.*	SBD	MSI	RsC	2013
Lihong Kang *et al.*	SBD	MSI	RsC	2014
Jiaojiao Tian *et al.*	SBD	MSI	RsC	2014
Kovacs and Ali Ozgun Ok	SBD	MSI	RsC	2015
Yansheng Li *et al.*	SBD	GSI	RsC	2015
Caglar Senaras and Fatos T. Yarman Vural	SBD	MSI	RsC	2016
Gregoris Liasis and Stavros Stavrou	SBD	GSI	RsC	2016
Gong Cheng and Junwei Han	ABD and SBD	GSI and MSI	RvC	2016
Ali Ozgun Ok	SBD	MSI	RsC	2016
D. Chaudhuri *et al.*	SBD	MSI	RsC	2016
N. Chandra and J. K. Ghosh	SBD	MSI	RsC	March 2017
N. Chandra and J. K. Ghosh	SBD	MSI	RsC	August 2017
Dimitrios Konstantinidis *et al.*	SBD	MSI	RsC	2017
N. L. Gavankar and S. K. Ghosh	SBD	MSI	RsC	2018
Masayu Norman *et al.*	SBD	MSI	RsC	2019
N. L. Gavankar and S. K. Ghosh	SBD	MSI	RsC	2019
S. Shirowzhan *et al.*	ABD	LDD	RsC	2020
X. Wang and P. Li	ABD	LDD	RsC	2020
Huiwei Jiang *et al.*	SBD	MSI	RsC	2020
Meng Chen *et al.*	SBD	MSI	RsC	2021
Khaled Moghalles *et al.*	SBD	MSI	RsC	2021
Christian Ayala *et al.*	SBD	MSI	RsC	2021

ABD, Airborne data; SBD, Spaceborne data; LDD, Lidar data; ELD, Elevation data; GSI, Grey scale images; MSI, Multispectral images; RsC, Research communication; RvC, Review communication.

of the proposed methods until 2003. Haala and Kada[31] present a review of the methods developed for building reconstruction using LIDAR and airborne elevation data. According to them, the reconstruction of buildings is based on three modules: (i) parametric shapes, (ii) segmentation and (iii) DSM simplification[31]. The LIDAR data have been used by Shirowzhan *et al.*[32] to determine the height of buildings using data mining techniques. 3D building extraction has a wide range of applications, such as urban expansion, estimation of population and urban climate.

Although several researchers have categorized the building detection methods (based on geometry, contours and shadow), it is hard to classify them due to their various applications. Here, we attempt to categorize and summarize the developed building detection methods.

Low-level feature-based methods

A method/approach to produce 3D hypotheses has been presented by Shufelt[33] using a single view for building extraction from aerial images. The author describes the impact of photogrammetric models with respect to PIVOT (perspective interpretation of vanishing points for objects in 3D). The potential of the method has been evaluated through quantitative as well as qualitative analysis of the results. Zhang[34] has attempted to extract buildings from an urban area. The method is divided into two parts: (i) multispectral classification and (ii) improving the classified results through matrix-based filtering which calculates the contrast, energy, entropy and homogeneity. The results of the matrix-based filtering were compared with the other available filtering methods. The proposed method was validated using the TM-SPOT merged dataset of Shanghai, China. A morphological based approach has been presented by Pesaresi and Benediktsson[35]. They performed the closing and opening operations using the reconstruction technique. The well-known watershed segmentation was used by these authors. The object and pixel-based classification methods have been widely used independently for land-cover detection. Generally, the object-based approach outperforms the pixel-based approach[36]. Therefore, a combination of object and pixel-based approaches has been described by Shackelford and Davis[37]. In the pixel-based classification, they employed maximum likelihood classifier and hierarchal fuzzy classifier that uses spectral and spatial information. The IKONOS imagery was classified into seven classes (buildings, roads, trees, grass, water, shadow and bare soil). It was observed that the hierarchal fuzzy classifier produced better results in comparison to the maximum likelihood classifier. Further, the results of pixel-based classification were refined within the object-based classification, which was performed using the theories of fuzzy logic and multi-resolution segmentation, which includes information related to spectral and spatial heterogeneity.

Benediktsson *et al.*[38] studied the mathematical morphological operations for feature extraction and classification of high-resolution satellite (HRS) images. They also explored the areas of neural networks for the classification of high-resolution IKONOS and IRS-1C images. Two approaches were used, namely discriminant analysis feature extraction (DAFE) and decision boundary feature extraction (DBFE) for feature extraction within the neural network. The test images were classified into seven categories, namely small buildings, large buildings, roads outside the urban region, roads within the urban region, open space outside the urban region, open space within the urban region and wastelands.

A system to detect houses and streets from multispectral satellite images was introduced by Unsalan and Boyer[30]. It contains four components: (i) processing and analysis of multispectral information, (ii) segmentation of the input data using *k*-means clustering algorithm through the combination of spectral and spatial features, (iii) decomposing of the segmented image by binary balloon algorithm and (iv) implementation of the graph-based theoretical algorithm for detecting houses and streets from IKONOS images. The developed system is valuable for automated map generation. The high-level and low-level geometry features of the input images were used to detect man-made objects from satellite images[39]. Further, these features were classified using the supervised learning approach, i.e. support vector machine (SVM). The proposed method was tested on SPOT5 THR images having ten classes. Genetic algorithms are generally used in search problems and are now a standard optimization technique with applications in different fields[40]. An adaptive fuzzy-based genetic algorithm has been proposed by Sumer and Turker[40] to determine the textural and spectral attributes from different bands (red, green, blue and near-infrared) of the image through Fisher's linear discriminant analysis, which is widely used in machine learning and statistics. Then the various operations (cross, selection and mutation) of genetic algorithms are performed. The performance of these operations is improved using the fuzzy logic controller. Lastly, the morphological operations (opening and closing) are carried out to complete the stages of post-processing. The validation of the proposed approach is done on ten test scenes (having different characteristics) of Turkey.

To detect buildings from the QuickBird satellite image, a novel method was proposed based on the theories of decision fusion in which after segmentation (using a mean-shift algorithm), various features of the image such as shape, colour and texture were classified within the newly proposed framework known as fuzzy stacked generalization[41]. Further, the potential of the proposed method was identified by a comparison of the results with different machine learning algorithms. A geometrical feature plays an important role in object detection. Therefore a method was proposed using height information to determine the building change detection through stereo images[42]. The method helped to generate digital surface models of the stereo images of two test regions (an urban region in Germany and an industrial region in Korea) and improved accuracy through the fusion theories of Dempster–Shafer.

Zhang *et al.*[43] developed another morphologically based framework. The structure of the buildings was determined using the morphological building index (MBI) method. During post-processing, the morphological spatial pattern was utilized to improve the results obtained from MBI. The images of WorldView-2 and GeoEye-1 were used to perform the experiment for demonstrating the robustness of the developed framework.

A two-staged model for detecting buildings from multispectral satellite images of QuickBird and WorldView-2 has been proposed[5]. In the first module, the model concatenates the local binary patterns (LBP) and histogram of oriented gradients (HOG) features, introducing a distance function that is trained using the well-known supervised learning algorithm, i.e. SVM for calculating the distance between LBP and HOG descriptors. The EM algorithm is employed in the second module known as 'region refinement' for detecting the rectangular-shaped regions representing buildings[44].

Recently, an automated method to extract footprints of buildings of dissimilar size and shape from HRS utilizing mathematical morphological operations has been proposed[10]. Another approach for detecting building footprints has been presented using an object-based method concentrating on shape parameters[11]. An IKONOS multispectral image was used to determine the completeness and correctness of the obtained results. Similarly, an object-based approach was used for extracting building footprints from the Worldview 3 image[45]. A segmentation-based approach that integrates the spatial plateau objective function and Taguchi statistical method was proposed to detect buildings from Worldview 3 images[45]. The various parameters (shape, scale and compactness) for segmentation classified the images into five classes, namely roads, buildings, trees, grass and water using the eCognition software.

Snake models

The snake model, also known as active contour model, was initially developed by Kass *et al.*[46]. The model includes dynamics curves present in the image for capturing its features. The motion of the curve is led by the external and internal forces, i.e. whenever the minimum energy state is achieved, the curve reaches the desired image boundaries. Snake models are divided into two categories: geometrical and parametric snakes. Geometrical snakes are referred to as zero-level sets in which updating is performed on the surface function in the image domain[47]. Geometrical snakes are further divided into two groups: region-based and edge-based active contours. Region-based active contours depend on the spatial properties (texture and intensity) of the objects. This method relies on the Mumford–Shah function for image segmentation[48–52]. The boundaries of the objects are located using the gradient information in edge-based active contours[53–59]. Conversely, parametric snakes are described as parameterized contours in which the evolution of a snake is accomplished on the predefined control points. A key limitation of this method is that it cannot change the topologies during the evolution of the snake and the contour must be near the desired boundary of the object[47,60,61]. The snake model has applications in image segmentation, contour location, edge detection and visual tracking[47].

The boundaries of buildings from LIDAR data were estimated and their accurate position was determined using the snake model[62]. Buildings were also detected from QuickBird images using a semi-automated algorithm[63]. Initially, a point within the boundary of the buildings was selected and then accurate boundaries of the buildings were extracted by reproducing the curve through an iterative approach. A traditional snake model was modified based on the geometric and radiometric characteristics of the buildings in aerial images using two parameters, i.e. selecting the initial seeds and external energy function[64]. This method is capable of assessing the shape of buildings. However, it is unable to extract the buildings present in urban regions. The shape accuracy achieved for detected buildings is 83.60%. An improved Chan–Vese model for extracting man-made objects from seven aerial images was proposed by Cao and Yang[65]. The method is implemented using fractal error metrics and a three-staged segmentation algorithm. Man-made objects are detected by changing the active contours. Prior knowledge of the shape of the buildings was incorporated with active contours for detecting buildings from the satellite as well as aerial images using level set-based segmentation methods[66]. The accuracy obtained was more than 80%. Recently a new level set method has been developed for calculating the energy function to detect buildings from the high-resolution aerial images of Lavasan (central Iran)[67]. This snake model can detect the boundaries of buildings, avoiding the edges of other objects present in the image. This approach requires additional information (height) for detecting buildings. The completeness and correctness of the extracted buildings are reported as 80% and 96% respectively.

A novel approach for extracting buildings was proposed utilizing an active contour model along with the colour feature[7]. The development of the model was carried out in three stages: (i) initialization of active contours, (ii) representation of HSV and RGB colour spaces and (iii) optimization of the proposed model. The model was assessed on 96 Google Earth images of different countries. It gave better results in comparison with other active contour-based models.

Graph-based methods

Krishnamachari and Chellappa[68] used Markov random field (MRF) for grouping the line segments to delineate buildings of particular shapes (rectangular). Later, active contours were utilized for improving the shapes of the segments. This method was tested on aerial images, including a qualitative assessment of the results. A robust method for building detection has been proposed by Kim and Muller[69] and implemented in four stages. First, line extraction; second, generation of line relation graph; third, generating building hypothesis based on the graph structure, and lastly, verification of building hypothesis. The robustness of the algorithm was determined by considering the geometrical and mathematical relations during the generation of the hypothesis. The method was validated using aerial photographs. However, the method was found to be applicable only to buildings of specific shapes. A right-angle graph method was proposed for detecting right angled-shaped buildings[70]. It was based on pose clustering, which is a voting process (the combination of voting elements and voting rules). During hypothesis generation, right-angle edges and Hough space of the buildings were incorporated. The model was validated using real aerial images and synthetic images. The building detection percentage was more than 80. This method could only be used for regular buildings in urban regions. The graph-based method was also used to detect streets and houses of North America[30]. A combination of 2D and 3D information from the airborne and synthetic images was used to detect building rooftops[71]. The MRF provides the dependencies between the various hypotheses.

To detect buildings and urban regions from the IKONOS imagery, the scale-invariant feature transform (SIFT) approach has been proposed[2]. SIFT is a suitable approach for identifying objects in different conditions; however, its key parameters are not strong enough to detect man-made structures. Therefore, the tools and solutions have been incorporated from graph theory (graph cut and graph matching). The system gave promising results on a set of 28 test images of different sites. Another graph-based method was presented for extracting buildings from HRS images[72]. The process of extraction involves two steps: (i) region growing and (ii) Hough transformation. The output of the proposed method proves its effectiveness, particularly for rectangular-shaped rooftops. Classification of RSIs has been an active practice in several applications. Schindler[73] has summarized the various classification methods used in remote sensing. He reviewed the random field and local filtering models used in the fields other than remote sensing. In particular, he proposes that ‘smoothness’ plays a vital role in improving the accuracy of the classification results. The datasets from two different sites (Graz in Austria and Zurich in Switzerland) were used for experimentation to evaluate (qualitative and quantitative) the results.

A built-up region generally contains natural as well as man-made objects[74]. These regions contain a large amount of structural information, which is useful for their detection due to their wide range of applications. A new block-based approach has been suggested for detecting built-up regions[74]. The overall method contains three key steps: (i) Multiterminal learning techniques for including the multiple features of the input data. (ii) The results of image interpretation are combined using several block sizes, which is technically known as ‘multifield integrating’. (iii) The results of multifield integrating serve as an input for pixel-level analysis termed as ‘multihypothesis voting’. The satellite images of GF-1 and ZY-3 were used for experimentation and validation. The results were also compared with the methods developed by various researchers[75–77].

Shadow detection-based methods

The shadow plays a vital role in detecting the objects from aerial and HRS images. A study used the low-level segmentation method for detecting buildings from aerial images of suburban areas[78]. Four methods were used to perform shadow analysis to determine the relation between buildings and their respective shadows[79]. These methods were also tested on high-resolution aerial images of suburban regions. Collated features have been used to perform visual tasks such as shape description, matching and active vision[80]. Also, perceptual grouping was utilized for detecting 3D structures from aerial images. The edge detection and region-growing techniques were integrated to extract buildings from aerial images[81]. Here the shadow acts as the key to improving the accuracy of the detected buildings. The study also proposed an algorithm that improved the error caused due to segmentation. The delineation and detection of man-made objects have been an important area of research in land-use analysis and cartography[82]. The study used stereo as well as monocular images for extracting the man-made objects by introducing the information fusion technique. Image geometry has been an active area providing useful information in detecting man-made objects from satellite and aerial images[83]. The four test images of Fort Hood were considered, which included different types of buildings such as peak roofs, *L*-shaped, flat and rectangular. The set of horizontal and vertical attributes was taken into account for developing the hypothesis for detecting man-made features. The output of these methodologies was represented for oblique and nadir imagery. Further, these results were evaluated and tested using the manually prepared ground truth. Later, a six-stage process of detecting buildings was introduced from single-intensity images[84]. The study used projective and geometric constraints to generate the hypothesis to detect rooftops from the images. The stages described are[84]: (a) linear feature extraction, (b) generating hypothesis, (c) selecting hypothesis, (d) verifying hypothesis, (e) 3D analysis and (f) 3D description of the scene. However, this technique detect can only rectilinear-shaped buildings having flat rooftops. Colour is also a key element that is being introduced in the field of object detection from satellite imagery. There are different colour spaces such as RGB, CMY, YIQ and YUV. Gevers and Smeulders[85] focused on intensity, hue and saturation for object recognition. The accuracy of the experimental results of the detected objects was evaluated using 500 images, which were taken from 3D man-made objects. A robust approach has been proposed which detects buildings from orthophotos by exploring Bayesian networks[86]. Figure 1 shows the important features of the designed Bayesian network for detecting

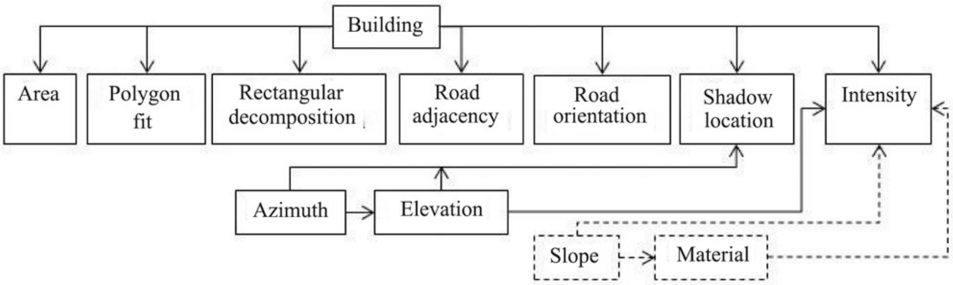

Figure 1. Elements of the Bayesian network[86].

buildings. These features are used in developing the building hypothesis. The Bayesian networks have been used in different areas, namely medical diagnosis, risk assessment and forecasting.

The well-known phenomenon of Rayleigh scattering was applied to segment the aerial coloured images[87]. This technique is simple and robust because the mathematical model represents a physical model. In August 1999, an earthquake struck Turkey, which caused large-scale destruction[88]. Therefore, the post-event aerial images were digitally analysed and processed using shadow information to determine the collapsed buildings. The overall process was performed in three stages: (a) preprocessing, (b) shadow casting and (c) detecting the collapsed buildings. The method presented by Sarabandi *et al.*[89] has a few limitations, such as the detection of buildings with complex shapes and those near vegetation area. The method proposed by Peng and Liu[90] detects buildings from monocular aerial images of urban areas in China. Objects such as roads, grasslands, parking areas and buildings are densely distributed in this region. The buildings are detected in two phases; first, extraction and verification of sunshine-parts and second, extraction of self-shadow parts. The first phase consists of region-based segmentation, candidate verification using context and radiometric parameters, using context and features for region-based refinement, and improving building contours. The second phase includes the development of a mathematical model and exploring self-shadow. The overall process was carried out without prior knowledge of illumination.

An automated de-shadowing technique has been implemented in five different colour spaces such as HSI, HCV, HSV, YC_bC_r and YIQ models to detect buildings from colour images[91]. This method is processed in four steps: colour transformation, shadow segmentation, shape preservation and shadow compensation. The study also presents a comparative evaluation of the results obtained in different colour spaces. The detection and identification of building rooftops have been an active area of research in computer vision and remote sensing. An exceptional method which combines 2D and 3D information for detecting building rooftops from RSIs has been presented[71]. Being a stochastic approach, contour-based grouping has been used for generating the hypothesis for rooftops. However, the dependencies and relationships between different hypotheses are represented using the well-known MRF model. The proposed methodology was used for detecting buildings of various colours, shapes, and heights from a set of airborne and synthetic images.

An efficient successive thresholding scheme was proposed to detect shadows from aerial images[92]. This approach improved and updated the ration map obtained from the exponential function using Tsai's algorithm. The subjective and objective evaluation of the results was done to determine the accuracy of the detected shadows. A mathematical, morphological-based method for building detection has been proposed[93]. First, watershed segmentation is used for partitioning similar regions of the IKONOS panchromatic image. Then the shadow regions are clustered using minimum spanning trees. The experimental results of the proposed method were able to detect buildings of complex shapes and different colours. A multispectral shadow detection algorithm incorporates the benefits of the near-infrared bands[94]. The proposed method has been evaluated using three IKONOS images and one GeoEye image of 1 m and 0.5 m of spatial resolution respectively. Based on visual examination, it was reported that the proposed approach has the potential to detect natural as well as artificial shadows.

Izadi and Saeedi[95] proposed a method for (i) 2D rooftop detection and (ii) 3D building estimation. They employed different image primitives (line intersection detection and line linking) in the first part for examining their relationship using a graph-based method for creating and refining the hypothesis. However, the second part of the proposed system was further divided into five steps: (i) acquisition geometry, (ii) shadow segmentation, (iii) shadow prediction, (iv) creating fuzzy rules and (v) height estimation. The potential and effectiveness of the presented system were evaluated with 20 QuickBird images. A novel automated approach for detecting buildings was introduced from VHR optical satellite images[3]. The study used two key algorithms, i.e. fuzzy landscape generation[96] and grab-cut partitioning, to extract buildings from the images of 20 different sites captured from two different satellites (QuickBird and GeoEye-1). The results were evaluated using the pixel and object-based approach. However, buildings whose shadows were not visible were missing in the generated output.

A supervised classification (enhanced parallelepiped) approach has been presented for detecting buildings from Google Earth images[97]. The performance of the proposed method was evaluated using the object and pixel-based methods. Using the advantages of supervised and unsupervised learning algorithms, a novel framework known as 'self-supervised decision fusion' was proposed[98]. This framework is divided into three components: (i) information extraction, (ii) development of an algorithm for selecting the negative and positive samples required for training, and (iii) implementation of a decision approach for classification at the base and meta layer. The results of the proposed framework were validated over 19 test sites (multispectral images of QuickBird, WorldView 2, and GeoEye-1), which were further compared with the other algorithms to prove the viability of the method.

Again the shadow information proved to be an important factor for detecting objects (buildings) from RSIs. The framework introduced by Manno-Kovacs and Ok[99] combines the knowledge of urban regions (with the help of graph cut) and shadow information for detecting buildings. The reliability and quality of the proposed method were assessed with 14 test images of IKONOS-2 and QuickBird. The obtained results were compared with those of other methods[3,96,100] to confirm the superiority of the proposed method. An unsupervised method to extract buildings along with roads from HRS image has been presented[101]. The complete process consists of three stages: (i) detecting initial building areas through local processing, (ii) detecting initial building areas through global processing, and (iii) detecting buildings and roads simultaneously. The performance of the method was tested over the 12 multispectral test images of GeoEye-1. A morphological based framework has been suggested which consists of five steps[1]: (i) morphological improvement, (ii) clustering using a multispeed-based approach, (iii) shadow detection, (iv) minimizing the false alarms and (v) segmentation. The proposed framework was evaluated using the QuickBird and IKONOS images. A comparative evaluation of the obtained results with other methods[84,102] proved the efficiency of the proposed framework.

Cognition-based methods

The objective of the cognitive model introduced by Zhang *et al.*[103] is to assess and analyse the damage caused to buildings by the RSIs. It also aims to understand the underlying cognitive process and make use of the knowledge iteratively, which is necessary for interpreting the imagery. This model is based on the principle of fuzzy theory and cognitive theory for interpreting the satellite images and extracting the information related to the damage. The fuzzy logic approach is capable of emulating human thinking and it also considers all linguistic rules. The fuzzy classification method is widely used for information extraction from images. The cognitive model uses human cognitive parameters such as perception by visually interpreting the pre- and post-earthquake images to determine the changes[103]. The cognitive model also uses reasoning as a key cognitive parameter by providing a semantic meaning to the damaged objects. There is a different process of object recognition and image understanding based on fuzzy theory and cognitive theory. Therefore, this model simulates the process of interpreting RSIs by human beings. The cognitive model for damage assessment used the HRS images of QuickBird and IKONOS, which have a resolution under 1 m. The overall method has been implemented in three steps[103]. In the first step, low-level features such as texture, colour, shape and tone are extracted using image processing and object recognition techniques. There are two methods to obtain the low-level features. First, the original image is segmented which consists of object attributes and features. Second, filters are also used to determine the feature of the particular object in an image. In the second step, semantic features (close-to, part-of, is-a, temp-rel, and con-of) are determined in a top-down manner using a knowledge base that consists of the predefined object detectors. In the third step, integration of these features is performed using a fuzzy logic approach through membership function for image understanding and object recognition. This procedure is similar to the way human beings understand images[103]. The key advantage of the model is that the process of extraction of semantic features from the satellite images is iterative and not mono-directional. Therefore, each step of the model can connect easily. As this model incorporates the cognitive process for damage assessment, therefore, the defined rules and knowledge can be reused easily. However, the cognitive model for damage assessment needs to be tested on the aerial images to determine the quality of the model in terms of accuracy. Chandra and Ghosh[8] aimed to emulate human cognitive processes by integrating cognitive task analysis (CTA) for information extraction from HRS images. A deep understanding of human cognitive capabilities is required to automate the method of information retrieval from HRS images. The authors have used theories from cognitive system engineering, which have been combined with geospatial studies. First, preliminary knowledge about the cognitive processes which human beings acquire during the interpretation of satellite images was collected. Then, this knowledge was represented in the form of rules based on the visual interpretation of the images by human beings. During knowledge elicitation, these rules were used to extract buildings from HRS images employing the mixture tuned matched filtering algorithm. Later, the method was tested using 14 HRS images of an urban area (sample result shown in Figure 2). A cognitive-based automated approach for detecting buildings

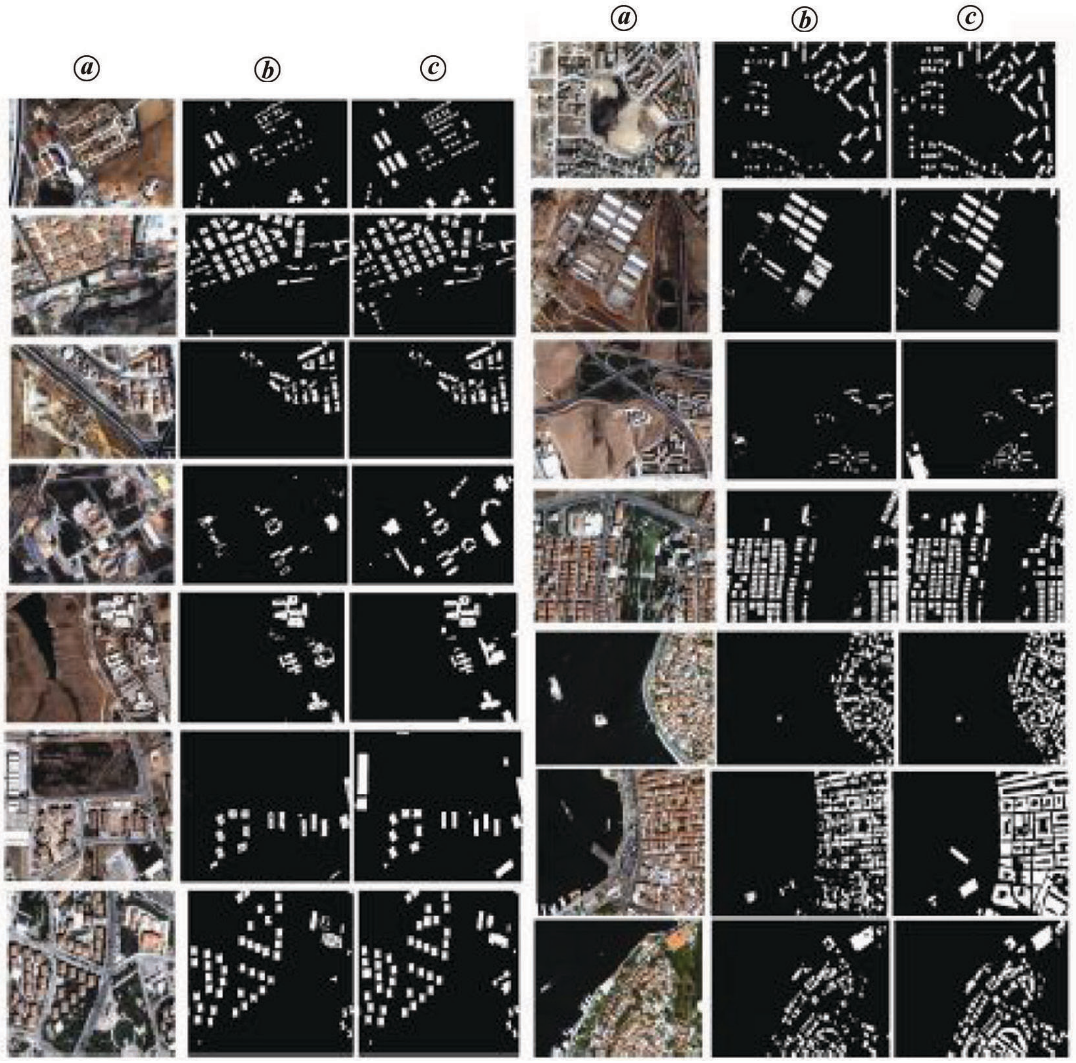

Figure 2. Results of cognitive-based methods[8]: ***a***, input images; ***b***, output images; ***c***, reference images.

from VHR multispectral images has been proposed[9]. VHR satellite imagery is a valuable source of information for the extraction of geospatial information. The cognitive processes used by human beings are emulated and incorporated with CTA to detect buildings from VHR multispectral images. CTA is carried out in five stages: (i) preliminary knowledge collection, (ii) knowledge representation, (iii) knowledge elicitation, (iv) verification and analysis of results, and (v) formatting output for various applications. The performance of the proposed method was assessed over 14 test images of an urban region.

Chandra and co-workers have presented building detection from aerial and satellite images[104,105]. A similar approach has been implemented by Chandra and co-workers[106,107] to detect road networks from HRS images using the dataset presented in Das *et al.*[108]. To analyse the satellite images cognitively, Dong *et al.*[109] utilized the eye-tracking technology to record eye movements and the response of the participants to stimuli. The process of human-based segmentation of HRS images has been studied using eye-tracking technology[110,111].

The role of cognition and visual perception in the analysis of RSIs has been examined in some studies[112,113]. The geo-visual analytical methods were implemented to monitor the forest disturbance by integrating the computational as well as human efforts. The cognition-based geographic information system has been summarized in Montello and Freundschuh[114]. The cognitive factors present in the interpretation of aerial photographs are described by Bianchetti and MacEachren[115]. The authors present details of the literature from 1922 to 1960 about the cognitive views on the interpretation of aerial photographs[115,116]. The cognitive factors required to interpret and prepare of maps have also been mentioned[117]. A cognitive approach for identifying the damage caused due to a hurricane from an RSI has been showcased[118]. A cognitive approach for detecting buildings has also been proposed[119]. This model employs the theories of hierarchical cognition, which are defined in three layers: (i) visual cognition (implementation of image segmentation), (ii) logical cognition (implementation of neural network and fuzzy logic), and (iii) psychological cognition (identification of features). The overall method is computed with the help of prior knowledge used for developing production rules[120,121]. The validation of the proposed method has been performed using PolSAR images.

Deep learning models

Deep learning, the subset of machine learning and artificial intelligence, has been a key breakthrough in the last few years due to its wide range of applications (object detection, semantic segmentation and image classification) in computer vision and remote sensing studies[122]. In recent years, convolutional neural network (CNN) has been widely used in remote sensing[123]. The most common structures of CNN include VGGNet, AlexNet, GoogLeNet and ResNet. This section discusses the deep learning-based methods proposed in the last few years for object detection (buildings) from satellite images.

A deep CNN was employed to develop an automated framework for building detection from VHR RSIs of WorldView-2 and QuickBird having eight and four spectral bands respectively[124]. The framework used the supervised classification approach for training and MRF to detect the labels. The ImageNet framework was utilized to detect buildings through trained data. A CNN model was trained to classify multispectral images in order to identify buildings in a particular patch[125]. The Landsat 8 images were used to evaluate the developed model. CNN was also used to detect buildings from HRS images[126]. The dataset was obtained from Bin Maps (8408 tile imagery of Myanmar) and the proposed model was implemented in Deep-Learn-Toolbox and GNU octave. The overall accuracy obtained was 98%, whereas the producer's and user's accuracy was estimated to be 35% and 48% respectively. An approach with reduced complexity has been presented for classifying and extracting buildings from synthetic aperture radar (SAR) imagery[127]. The modified approach employs CNN and fully-connected-feed-forward-deep-network (FDN) to detect buildings. The images were obtained from two sensors, i.e. airborne SAR (Shifang and Dujiangyan, China) and TerraSAR-X (Spain, Barcelona, Japan and Sendai). The produced accuracy was greater than 92%. The conditional random field model used boundary/edge localization to improve accuracy[128]. The SpaceNet dataset was used to validate the proposed method with greater than 92% accuracy. Further, a method which first calculates NDVI was presented[129], which later integrated the information within Res-U-Net. The proposed method was validated using the ISPRS-2D-semantic-labelling dataset of Germany (Vaihingen and Potsdam). The f1-score obtained using Postdam and Vaihingen datasets was 0.9390 and 0.9515 respectively. Another method used binary distance transformation approach to improve data labelling and U-Net model for detecting buildings from multispectral images[130]. The images of four cities (Vegas, Shanghai, Paris and Khartoum) included in the SpaceNet challenge dataset were used to evaluate the model. The maximum f1 score obtained was 0.883 for Vegas; however, the minimum was 0.584 for Khartoum. To integrate the structure-based information of objects (buildings), the Xception-module was replaced with the U-Net encoder for extracting buildings[131]. This approach is known as the multitasking learning-based method. Massachusetts (151 images) and Vaihingen (33 images) building detection datasets were employed for validation. The overall accuracy for the Massachusetts and Vaihingen datasets was 94.23% and 96.53% respectively. A multi-source-based method for building extraction using U-Net has been presented[123]. The authors developed their own dataset known as WHU building detection dataset containing 220,000 buildings in aerial images of New Zealand. The proposed model was also validated with other well-known datasets such as ISPRS, INRIA and Massachusetts Building Detection dataset. The quantitative and comparative evaluation proved the potential of the proposed method. The buildings from Sentinel 1 SAR images were extracted to explore the capability of the U-Net algorithm[132]. The proposed method is based on CNN. The authors also validated the model with multispectral images of Sentinel-2. The study area was located in the Netherlands. The overall accuracy obtained was more than 80%. The images of unmanned aerial vehicles were employed to detect buildings[133]. The faster R-CNN model was trained with 800 images and the accuracy obtained on 200 test images was 92.3%. A dense-residual-neural network (DR-Net) has been presented, which is a combination of three models, namely densely connected CNN, deep-labv3+Net decoder or encoder, and residual network[134]. The proposed model included less number of parameters (9 million) in comparison with the BRR-Net (17 million). The model revealed increased f1 scores on both the WHU (1.4%) and Massachusetts (2.9%) building detection datasets. A multitasking-based method for semantic segmentation of buildings was proposed[135]. In this method, the performance of U-Net model was improved through an encoder (one) and decoder (two). In addition, a joint-less-function (collection of mean-square-error and negative-log-likelihood) was introduced, which is capable of operating two tasks together. The method was evaluated on ISPRS-2D-semantic labeling dataset. The f1-score estimated was 94.53%, which was greater than those of other methods (DAN, MFRN and Deep-Lab-V3). Recently, Sentinel 1 and Sentinel 2 datasets have been used with Open Street Map to train the U-Net model to detect both buildings and roads[122]. The training and testing datasets included 31 and 13 zones of Spanish cities. The qualitative and quantitative outputs proved the potential of the method. Thus on the basis of the quantitative results, it has been observed that deep learning is an efficient method for detection of buildings from RSIs.

Table 1 provides a description of the building detection methods.

Dataset and evaluation metrics

The potential of the methods developed for detecting buildings from satellite images has been evaluated with the available benchmark datasets. This section describes the datasets used by several researchers for qualitative and quantitative evaluation of their proposed methods. They have used satellite and aerial images. Each dataset discussed includes the following details: (a) number of test images, (b) ground-truth images, (c) name of satellite/sensor, (d) number of bands, (e) size and (f) location. The first building detection benchmark is available on the website: http://biz.nevsehir.edu.tr/ozgunok/en/408. The size of the dataset is 33.6 MB. The building detection dataset consists of 14 images obtained from two VHR satellites, namely QuickBird and IKONOS-2, having a resolution of 0.60 and 1 m respectively. Four images were obtained from IKONOS-2 and ten from QuickBird. All the images consist of three bands (blue, green and red) with a radiometric resolution of 11 bits per band. The sun azimuth and sun zenith of 14 images of the building detection dataset ranged from 144.8494 to 157.1640 and 24.6573 to 31.8230 respectively[96]. This dataset has been employed to validate the developed cognitive methods for building detection[8,9]. It has also been used for evaluating the methods proposed by various researchers[96,99,101]. Second, the SZTAKIINRIA benchmark dataset contains the 655 rectangular footprints of the buildings in nine aerial/satellite images captured from IKONOS, QuickBird and Google Earth[136]. These images were taken from Bodensee (Germany), Manchester (United Kingdom), Normandy, Szada, Budapest, and Cot d'Azur. The methods described earlier in text have used this dataset for validation[100]. Third, the Massachusetts Buildings dataset contains 151 aerial images of the suburban and urban regions of Boston, USA[137,138]. The images are 1500 × 1500 pixels in size. The area covered by each image is 2.25 km^2 at a resolution of 1 m^2/pixel. The dataset is divided into three categories: training data (1108 images), testing data (49 images) and validation data (14 images). It has been used to validate the proposed machine learning-based approach for building detection from aerial images[137,138]. Fourth, Inria's dataset was used for pixel-wise labelling of the aerial images[139]. It contains aerial orthorectified coloured images with geographical coverage of 810 km^2 with a spatial resolution of 0.3 m. The dataset is divided into two subsets: (i) training dataset (405 km^2) and (ii) testing dataset (405 km^2). It consists of various cities such as Chicago, Austin, Kitsap County, Western Tyrol, Vienna, Innsbruck, San Francisco, Bellingham, Eastern Tyrol, and Bloomington. Each region in the training and testing datasets contains 36 tiles (5000*5000 pixels in size) covering an area of 1500*1500 m. The semantic labelling-based method has utilized this dataset for validation[139]. Fifth, the WHU dataset consists of more than 222,000 manually edited aerial as well as satellite images[140]. This dataset is divided into four categories: (1) the aerial images which include independent buildings in Christchurch, New Zealand, with a ground resolution of 0.075 m. The key features of these images are that they have been manually edited and are available on the official website of the New Zealand Land Information Services (https://data.linz.govt.nz/layer/51932-christchurch-post-earthquake-01m-urban-aerial-photos-24-february-2011/). Further, most of the dataset is down-sampled to 187,000 buildings (0.3 m of ground resolution) and cropped into 8,189 tiles with the size of 512*512 pixels. These samples are further divided into three parts: (i) training data which include 130,500 buildings having 4736 tiles; (ii) validation data with 14,500 buildings having 1036 tiles and (iii) testing data with 42,000 buildings having 2416 tiles. (2) Satellite dataset I contains 204 images of global cities (New York, Milan, Venice, Cordoba, Santiago, Cairo, Wuhan, Taiwan, Los Angeles and Ottawa) with resolutions ranging from 0.3 to 2.5 m. As the images were taken from different satellite sensors (IKONOS QuickBird, WorldView series and ZY-3) resulting in differences in radiometric and atmospheric corrections, atmospheric conditions, however the multispectral and panchromatic fusion algorithms still made the dataset challenging for evaluating the effectiveness of the proposed building detection method[140]. (3) Satellite dataset II which includes 29,085 images of six neighbouring satellites from East Asia with geographical coverage of 550 km^2 and ground resolution of 2.7 m. The training dataset contains 21,556 buildings with 13,662 tiles, whereas the testing dataset includes 7529 buildings with 3726 tiles. (4) The building change detection dataset includes 12,796 buildings that were rebuilt due to the damage caused by an earthquake (6.3 magnitude) in February 2011. The aerial images were captured in April 2012, covering an area of 20.5 km^2. In 2016, this database was updated by adding 3281 buildings. The WHU dataset has been used to validate the developed deep learning-based method for building detection from aerial and satellite imagery[140]. Sixth, the high-resolution aerial image of Graz city mainly consists of buildings[141]. The size of this dataset is 512*511 pixels. It is obtained from UltraCamD from Microsoft Photogrammetry having three colour channels, i.e. red, green and blue. UltraCamD is capable of delivering images of size 3680*7500 pixels along and across the track respectively. The high-resolution aerial images have been used for testing the performance of the hierarchical pseudo-conditional random field model for building detection from aerial imagery[141]. The Graz dataset has been used to compare the labelling methods[73]. There are several other datasets which include the multiple

Table 2. List of available building detection datasets

Building detection datasets	Source
SZTAKI-INRIA	http://mplab.sztaki.hu/remotesensing/building_benchmark.html
Massachusetts buildings dataset	https://www.cs.toronto.edu/~vmnih/data/
Inria aerial image labeling dataset	https://project.inria.fr/aerialimagelabeling/
WHU building dataset	http://gpcv.whu.edu.cn/data/building_dataset.html
ISPRS	https://www2.isprs.org/commissions/comm2/wg4/benchmark/detectionand-reconstruction/
SpaceNet 2	https://spacenet.ai/spacenet-buildings-dataset-v2/

classes along with the buildings. Details of such datasets are available in the literature[142–145]. Table 2 provides a list of a few benchmark datasets.

The methods described in the previous sections have used the dataset discussed here. These methods must be validated to prove their potential. Several methods are described in the literature to determine the accuracy of the results. Therefore, the methods which can be used for quantitative evaluation have been discussed here. In general, evaluation and assessment of the results are performed by comparing the results of a method with the manually prepared reference data, also known as ground-truth data. To evaluate the results three standard quality measures, i.e. precision, recall, and f-score given in eqs (1), (2) and (3) respectively are best suited[3,96,111]. Precision (also known as the positive predictive value) represents how many of the selected instances are relevant, whereas recall (also known as sensitivity) represents how many significant/relevant instances are being selected from a total number of instances. However, the f-score represents the harmonic mean of positive predictive value and sensitivity. Precision and recall are widely used in information retrieval, binary classification and pattern recognition to determine how well the detected objects correspond to the reference datasets. Precision determines the false positive in an algorithm; however, recall determines the objects correctly detected by the algorithm.

$$\text{Precision} = \frac{\|\text{TP}\|}{\|\text{TP}\| + \|\text{FP}\|} \tag{1}$$

$$\text{Recall} = \frac{\|\text{TP}\|}{\|\text{TP}\| + \|\text{FN}\|} \tag{2}$$

$$f\text{-score} = \frac{(2 \times \text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \tag{3}$$

During assessment, all the pixels of an image are classified into three different classes, namely true positive (TP), false positive (FP) and false negative (FN)[3]. TP indicates a pixel that is labelled as a building by the proposed method and represents the building in the ground-truth dataset. FP signifies a pixel that does not represent any of the pixels labelled as buildings in the ground-truth dataset. FN represents a pixel that is labelled as a building in the ground-truth dataset, but is not available in the proposed method. In eqs (1) and (2), ||.|| denotes the number of pixels assigned to each class and the *f*-score is the combination of precision and recall into a single score. Several researchers have used these metrics for validating their results[8,9,98–100].

Some researchers define the terminologies mentioned in eqs (1)–(3) as shape accuracy, completeness and correctness[67,132]. Shape accuracy is calculated using eq. (4), where X_1 represents the true building region and X_2 denotes the corresponding detected values. Completeness defines the ratio of the detected buildings to the total number of buildings available in the imagery. Correctness is the ratio of truly detected buildings to the total number of buildings. Correctness represents accuracy based on boundary extraction.

$$\text{Shape accuracy} = \left(1 - \frac{|X_1 - X_2|}{X}\right) \times 100 \tag{4}$$

The other evaluation terminologies used in the building detection approach are branching factor[33] (eq. (5)), miss factor[33] (eq. (6)), percentage[33] of building detection (eq. (7)), quality percentage[33] (eq. (8)), kappa coefficient[40] (*k*) (eq. (9)) and chance agreement[40] (eq. (10)).

$$\text{Branching factor} = \frac{\text{FP}}{\text{TP}} \tag{5}$$

$$\text{Miss factor} = \frac{\text{FN}}{\text{TP}} \tag{6}$$

$$\text{Percentage of building detection} = 100 \times \frac{\text{TP}}{(\text{TP + FN})} \tag{7}$$

$$\text{Quality percentage} = 100 \times \frac{\text{TP}}{(\text{TP + FP + FN})} \tag{8}$$

$$\text{Kappa coefficient } (k)$$

$$= \frac{(\text{TP + TN})\cdot(\text{TP + TN + FP + FN}) - \text{chance agreement}}{(\text{TP + TN + FP + FN})^2 - \text{chance agreement}} \tag{9}$$

$$\text{Chance agreement} = (\text{TP} + \text{FP}) \cdot (\text{TP} + \text{FN}) \cdot (\text{TN} + \text{FN}) \cdot (\text{TN} + \text{FP}) \qquad (10)$$

The receiver operating characteristics (ROC) curve analysis is a well-known pixel-based evaluation technique[42]. ROC generally represents the relationship of TP pixels against FP pixels. The area under the ROC curve is used for evaluating the quality of each change index and the produced output.

Conclusion

This study categorizes and summarizes the algorithms/ methods/techniques of building detection from RSIs in the last few years.

The low-level feature-based methods utilize the descriptors such as colour, geometrical properties, texture, contrast, shape, regularity and images to detect objects. These descriptors are the basic features which include information related to the visual properties of an imagery. These methods are generally the effects of high-level features and are easy to define and implement; however, semantically they are less meaningful. The snake model determines the boundaries of a particular shape present in an image. It is the best fit in the condition where the approximate/estimated shape of the boundaries is known. Snake models can be used to monitor dynamic objects, while the process of convergence defines accuracy. The graph-based methods generate the hypothesis and topological relationships which simplify the task of detecting objects from an image. These models may sometimes be time-consuming and highly complex. The shadow detection-based methods are categorized in feature-based taxonomy in particular, spatial (texture or geometrical) and spectral (physical or chromaticity). These methods are robust but computationally expensive. The recently proposed cognitive-based methods explore the human cognitive capabilities such as reasoning and perception to detect objects from RSIs. These models determine and monitor human brain activity involved in object detection. In future, these models will mimic the process of object detection as in humans. On the other hand, deep learning-based methods learn directly from the data (image, sound and text) to produce output. These models are flexible and can be used to find solutions to complex problems in the future. They also have a range of applications. Moreover, a high volume of dataset is required to produce better accuracy when compared to traditional methods. The training is highly expensive due to the hundreds of machines. The present study also describes details of the benchmark datasets used by different authors for detecting buildings from RSIs. It also describes the evaluation metrics used in previous studies for quantitative evaluation of the results. The developed methods are found to be robust and efficient, however, they have some limitations. Therefore, researchers must focus on the following in the future.

(i) Improvement in detecting the boundaries of buildings.
(ii) Some of the methods are unable to detect the building regions clearly due to variation in the shape, size, density and colour of the buildings.
(iii) The cognitive-based approach must be tested on the images covering a large geographical area.
(iv) Few techniques fail to extract buildings having mixed rooftops (texture or shade).
(v) The impact of haze or snow on building detection must be studied.
(vi) Improvement in partially detected buildings due to noise.
(vii) To improve the accuracy of the proposed algorithms, the quality of the training data must be improved.
(viii) The developed models perform well for an urban region where buildings are densely populated, whereas they fail to provide satisfactory results in rural areas, where the buildings are sparsely located. Therefore, a generalized model must be developed which can deliver effective results for urban as well as rural areas.

In conclusion, the development of an automated model with negligible human involvement is yet a challenging and significant task in the field of computer vision and remote sensing.

1. Chaudhuri, D., Kushwaha, N. K., Samal, A. and Agarwal, R. C., Automatic building detection from high-resolution satellite images based on morphology and internal gray variance. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sensing*, 2015, **9**(5), 1767–1779.
2. Sirmacek, B. and Unsalan, C., Urban-area and building detection using SIFT keypoints and graph theory. *IEEE Trans. Geosci. Remote Sensing*, 2009, **47**(4), 1156–1167.
3. Ok, A. O., Automated detection of buildings from single VHR multispectral images using shadow information and graph cuts. *ISPRS J. Photogramm. Remote Sensing*, 2013, **86**, 21–40.
4. Sirmacek, B. and Unsalan, C., A probabilistic framework to detect buildings in aerial and satellite images. *IEEE Trans. Geosci. Remote Sensing*, 2010, **49**(1), 211–221.
5. Konstantinidis, D., Stathaki, T., Argyriou, V. and Grammalidis, N., Building detection using enhanced HOG–LBP features and region refinement processes. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sensing*, 2016, **10**(3), 888–905.
6. Cheng, G. and Han, J., A survey on object detection in optical remote sensing images. *ISPRS J. Photogramm. Remote Sensing*, 2016, **117**, 11–28.
7. Liasis, G. and Stavrou, S., Building extraction in satellite images using active contours and colour features. *Int. J. Remote Sensing*, 2016, **37**(5), 1127–1153.
8. Chandra, N. and Ghosh, J. K., A cognitive method for building detection from high-resolution satellite images. *Curr. Sci.*, 2017, **112**(5), 1038–1044.
9. Chandra, N. and Ghosh, J. K., A cognitive viewpoint on building detection from remotely sensed multispectral images. *IETE J. Res.*, 2018, **64**(2), 165–175.
10. Gavankar, N. L. and Ghosh, S. K., Automatic building footprint extraction from high-resolution satellite image using mathematical morphology. *Eur. J. Remote Sensing*, 2018, **51**(1), 182–193.

11. Gavankar, N. L. and Ghosh, S. K., Object based building footprint detection from high resolution multispectral satellite image using *K*-means clustering algorithm and shape parameters. *Geocarto Int.*, 2019, **34**(6), 626–643.
12. Jiang, X., He, Y., Li, G., Liu, Y. and Zhang, X. P., Building damage detection via superpixel-based belief fusion of space-borne SAR and optical images. *IEEE Sens. J.*, 2019, **20**(4), 2008–2022.
13. Mayer, H., Automatic object extraction from aerial imagery – a survey focusing on buildings. *Comput. Vision Image Understand.*, 1999, **74**(2), 138–149.
14. Baltsavias, E. P., Object extraction and revision by image analysis using existing geodata and knowledge: current status and steps towards operational systems. *ISPRS J. Photogramm. Remote Sensing*, 2004, **58**(3–4), 129–151.
15. Brenner, C., Building reconstruction from images and laser scanning. *Int. J. Appl. Earth Obs. Geoinf.*, 2005, **6**(3–4), 187–198.
16. Gruen, A. and Dan, H., TOBAGO – a topology builder for the automated generation of building models. In *Automatic Extraction of Man-made Objects from Aerial and Space Images (II)*, Birkhauser, Basel, Switzerland, 1997, pp. 149–160.
17. Gruen, A. and Wang, X., CC-Modeler: a topology generator for 3-D city models. *ISPRS J. Photogramm. Remote Sensing*, 1998, **53**(5), 286–295.
18. Gulch, E., Müller, H., Labe, T. and Ragia, L., On the performance of semiautomatic building extraction. *Int. Arch. Photogramm. Remote Sensing*, 1998, **32**, 331–338.
19. Gulch, E., Muller, H. and Labe, T., Integration of automatic processes into semiautomatic building extraction. *Int. Arch. Photogramm. Remote Sensing*, 1999, **32**(3; SECT 2W5), 177–186.
20. Haala, N., *Building Reconstruction by Combining Picture and Height Data*, Bavarian Academy of Sciences, 1996, p. 12.
21. Henricsson, O. and Baltsavias, E., 3-D building reconstruction with ARUBA: a qualitative and quantitative evaluation. In *Automatic Extraction of Man-made Objects from Aerial and Space Images (II)*, Birkhauser, Basel, Switzerland, 1997, pp. 65–76.
22. Fischer, A., Kolbe, T. H., Lang, F., Cremers, A. B., Forstner, W., Plumer, L. and Steinhage, V., Extracting buildings from aerial images using hierarchical aggregation in 2D and 3D. *Comput. Vision Image Understand.*, 1998 **72**(2), 185–203.
23. Baillard, C. and Zisserman, A., Automatic reconstruction of piecewise planar models from multiple views. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149), 1999, vol. 2, pp. 559–565.
24. Suveg, I. and Vosselman, G., Reconstruction of 3D building models from aerial images and maps. *ISPRS J. Photogramm. Remote Sensing*, 2004, **58**(3–4), 202–224.
25. Brenner, C. and Haala, N., Rapid acquisition of virtual reality city models from multiple data sources. *Int. Arch. Photogramm. Remote Sensing*, 1998, **32**, 323–330.
26. Brenner, C., Towards fully automatic generation of city models. *Int. Arch. Photogramm. Remote Sensing*, 2000, **33**(B3/1; PART 3), 84–92.
27. Vosselman, G., Building reconstruction using planar faces in very high density height data. *Int. Arch. Photogramm. Remote Sensing*, 1999, **32**(3; SECT 2W5), 87–94.
28. Vosselman, G. and Dijkman, S., 3D building model reconstruction from point clouds and ground plans. *Int. Arch. Photogramm. Remote Sensing Spat. Inf. Sci.*, 2001, **34**(3/W4), 37–44.
29. Rottensteiner, F. and Briese, C., Automatic generation of building models from LIDAR data and the integration of aerial images, ISPRS, Dresden, 2003, vol. XXXIV.
30. Unsalan, C. and Boyer, K. L., A system to detect houses and residential street networks in multispectral satellite images. *Comput. Vision Image Understand.*, 2005, **98**(3), 423–461.
31. Haala, N. and Kada, M., An update on automatic 3D building reconstruction. *ISPRS J. Photogramm. Remote Sensing*, 2010, **65**(6), 570–580.
32. Shirowzhan, S., Lim, S., Trinder, J., Li, H. and Sepasgozar, S. M. E., Data mining for recognition of spatial distribution patterns of building heights using airborne lidar data. *Adv. Eng. Informat.*, 2020, **43**, 101033.
33. Shufelt, J. A., Exploiting photogrammetric methods for building extraction in aerial images. *Int. Arch. Photogramm. Remote Sensing*, 1996, **31**(B6), 74–79.
34. Zhang, Y., Optimisation of building detection in satellite images by combining multispectral classification and texture filtering. *ISPRS J. Photogramm. Remote Sensing*, 1999, **54**(1), 50–60.
35. Pesaresi, M. and Benediktsson, J. A., A new approach for the morphological segmentation of high-resolution satellite imagery. *IEEE Trans. Geosci. Remote Sensing*, 2001, **39**(2), 309–320.
36. Wang, X. and Li, P., Extraction of urban building damage using spectral, height and corner information from VHR satellite images and airborne LiDAR data. *ISPRS J. Photogramm. Remote Sensing*, 2020, **159**, 322–336.
37. Shackelford, A. K. and Davis, C. H., A combined fuzzy pixel-based and object-based approach for classification of high-resolution multispectral data over urban areas. *IEEE Trans. Geosci. Remote Sensing*, 2003, **41**(10), 2354–2363.
38. Benediktsson, J. A., Pesaresi, M. and Amason, K., Classification and feature extraction for remote sensing images from urban areas based on morphological transformations. *IEEE Trans. Geosci. Remote Sensing*, 2003, **41**(9), 1940–1949.
39. Inglada, J., Automatic recognition of man-made objects in high resolution optical remote sensing images by SVM classification of geometric image features. *ISPRS J. Photogramm. Remote Sensing*, 2007, **62**(3), 236–248.
40. Sumer, E. and Turker, M., An adaptive fuzzy-genetic algorithm approach for building detection using high-resolution satellite images. *Comput., Environ. Urban Syst.*, 2013, **39**, 48–62.
41. Senaras, C., Ozay, M. and Vural, F. T. Y., Building detection with decision fusion. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sensing*, 2013, **6**(3), 1295–1304.
42. Tian, J., Cui, S. and Reinartz, P., Building change detection based on satellite stereo imagery and digital surface models. *IEEE Trans. Geosci. Remote Sensing*, 2013, **52**(1), 406–417.
43. Zhang, Q., Huang, X. and Zhang, G., A morphological building detection framework for high-resolution optical imagery over urban areas. *IEEE Geosci. Remote Sensing Lett.*, 2016, **13**(9), 1388–1392.
44. Dempster, A. P., Laird, N. M. and Rubin, D. B., Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc.: Series B*, 1977, **39**(1), 1–22.
45. Norman, M., Mohd Shafri, H. Z., Idrees, M. O., Mansor, S. and Yusuf, B., Spatio-statistical optimization of image segmentation process for building footprint extraction using very high-resolution WorldView 3 satellite data. *Geocarto Int.*, 2019, 1–24.
46. Kass, M., Witkin, A. and Terzopoulos, D., Snakes: active contour models. *Int. J. Comput. Vision*, 1988, **1**(4), 321–331.
47. Li, C., Liu, J. and Fox, M. D., Segmentation of external force field for automatic initialization and splitting of snakes. *Pattern Recogn.*, 2005, **38**(11), 1947–1960.
48. Chan, T. F. and Vese, L., Active contours without edges. *IEEE Trans. Image Process.*, 2001, **10**(2), 266–277.
49. Chan, T. F., Sandberg, B. Y. and Vese, L. A., Active contours without edges for vector-valued images. *J. Vis. Commun. Image Represent.*, 2000, **11**(2), 130–141.
50. Lie, J., Lysaker, M. and Tai, X. C., A binary level set model and some applications to Mumford-Shah image segmentation. *IEEE Trans. Image Process.*, 2006, **15**(5), 1171–1181.
51. Brox, T. and Weickert, J., Level set based image segmentation with multiple regions. In Joint Pattern Recognition Symposium, Springer, Berlin, Germany, 2004, pp. 415–423.

52. Chen, L., Zhou, Y., Wang, Y. and Yang, J., GACV: geodesic-aided C–V method. *Pattern Recogn.*, 2006, **39**(7), 1391–1395.
53. Caselles, V., Kimmel, R. and Sapiro, G., Geodesic active contours. In Proceedings of IEEE International Conference on Computer Vision, Cambridge, MA, USA, 1995, pp. 694–699.
54. Caselles, V., Kimmel, R. and Sapiro, G., Geodesic active contours. *Int. J. Comput. Vision*, 1997, **22**(1), 61–79.
55. Yezzi, A., Kichenassamy, S., Kumar, A., Olver, P. and Tannenbaum, A., A geometric snake model for segmentation of medical imagery. *IEEE Trans. Med. Imaging*, 1997, **16**(2), 199–209.
56. Siddiqi, K., Lauziere, Y. B., Tannenbaum, A. and Zucker, S. W., Area and length minimizing flows for shape segmentation. *IEEE Trans. Image Process.*, 1998, **7**(3), 433–443.
57. Vasilevskiy, A. and Siddiqi, K., Flux maximizing geometric flows. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002, **24**(12), 1565–1578.
58. Pi, L., Fan, J. and Shen, C., Color image segmentation for objects of interest with modified geodesic active contour method. *J. Math. Imaging Vision*, 2007, **27**(1), 51–57.
59. Zheng, Y., Li, G., Sun, X. and Zhou, X., A geometric active contour model without re-initialization for color images. *Image Vision Comput.*, 2009, **27**(9), 1411–1417.
60. Hou, Z. and Han, C., Force field analysis snake: an improved parametric active contour model. *Pattern Recogn. Lett.*, 2005, **26**(5), 513–526.
61. Yan, P. and Kassim, A. A., Segmentation of volumetric MRA images by using capillary active contour. *Med. Image Anal.*, 2006, **10**(3), 317–329.
62. Guo, T. and Yasuoka, Y., Snake-based approach for building extraction from high-resolution satellite images and height data in urban areas. In Proceedings of the 23rd Asian Conference on Remote Sensing, Vienna, Austria, 2002, pp. 25–29.
63. Mayunga, S. D., Zhang, Y. and Coleman, D. J., Semi-automatic building extraction utilizing Quickbird imagery. In Proceedings of the ISPRS Workshop CMRT, 2005, vol. 13, pp. 1–136.
64. Peng, J., Zhang, D. and Liu, Y., An improved snake model for building detection from urban aerial images. *Pattern Recogn. Lett.*, 2005, **26**(5), 587–595.
65. Cao, G. and Yang, X., Man-made object detection in aerial images using multistage level set evolution. *Int. J. Remote Sensing*, 2007, **28**(8), 1747–1757.
66. Karantzalos, K. and Paragios, N., Recognition-driven two-dimensional competing priors toward automatic and accurate building detection. *IEEE Trans. Geosci. Remote Sensing*, 2009, **47**(1), 133–144.
67. Ahmadi, S., Zoej, M. V., Ebadi, H., Moghaddam, H. A. and Mohammadzadeh, A., Automatic urban building boundary extraction from high resolution aerial images using an innovative model of active contours. *Int. J. Appl. Earth Obs. Geoinf.*, 2010, **12**(3), 150–157.
68. Krishnamachari, S. and Chellappa, R., Delineating buildings by grouping lines with MRFs. *IEEE Trans. Image Process.*, 1996, **5**(1), 164–168.
69. Kim, T. and Muller, J. P., Development of a graph-based approach for building detection. *Image Vision Comput.*, 1999, **17**(1), 3–14.
70. Croitoru, A. and Doytsher, Y., Monocular right-angle building hypothesis generation in regularized urban areas by pose clustering. *Photogramm. Eng. Remote Sensing*, 2003, **69**(2), 151–169.
71. Katartzis, A. and Sahli, H., A stochastic framework for the identification of building rooftops using a single remote sensing image. *IEEE Trans. Geosci. Remote Sensing*, 2007, **46**(1), 259–271.
72. Cui, S., Yan, Q., Reinartz, P. and Mansour, N., Graph search and its application in building extraction from high resolution remote sensing imagery. *Search Algorithms and Applications*, 2011.
73. Schindler, K., An overview and comparison of smooth labeling methods for land-cover classification. *IEEE Trans. Geosci. Remote Sensing*, 2012, **50**(11), 4534–4545.
74. Li, Y., Tan, Y., Li, Y., Qi, S. and Tian, J., Built-up area detection from satellite images using multikernel learning, multifield integrating, and multihypothesis voting. *IEEE Geosci. Remote Sensing Lett.*, 2015, **12**(6), 1190–1194.
75. Smits, P. C. and Annoni, A., Updating land-cover maps by using texture information from very high-resolution space-borne imagery. *IEEE Trans. Geosci. Remote Sensing*, 1999, **37**(3), 1244–1254.
76. Weizman, L. and Goldberger, J., Urban-area segmentation using visual words. *IEEE Geosci. Remote Sensing Lett.*, 2009, **6**(3), 388–392.
77. Tao, C., Tan, Y., Yu, J. G. and Tian, J., Urban area detection using multiple kernel learning and graph cut. In IEEE International Geoscience and Remote Sensing Symposium, 2012, pp. 83–86.
78. Huertas, A. and Nevatia, R., Detecting buildings in aerial images. *Comput. Vision, Graph. Image Process.*, 1988, **41**(2), 131–152.
79. Irvin, R. B. and McKeown, D. M., Methods for exploiting the relationship between buildings and their shadows in aerial imagery. *IEEE Trans. Syst. Man, Cybern.*, 1989, **19**(6), 1564–1575.
80. Mohan, R. and Nevatia, R., Using perceptual organization to extract 3D structures. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1989, **11**(11), 1121–1139.
81. Liow, Y. T. and Pavlidis, T., Use of shadows for extracting buildings in aerial images. In *Structural Pattern Analysis*, 1990, pp. 165–180.
82. Shufelt, J. and McKeown, D. M., Fusion of monocular cues to detect man-made structures in aerial imagery. *CVGIP: Image Understand.*, 1993, **57**(3), 307–330.
83. McGlone, J. C. and Shufelt, J. A., Projective and object space geometry for monocular building extraction (No. CMU-CS-94-118). Carnegie-Mellon University of Pittsburgh PA, Department of Computer Science, USA, 1994.
84. Lin, C. and Nevatia, R., Building detection and description from a single intensity image. *Comput. Vision Image Understand.*, 1998, **72**(2), 101–121.
85. Gevers, T. and Smeulders, A. W., Color-based object recognition. *Pattern Recogn.*, 1999, **32**(3), 453–464.
86. Stassopoulou, A. and Caelli, T., Building detection using Bayesian networks. *Int. J. Pattern Recogn. Artif. Intell.*, 2000, **14**(6), 715–733.
87. Polidorio, A. M., Flores, F. C., Imai, N. N., Tommaselli, A. M. and Franco, C., Automatic shadow segmentation in aerial color images. In 16th IEEE Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI 2003), 2003, pp. 270–277.
88. Turker, M. U. S. T. A. F. A. and San, B. T., Detection of collapsed buildings caused by the 1999 Izmit, Turkey earthquake through digital analysis of post-event aerial photographs. *Int. J. Remote Sensing*, 2004, **25**(21), 4701–4714.
89. Sarabandi, P., Yamazaki, F., Matsuoka, M. and Kiremidjian, A., Shadow detection and radiometric restoration in satellite high resolution images. In IEEE International Geoscience and Remote Sensing Symposium, 2004, vol. 6, pp. 3744–3747.
90. Peng, J. and Liu, Y. C., Model and context-driven building extraction in dense urban aerial images. *Int. J. Remote Sensing*, 2005, **26**(7), 1289–1307.
91. Tsai, V. J., A comparative study on shadow compensation of color aerial images in invariant color models. *IEEE Trans. Geosci. Remote Sensing*, 2006, **44**(6), 1661–1671.
92. Chung, K. L., Lin, Y. R. and Huang, Y. H., Efficient shadow detection of color aerial images based on successive thresholding scheme. *IEEE Trans. Geosci. Remote Sensing*, 2008, **47**(2), 671–682.
93. Akçay, H. G. and Aksoy, S., Building detection using directional spatial constraints. In IEEE International Geoscience and Remote Sensing Symposium, 2010, pp. 1932–1935.
94. Teke, M., Başeski, E., Ok, A. Ö., Yuksel, B. and Şenaras, Ç., Multispectral false color shadow detection. In ISPRS Conference

on Photogrammetric Image Analysis, Springer, Berlin, Germany, 2011, pp. 109–119.

95. Izadi, M. and Saeedi, P., Three-dimensional polygonal building model estimation from single satellite images. *IEEE Trans. Geosci. Remote Sensing*, 2011, **50**(6), 2254–2272.
96. Ok, A. O., Senaras, C. and Yuksel, B., Automated detection of arbitrarily shaped buildings in complex environments from monocular VHR optical satellite imagery. *IEEE Trans. Geosci. Remote Sensing*, 2012, **51**(3), 1701–1717.
97. Ghaffarian, S., Automatic building detection based on supervised classification using high resolution Google Earth images. *Int. Arch. Photogramm., Remote Sensing Spat. Inf. Sci.*, 2014, **40**(3), 101.
98. Senaras, C. and Vural, F. T. Y., A self-supervised decision fusion framework for building detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sensing*, 2015, **9**(5), 1780–1791.
99. Manno-Kovacs, A. and Ok, A. O., Building detection from monocular VHR images by integrated urban area knowledge. *IEEE Geosci. Remote Sensing Lett.*, 2015, **12**(10), 2140–2144.
100. Benedek, C., Descombes, X. and Zerubia, J., Building detection in a single remotely sensed image with a point process of rectangles. In 20th IEEE International Conference on Pattern Recognition, 2010, pp. 1417–1420.
101. Ok, A. O., Automated detection of buildings and roads in urban areas from VHR satellite images. *J. Geodesy Geoinf.*, 2016, **3**(1), 29–38.
102. Liu, W. and Prinet, V., Building detection from high-resolution satellite image using probability model. In Proceedings IEEE International Geoscience and Remote Sensing Symposium (IGARSS'05), 2005, vol. 6, pp. 3888–3891.
103. Zhang, C., Wang, T., Liu, X. and Zhang, S., Cognitive model based method for earthquake damage assessment from high-resolution satellite images: a study following the WenChuan earthquake. In Sixth IEEE International Conference on Natural Computation, 2010, vol. 4, pp. 2079–2083.
104. Chandra, N., Sharma, A. and Ghosh, J. K., A cognitive method for object detection from aerial image. In IEEE International Conference on Computing, Communication and Automation (ICCCA), 2016, pp. 327–330.
105. Chandra, N., Ghosh, J. K. and Sharma, A., A cognitive based approach for building detection from high resolution satellite images. In IEEE International Conference on Advances in Computing, Communication and Automation (ICACCA) (Spring), 2016, pp. 1–5.
106. Chandra, N., Ghosh, J. K. and Sharma, A., A cognitive framework for road detection from high-resolution satellite images. *Geocarto Int.*, 2019, **34**(8), 909–924.
107. Chandra, N. and Ghosh, J. K., A cognitive perspective on road network extraction from high resolution satellite images. In IEEE Second International Conference on Next Generation Computing Technologies (NGCT), 2016, pp. 772–776.
108. Das, S., Mirnalinee, T. T. and Varghese, K., Use of salient features for the design of a multistage framework to extract roads from high-resolution multispectral satellite images. *IEEE Trans. Geosci. Remote sensing*, 2011, **49**(10), 3906–3931.
109. Dong, W., Liao, H., Roth, R. E. and Wang, S., Eye tracking to explore the potential of enhanced imagery basemaps in web mapping. *Cartograph. J.*, 2014, **51**(4), 313–329.
110. Sharma, A., Kumar Ghosh, J. and Kolay, S., Fixation data analysis for complex high-resolution satellite images. *Geocarto Int.*, 2019, 1–22.
111. Sharma, A., Ghosh, J. and Kolay, S., Reference data preparation for complex satellite image segmentation. *IET Image Processing*, 2019.
112. Bianchetti, R. A., Considering visual perception and cognition in the analysis of remotely sensed images. In Conference on Spatial Information Theory, 2013.
113. White, R. A., Coltekin, A. and Hoffman, R. R. (eds), *Remote Sensing and Cognition: Human Factors in Image Interpretation*, CRC Press, 2018.
114. Montello, D. R. and Freundschuh, S., Cognition of geographic information. *A Research Agenda for Geographic Information Science*, 2005, 61–91.
115. Bianchetti, R. A. and MacEachren, A. M., Cognitive themes emerging from air photo interpretation texts published to 1960. *ISPRS Int. J. Geo-Inf.*, 2015, **4**(2), 551–571.
116. Bianchetti, R. A., Looking back to inform the future: the role of cognition in forest disturbance characterization from remote sensing imagery, Doctoral Thesis, The Pennsylvania State University, 2014.
117. Barkowsky, T. and Freksa, C., Cognitive requirements on making and interpreting maps. In International Conference on Spatial Information Theory, Springer, Berlin, Germany, 1997, pp. 347–361.
118. Battersby, S. E., Hodgson, M. E. and Wang, J., Spatial resolution imagery requirements for identifying structure damage in a hurricane disaster. *Photogramm. Eng. Remote Sensing*, 2012, **78**(6), 625–635.
119. Kang, L., Zou, B., Zhang, Y. and Zhang, L., Building detection based on human visual cognition mechanism using PolSAR images. In IEEE Geoscience and Remote Sensing Symposium, 2014, pp. 2742–2745.
120. Zou, B., Zhang, Y., Wang, C. and Cheng, Y., Building cognition method based on human images cognition mechanism in high resolution PolSAR images. In IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2015, pp. 3223–3226.
121. Zou, B., Xu, X., Zhang, L. and Song, C., High-resolution PolSAR image interpretation based on human image cognition mechanism. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sensing*, 2018, **11**(11), 4259–4269.
122. Ayala, C., Sesma, R., Aranda, C. and Galar, M., A deep learning approach to an enhanced building footprint and road detection in high-resolution satellite imagery. *Remote Sensing*, 2021, **13**(16), 3135.
123. Ji, S., Wei, S. and Lu, M., Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. *IEEE Trans. Geosci. Remote Sensing*, 2018, **57**(1), 574–586.
124. Vakalopoulou, M., Karantzalos, K., Komodakis, N. and Paragios, N., Building detection in very high resolution multispectral data with deep learning features. In IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2015, pp. 1873–1876.
125. Ishii, T., Simo-Serra, E., Iizuka, S., Mochizuki, Y., Sugimoto, A., Ishikawa, H. and Nakamura, R., Detection by classification of buildings in multispectral satellite imagery. In 23rd International Conference on Pattern Recognition (ICPR), 2016, pp. 3344–3349.
126. Miyazaki, H., Kuwata, K., Ohira, W., Guo, Z., Shao, X., Xu, Y. and Shibasaki, R., Development of an automated system for building detection from high-resolution satellite images. In IEEE Fourth International Workshop on Earth Observation and Remote Sensing Applications (EORSA), 2016, pp. 245–249.
127. Xu, Z., Wang, R., Zhang, H., Li, N. and Zhang, L., Building extraction from high-resolution SAR imagery based on deep neural networks. *Remote Sensing Lett.*, 2017, **8**(9), 888–896.
128. Vakalopoulou, M., Bus, N., Karantzalos, K. and Paragios, N., Integrating edge/boundary priors with classification scores for building detection in very high resolution data. In IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2017, pp. 3309–3312.
129. Xu, Y., Wu, L., Xie, Z. and Chen, Z., Building extraction in very high resolution remote sensing imagery using deep learning and guided filters. *Remote Sensing*, 2018, **10**(1), 144.
130. Prathap, G. and Afanasyev, I., Deep learning approach for building detection in satellite multispectral imagery. In IEEE International Conference on Intelligent Systems (IS), 2018, pp. 461–465.

131. Hui, J., Du, M., Ye, X., Qin, Q. and Sui, J., Effective building extraction from high-resolution remote sensing images with multitask driven deep neural network. *IEEE Geosci. Remote Sensing Lett.*, 2018, **16**(5), 786–790.

132. Emek, R. A. and Demir, N., Building detection from SAR images using U-Net deep learning method. *Int. Arch. Photogramm., Remote Sensing Spat. Inf. Sci.*, 2020, **44**, 215–218.

133. Zheng, L., Ai, P. and Wu, Y., Building recognition of UAV remote sensing images by deep learning. In IGARSS 2020 IEEE International Geoscience and Remote Sensing Symposium, 2020, pp. 1185–1188.

134. Chen, M. *et al.*, DRNet: an improved network for building extraction from high resolution remote sensing image. *Remote Sensing*, 2021, **13**(2), 294.

135. Moghalles, K., Li, H. C., Al-Huda, Z. and Hezzam, E. A., Multitask deep network for semantic segmentation of building in very high resolution imagery. In IEEE International Conference of Technology, Science and Administration (ICTSA), 2021, pp. 1–6.

136. Benedek, C., Descombes, X. and Zerubia, J., Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2011, **34**(1), 33–50.

137. Mnih, V., Machine Learning for Aerial Image Labeling, Doctor of Philosophes, Graduate Department of Computer Science, University of Toronto, Canada, 2013.

138. Khalel, A. and El-Saban, M., Automatic pixelwise object labeling for aerial imagery using stacked u-nets, 2018, preprint arXiv: 1803.04953.

139. Maggiori, E., Tarabalka, Y., Charpiat, G. and Alliez, P., Can semantic labeling methods generalize to any city? The Inria Aerial Image Labeling benchmark. In IEEE International Geoscience and Remote Sensing Symposium, 2017, pp. 3226–3229.

140. Ji, S., Wei, S. and Lu, M., Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. *IEEE Trans. Geosci. Remote Sensing*, *Austria*, 2018, **57**(1), 574–586.

141. Thuy, N. T., Object detection from aerial image. Doctoral dissertation, Graz University of Technology, 2009.

142. Yang, Y. and Newsam, S., Bag-of-visual-words and spatial extensions for land-use classification. In Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2010, pp. 270–279.

143. Zou, Q., Ni, L., Zhang, T. and Wang, Q., Deep learning based feature selection for remote sensing scene classification. *IEEE Geosci. Remote Sensing Lett.*, 2015, **12**(11), 2321–2325.

144. Cheng, G., Han, J. and Lu, X., Remote sensing image scene classification: Benchmark and state of the art. *Proc. IEEE*, 2017, **105**(10), 1865–1883.

145. Demir, I. *et al.*, A challenge to parse the earth through satellite images. 2018, preprint arXiv:1805.06561.

Received 20 September 2021; revised accepted 27 March 2022

doi: 10.18520/cs/v122/i11/1252-1267