A Review of Building Detection Methods from Remotely Sensed Images

Naveen Chandra, naveenchandra0408@gmail.com, Assistant Professor, Uttarakhand Technical University, India.

Himadri Vaidya, Himadrivaidya8@gmail.com, Assistant Professor, Uttarakhand Technical University, India.

Corresponding author and address
Dr. Naveen Chandra
Assistant Professor,
Uttarakhand Technical University,
E-mail: naveenchandra0408@gmail.com, +91-9760030791

Abstract

With the availability of high-resolution satellite (HRS) imagery, the new applications have developed for urban regions. Nowadays from novice to expert users, everybody is using HRS images for solving geospatial issues. Building detection from remote sensing images has been an active area of research due to its broad range of applications including city modeling, map updating, and urban monitoring. To process the image manually is a consuming and laborious task therefore the number of scientist and researchers have developed the methods that demand very less or no human efforts. In current years, building detection has made significant growth through various automated and semi-automated methods/algorithms/techniques suggested in various studies. Hence the number of articles and research publications has been communicated which concentrates on detecting buildings from HRS images. Therefore the objective of this article is to provide the review and efforts of such studies. In this paper, the building detection methods are categorized into five parts: (i) low-level feature-based methods, (ii) snake models, (iii) Graph-based methods, (iv) shadow detection based methods and (v) cognition based methods. We aim that this publication will prove to be constructive for the scientists/researchers/academicians working in this research domain for better and strong understanding.

Keywords: Building detection, classification, cognitive, morphological, remote sensing images.

1 Introduction

Remote sensing technology produces 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 the man-made and natural objects from the images acquired from the satellites (Landsat) due to their resolution. However, with the emergence of very high-resolution (VHR) satellite images (Quick bird and IKONOS); the difference between the objects present on the surface of the earth can be observed [2]. Remote sensing images (RSIs) are a key and valuable source of information to be utilized for object detection. It is found that more than 50% of the world population lives in the suburban and urban region, thus the accurate and reliable detection of buildings from RSIs is a prime task [3] for various applications such as urban mapping, military intelligence, map making, change monitoring, damage detection, estimation of population and land use/land cover analysis [4-12]. Due to several reasons, human experts are unable to label the buildings in RSIs [4], [9]. First, due to their complex geometrical properties (shape and size) second, buildings may be surrounded by other objects like trees [4], [9]. Third, remote sensing data capture the large geographical region hence labeling each building in the image is a tedious and time-consuming process [4], [9]. Fourth, the key interpretation elements (contrast, resolution, and illumination) may not prove to be enough for determining buildings from RSIs [4], [9]. This article provides a summary of several building detection techniques from RSIs of the last thirty years. The paper is organized as follows: Section 2 describes the building detection methods, section 3 shows the data set and evaluation metrics, section 4 contains the comparison and characteristics of building detection methods and section 5 contains the conclusion.

2 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. Object based approach creates the segments and characterize them through features (shape, spectral, and height). While the threshold based approach generates normalized difference vegetation index (NDVI) and digital surface model (DSM) to detect buildings.

The structure of buildings in 2-dimensions (2D) and 3-dimensions (3D) has a tremendous impact globally. Hence the methods/tools developed for extraction and detection of both are different. In the past, many methods/algorithms has been introduced for 2D building extraction however few review articles have been published describing the limitations and the capabilities. In [13] a review of the building detection methods developed until mid-1990 has been provided. The review consisted of the deep summation of the strategies and the models of the developed methods. The details of the developed methods are given in table 1 however; characteristics of building detection methods including the type, resolution, and complexity of data are given in table 2. A survey concentrating on the type of knowledge being used for object detection from satellite images is given in [14]. It also focused on the issues encountered while using and upgrading the existing knowledge. It also provided 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 concentrates on 2D level because the access to 3D data set is limited, and expensive. In [15] a review emphasizing on the light detection and ranging (LIDAR) based reconstruction approaches along with their achievements have been presented. It contains a detailed review of the semiautomatic [16-19] and automatic [20-29] reconstruction methods and their properties. Further, in [30] a review presented in [13] is extended by providing a comparative evaluation of the proposed methods until 2003. Recently in [31], a review investigated the methods developed for building reconstruction using LIDAR and airborne elevation data has been provided. The author states that the reconstruction of buildings is based on three modules [31]: (i) parametric shapes, (ii) segmentation and (iii) DSM simplification. The LIDAR data has been used in [32] for determining the height of the building through 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) however it is hard to classify them due to their different applications. However the following section attempts to categorize and summarize the developed the building detection methods.

2.1 Low level features based methods

In [33] a method to produce 3D hypotheses has been presented 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 the qualitative analysis of the results. The method presented in [34] tries to extract the 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 are 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 in [35]. The author performs the closing and opening operation using the reconstruction technique. The well-known watershed segmentation has been used by the author. 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, the combination of object and pixel-based approach has been described in [37]. Within the pixel-based classification, the author uses maximum likelihood classifier and hierarchal fuzzy classifier which use spectral and spatial information. The IKONOS imagery was classified into seven different 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 multiresolution segmentation which includes the information related to spectral and spatial heterogeneity.

Similar to the approach presented in [35] the author investigates the mathematical morphological operations for feature extraction and classification of HRS images [38]. However, the author also explores the areas of neural networks for the classification of high-resolution IKONOS and IRS-1C images. The author applies the two approaches namely DAFE (discriminant analysis feature extraction) and DBFE (decision boundary feature extraction) 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.

Further, a system to detect houses and streets from multispectral satellite images is introduced in [30]. The system 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 the segmented image by binary balloon algorithm and (iv) implementation of the graph-based theoretical algorithm for detecting houses and street from IKONOS image. The developed system is very valuable for automated map generation. In [39] the high level and low-level geometry feature of the input images have been used for detecting the man-made objects from satellite images. Further, these features are classified using the supervised learning approaches i.e. support vector machine. The proposed method is tested on SPOT5 THR images having ten classes. Genetic algorithms are generally used in search problems and have turned into a standard optimization technique with its application in different fields [40]. An adaptive fuzzy based genetic algorithm has been proposed in [40]. The author determines 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. Later 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 the buildings from the Quick-Bird satellite image a novel method has been proposed in [41]. This method is based on the theories of decision fusion in which after the segmentation (using a mean-shift algorithm) the various features of the image such as shape, color and texture are classified within the newly proposed framework known as Fuzzy Stacked Generalization. Further, the potential of the proposed method identified through the comparison of the results with different machine learning algorithms. A geometrical feature plays an important role in object detection. Therefore a method proposed in [42] uses height information for determining the building change detection through stereo images. The author generates the digital surface models of the stereo images of two test regions (an urban region in Germany and industrial region in Korea) and improves the accuracy through the fusion theories of Dempster–Shafer.

As proposed in [35] and [38] another morphologically based framework is provided in [43]. The structure of the buildings is determined using the morphological building index (MBi) method. Later, during post-processing, the morphological spatial pattern is utilized to improve the results obtained from MBI. The images of WorldView-2 and GeoEye-1 are used to execute 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 suggested in [5]. In the first module, the proposed model concatenates the LBP and HOG features introducing a distance function that is trained using well known supervised learning algorithm i.e., support vector machines for calculating the distance between LBP and HOG descriptors. Further, the EM algorithm [44], is employed within the second module termed as "region refinement" for detecting the rectangular-shaped regions representing buildings.

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

2.2 Snake-models

The snake model, also known as active-contour models, was initially developed by [46]. The model includes the dynamics curves that are present in the image for capturing the image 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 categories: geometrical and parametric snakes. Geometrical snakes are referred to as zero-level sets in which updating is performed on surface function in the image domain (47). Geometrical snakes are further divided into two parts: region-based and edge-based active contours. Region-based active contours are based 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 the parametric snake is that it is unable to change the topologies during the evolution of the snake and the contour must be near to the desired boundary of the object [47], [60-61]. The snake model has its applications in image segmentation, contour location, edge detection, and visual tracking [47].

In [62] boundaries of the buildings from LIDAR data are estimated and further, the accurate position is determined using the snake model. In [63] buildings were detected from Quick bird images using a semi-automated algorithm. Initially, a point within the boundary of buildings is selected and then accurate boundaries of the buildings are extracted by reproducing the curve through an iterative approach. In [64] a traditional snake model is modified based on the geometric and radiometric characteristics of the buildings in aerial images using two parameters i.e., selecting the initial seeds along with external energy function. This method is capable of assessing the shape of buildings however it is unable to extract the buildings present in the urban region. The shape accuracy achieved for detected buildings is 83.60%. An improved Chan–Vese model for extracting man-made objects from seven aerial imagery is presented in [65]. The method is implemented using fractal error metrics and three staged segmentation algorithm. Man-made objects are detected by changing the active contours. Later, 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, in [67] a new level set method is developed for calculating the energy function to detect buildings from the high-resolution aerial image of Lavasan (central Iran). This snake model is capable of detecting the boundaries of the buildings avoiding the edges of the other objects present in the image. Unlike another model, this approach does require additional information (height) for detecting buildings. The completeness and correctness of extracted buildings are reported as 80% and 96% respectively.

A novel approach for extracting buildings has been proposed in [7] which utilizes an active contour model along with the color feature. The development of the proposed model is carried out in three stages: (i) initialization of active contours, (ii) representation of HSV and RGB color spaces, and (iii) optimization of the proposed model. The proposed model is assessed on ninety-six google earth images of different countries. Besides, the proposed model showcases better results in comparison to other active contour-based models.

2.3 Graph-Based methods

The study presented in [68] used Markov Random Field (MRFs) for grouping the line segments to delineate the buildings of particular shapes (rectangular). Later, active contours were utilized for improving the shapes of the segments. The method was tested on aerial images including the qualitative assessment of the results. In [69] a robust method is proposed for building detection 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 is obtained by considering the geometrical and mathematical relations during the generation of the hypothesis. The method is validated on the aerial photograph. However, it is found that the method applies only to the buildings of specific shapes. A Right-Angle Graph (RAG) method is proposed in [70] for detecting right-angled shaped buildings. The overall method is based on pose clustering which is a voting process (the combination of voting elements and voting rules). During the hypothesis generation, right angle edges and Hough space of the buildings are incorporated. The model is validated using real aerial images and synthetic images. The building detection percentage is obtained as more than 80%. It is observed that this method can only be used for the regular buildings of an urban region. In [30] the graph-based method was also used to detect steer and houses of North America. Further, the combination of 2D and 3D information from the airborne and synthetic images was used to detect the building rooftop [71]. The Markov random field stated the dependencies between the various hypotheses.

To detect the buildings and urban regions from the IKONOS imagery [2] the author proposed the scale-invariant feature transform (SIFT) approach. The author states that SIFT is a strong approach for identifying objects if 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 produced promising results on the set of twenty-eight test images of different sites. Another graph-based method was presented in [72] for extracting building from HRS images. 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 the remote sensing images has been an active practice for several applications. The author in [73] describes the summary of the various classification methods used in remote sensing. The author reviews the random field and local filtering models that are used in the fields other than remote sensing. In particular, the author proposes that the "smoothness" plays a vital role in improving the accuracy of the classification results. The data set from two different sites (Graz-city of Austria and Zurich-city of Switzerland) was used for experimentation to evaluate (qualitative and quantitative) the results.

A built-up region generally contains the natural as well as man-made objects [74]. These regions contain a rich amount of structural information which is very useful for the detection of the built-up region due to its wide range of applications. A new block-based approach is suggested in [74] for detecting built-up regions. The overall method contains three keys 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 size which is technically known as "multifield integrating", and (iii) The results of multifield integrating serves as an input for pixel-level analysis termed as "Multihypothesis voting". The satellite images of GF-1 and ZY-3 and are used for experimentation and validation. Further, the results were also compared with the method proposed in [75]- [77].

2.4 Shadow Detection based method

The shadow plays a vital role in detecting the objects from aerial and HRS images. The study presented in [78] used the low-level segmentation method for detecting buildings from aerial images of suburban areas. In [79] four methods were used to perform the shadow analysis for determining the relation between the buildings and their respective shadows. These methods were also tested on high-resolution aerial images of the suburban regions. In [80] collated features were used to perform some visual tasks such as shape description, matching, and active vision. Also, perceptual grouping was utilized for detecting 3D structures from aerial images. The edge detection and region growing techniques were integrated with [81] to extract buildings from aerial images. Here the shadow act as an important key in improving the accuracy of detected buildings. The author also proposed an algorithm that improved the error caused due to segmentation. The delineation and detection of the objects created by human beings have been an important area of research in land use analysis and cartography [82]. The author used the stereo as well as the monocular images for extracting the mam made objects by introducing the information fusion technique. Image geometry has been an active area which provides useful information in detecting the man-made objects from satellite and aerial images [83]. The four test images of Fort Hood were taken which included the different types of buildings such as peak roof, L-shaped, flat, and rectangular. The set of horizontal and vertical attributes were taken into account for developing the hypothesis for detecting the 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 staged process of detecting buildings was introduced in [84] from single intensity images. The author used the projective and geometric constraints to generate the hypothesis to detect rooftops from the image. The stages described in [84] are: (a) linear feature extraction, (b) generating hypothesis (c) selecting hypothesis (d) verifying hypothesis (e) three-dimensional analysis, and (f) three-dimensional description of the scene. However, it was stated that this technique is only capable of detecting rectilinear shaped buildings having flat rooftops. Color is also the key element that is being introduced in the field of object detection from satellite imagery. There are different color spaces such as RGB, CMY, YIQ, and YUV. The author in [85] concentrates on intensity, hue, and saturation for object recognition. The accuracy of the experimental results of the detected objects was evaluated using the 500 images which were taken from three-dimensional man-made objects. In [86] a robust approach has been proposed which detects buildings from orthophotos by exploring Bayesian networks. The important features of the designed Bayesian network for detecting building by the author is shown in figure 1. 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.

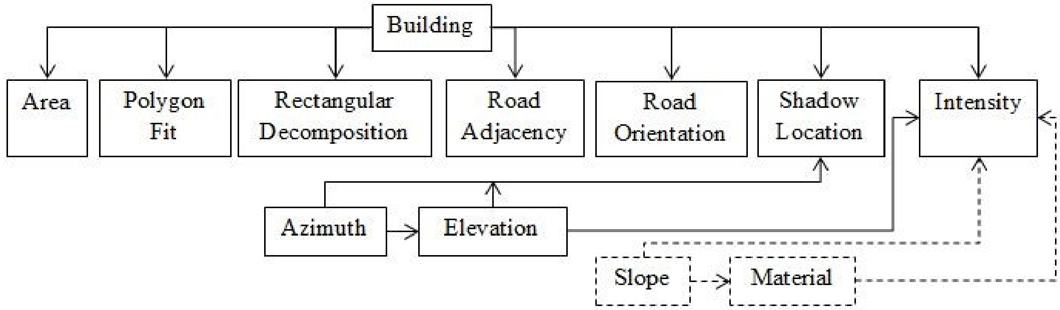

Figure 1: Elements of Bayesian networks [86]

Further, the well-known phenomenon “Rayleigh scattering effect” was applied in [87] to segment the aerial colored images. This technique was very simple and robust because the mathematical model represents a physical model. In August 1999, an earthquake struck in turkey which caused tremendous damage [88]. Therefore, the post-event aerial images were digitally analyzed 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 presented approach [89] has few limitations such as the detection of buildings with complex shapes and which were near to the vegetation area. The approach presented in [90] detects the buildings from monocular aerial images of urban areas of China. The objects such as roads, grassland, parking areas, and buildings are densely distributed in this region. The buildings are detected in two phases firstly extraction and verification of sunshine parts and secondly 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 the prior knowledge of illumination.

In [91] an automated de-shadowing technique is being implemented in five different color spaces such as HSI, HCV, HSV, YC_bC_r and YIQ models to detect buildings from color images. This method is processed in four steps: color transformation, shadow segmentation, shape preservation, and shadow compensation. The author also presents a comparative evaluation of the results obtained in different color 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 the two-dimensional and three- dimensional information for detected building rooftops from remote sensing image is presented in [71]. The author terms their approach as “Stochastic” as thy use contour-based grouping for generating the hypothesis for rooftops. However, the dependencies and relationship between different hypotheses are represented using the well-known Markov random field model. The proposed methodology was used for detecting buildings of variant color, shape, and height from the set of airborne and synthetic images.

In [92] an efficient successive thresholding scheme (STS) was proposed which was able to detect the shadows from the aerial images. This approach improved and updated the ration map obtained from the exponential function using the Tsai's algorithm. Later, 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 in [93]. Firstly the watershed segmentation is used for partitioning the 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 the buildings of complex shapes and different colors. A multi-spectral shadow detection algorithm is described in [94] which incorporated the benefits of the near-infrared bands. The proposed method has been evaluated using three IKONOS images and one GeoEye image of 1meter and 0.5 meters of spatial resolution respectively. Based on visual examination it was stated that the proposed approach has the potential of detecting natural as well as artificial shadows.

A system proposed in [95] is divided into two parts: (i) two-dimensional rooftop detection and (ii) three-dimensional building estimation. The author employs the different image primitives (line intersection detection and line linking) in the first part for examining the relationship between the image primitives using a graph-based method for creating and refining the hypothesis. However, the second part of the proposed system is 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 twenty quick-bird images (sample result shown in figure 5). A novel approach and automated approach for detecting buildings [3] were introduced from very high-resolution optical satellite images. The author used two key algorithms i.e. fuzzy landscape generation [96] and grab cut partitioning to extract the buildings from the images of twenty different sites captured from two different satellites (Quick-Bird and GeoEye-1). The results were evaluated using the pixel and object-based approach. However, the buildings whose shadow was not visible were missing in the generated output.

A supervised classification (enhanced parallelepiped) approach is presented in [97] for detecting buildings from google earth images. The performance of the presented method is evaluated using the object and pixel-based methods. Using the advantages of supervised and unsupervised learning algorithms a novel framework called "self-supervised decision fusion" is suggested in

[98]. The proposed 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 decision approach for classification at base and meta layer. The results of the proposed framework were validated over nineteen test sites (multispectral images of QuickBird, WorldView 2, and GeoEye-1) which were further compared with the other algorithm to prove the viability of the method.

Again the shadow information proved to be an important factor for detecting the object (buildings) from remote sensing images. The framework introduced in [99] combines the knowledge of the urban region (with the help of graph cut) and shadow information for detecting buildings. The reliability and quality of the proposed method are assessed with fourteen test images of IKONOS-2 and QuickBird. The obtained results were compared with the method proposed in [3], [96] and [100] to confirm the supremacy of the proposed method. Further, an unsupervised method to extract buildings along with roads HRS image is presented in [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 twelve multispectral test images of GeoEye-1. Later, a morphological based framework is suggested in [1] which consists of five steps: (i) morphological improvement, (ii) clustering using multiseed based approach, (iii) shadow detection, (iv) minimizing the false alarms and (v) segmentation. The proposed framework is evaluated using the QuickBird and IKONOS images. The comparative evaluation of the obtained results with other existing methods [84] and [102] proves the efficiency of the proposed framework.

2.5 Cognition based methods

The objective of the cognitive model presented in [103] is to assess and analyze the damage caused to buildings from the remotely sensed images. It also aims to understand the underlying cognitive process and attempts to make use of the knowledge iteratively which is necessary for interpreting imagery. This model is based on the principle of fuzzy theory and cognitive theory for interpreting the satellite image 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 for determining the changes [103]. The cognitive model also uses the reasoning as a key cognitive parameter by providing a semantic meaning to the objects which have been damaged. There is a different process of object recognition and image understanding, based on fuzzy theory and cognitive theory, therefore, this model attempt to simulate the process of interpreting the remotely sensed images by human beings. The cognitive model for damage assessment used the HRS image of Quick bird and IKONOS which have the resolution under 1m [103] The overall method has been implemented in three steps [103]. In the first step, low-level features such as texture, color, shape, and tone are extracted by using image processing and object recognition techniques. There exist two methods of getting low-level features firstly, the original image is segmented which consists of object attributes and features secondly, by using filters 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 fuzzy logic approaches through membership function for image understanding and object recognition. This procedure is similar to the way human being carries out the process of image understanding [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 damaged assessment needs to be tested on the aerial images to determine the quality of the model in terms of accuracy. In [8] the author aims to emulate human cognitive processes by integrating cognitive task analysis for information extraction from high-resolution satellite images. A deep understanding of human cognitive capabilities is required to automate the method of information retrieval from HRS images. The author draws the theories from cognitive system engineering which are combined with the studies of geospatial studies. First, the preliminary knowledge about the cognitive processes which human beings acquire during the interpretation of satellite images is collected. Then, knowledge is represented in the form of rules which are based on the visual interpretation of the images by human beings. During knowledge elicitation, these rules are used to extract buildings from HRS images utilizing the mixture tuned matched filtering algorithm. Later, the method is tested using 14 HRS images of an urban area (sample result shown in figure 2). In [9] the author introduces a new cognitive-based automated approach for detecting buildings from very high-resolution (VHR) multispectral images. VHR satellite imagery is a valuable source of information to many researchers working in the extraction of geospatial information. The cognitive processes used by humans are emulated and incorporated with cognitive task analysis (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 is assessed over 14 test images of an urban region.

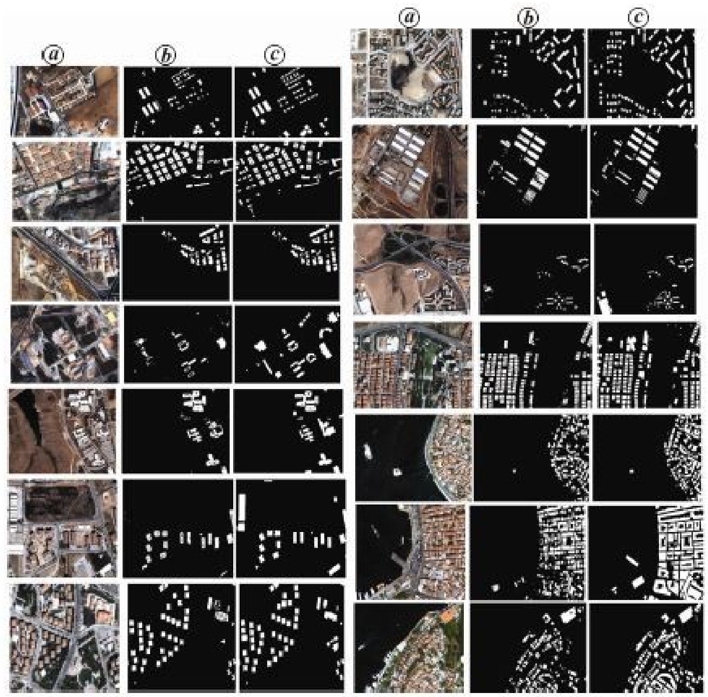

Figure 2: Results of cognitive-based methods [8]: (a) test images; (b) detected buildings; (c) corresponding ground truth data

Similar to the method proposed in [8] and [9] the building detection from aerial and satellite images has been presented in [104] and [105] respectively. To add, a similar approach has been implemented in [106] and [107] detecting road networks from HRS images using the data set presented in [108]. To analyze the satellite image cognitively the method suggested in [109] utilizes the eye-tracking technology to record the eye movements and the response of the participants to stimuli. Similarly in [110] and [111], the process of human-based segmentation HRS images has been experimented using eye-tracking technology.

The author in [112] and [113] suggested their views associated with the role of cognition and visual perception in the analysis of remote sensing images. 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 [114]. The article provided in [115] describes the cognitive factors present in the interpretation of the aerial photograph. The author also presents details of the literature available from 1922-1960 about the cognitive views on the interpretation of aerial photographs [115] and [116]. Besides, the cognitive factors which were needed in interpretation and preparing maps in suggested in [117]. A cognitive approach in identifying the damage caused due to hurricane from remote sensing image is showcased in [118]. The cognitive approach for detecting buildings has been proposed in [119]. The 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] and [121]. The validation of the proposed method is performed using PolSAR images.

2.6 Deep learning models

Deep learning the subset of machine learning and artificial intelligence is a key breakthrough in 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 includes VGGNet, AlexNet, GoogLenet, and ResNet. This section consists the of deep learning based methods proposed in last few years for object detection in particular buildings from satellite images.

A deep CNN was employed in [124] to develop and automated framework for building detection from very high resolution remote sensing images of WorldView-2 and QuickBird containing eight and four spectral bands respectively. The framework used the supervised classification approach for training and Markov-Random-field to detect the labels. The ImageNet framework was utilized to detect buildings through trained data. In [125] a CNN model was trained to classify multi-spectral imagery to identity the buildings in a particular patch. The images of Landsat 8 were used to evaluate the developed model. Similarly CNN was also used in [126] to detect buildings from HRS images. The data set 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% while the producer's and user's accuracy calculated was 35% and 48% respectively. In [127] and approach with reduced complexity was presented for classifying and extracting buildings from Synthetic Aperture Radar (SAR) imagery. 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., air-borne SAR (Shifang and Dujiangyan, China) and TerraSAR-X(Spain, Barcelona, Japan and Sendai). The produced accuracy was greater than 92%. The Conditional Random Field (CRF) model used boundary/edge localization to increase the accuracy [128]. The SpaceNet data set was used to validate the proposed method with greater than 92% of accuracy. Further, a new method which first calculates the NDVI was presented in [129] which later integrated that information within Res-U-Net. The proposed method was validated through ISPRS-2D-semantic-labelling data set of Germany (Vaihingen and Potsdam). The f1-score obtained postdam and Vaihingen data set was 0.9390 and 0.9515 respectively. Another method presented in [130] used Binary Distance Transformation approach to improve data labeling and U-Net model for detecting buildings from multi-spectral images. The images of four cities (Vegas, Shanghai, Paris, and Khartoum) included in SpaceNet challenge dataset are used to evaluate the the model. The maximum f1 score obtained was 0.883 of Vegas however the minimum was 0.584 of Khartoum. To integrate the structure-based information of object (buildings), the Xception-module were replaced with vanilla-U-Net-encoder for extracting buildings [131]. This approach was termed as multitasking learning based method. The Massachusetts (151 images) and Vaihingen (33 images) building detection data set were employed for validation. The overall accuracy calculated for Massachusetts and Vaihingen data set was 94.23% and 96.53% respectively. A multi-source based method for building extraction using U-Net is presented in [123]. The author created their own data set known as WHU building detection data set containing 220 000 buildings in aerial images of New Zeland. The proposed model was also validated with the other well-known data set such as ISPRS, INRIA, and Massachusetts building detection data set. The quantitative and the comparative evaluation proved the potential of the proposed method. In [132] the buildings from Sentinel 1 SAR images were extracted to explore the capability of Unet algorithm. The proposed method is based on CNN. The author 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 vehicle (UAV) were employed to detect the buildings [133]. The faster R-CNN model was trained with 800 images and the accuracy obtained on 200 test images was 92.3%. A new approach i.e, dense-residual-neural-network (DR-Net) was presented in [134]. It is the combination of three models namely, densely-connected CNN, deep-labv3+Net decoder or encoder and residual network. The proposed model included the less number of parameters (nine million) in comparison with BRR-Net (seventeen million). The model shows the increased f1 score on both the WHU (1.4%) and Massachusetts (2.9%) building detection data sets. A multi-tasking based method for semantic-segmentation of buildings is defined in [135]. In this method the performance of U-Net model has been improved through encoder (one) and decoders (two). In addition, a joint-less-function (collection of Mean-Square-Error and Negative-Log-Likelihood) is introduced which is capable of operating two tasks together. The method is evaluated on ISPRS-2D-semantic labeling data set. The f1-score estimated was 94.53% which was greater than the other methods (DAN, MFRN, and Deep-Lab-V3). Recently in [122] Sentinel 1 and Sentinel 2 data set are used within open street map to train U net model to detect both buildings and roads. The training and testing data set includes thirty one and thirteen zones of Spanish cities. The qualitative and quantitative output proved the potential of the method. Thus on the basis of the quantitative results it has been observed that deep learning has proven to be an efficient method for detection of buildings from RSIs.

The description of the building detection methods is given in table 1.

Table 1. Description of previous work on building detection methods from satellite images [8] and [9] (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

Author	Data Type	Image Type	Article Type	Year
Huertas & Nevatia	ABD	GSI	RsC	1988
Irvin & Mckeown				1989
Liow & Pavlidis				1990
Shufelt & Mckeown				1993
McGlone & Shufelt				1994
Weinder & Forstner				1995
Krishnamachari & Chellappa				1996
Baillard	ELD			1998
Zang	SBD	MSI		1999
Helmut Mayer	SBD & ABD	GSI & MSI	RvC	1999
Stassopoulou & Caelli	ABD	GSI	RsC	2000
Cord et al	ELD	GSI		2001

Ruther et al	ELD	GSI		2002
Lee et al	SBD	GSI		2003
Benediktsson et al	SBD	GSI		2004
Baltsavias	SBD & ABD	GSI & MSI	RvC	2004
J. Peng & Y. C. Liu	ABD	GSI	RsC	2005
Unsalan and Boyer	SBD & ABD	GSI & MSI	RvC & RsC	2005
Brenner	SBD & ABD	MSI & LDD	RvC & RsC	2005
Hongjian & Shiqiang	LDD	GSI	RsC	2006
Sohn & Dowman	SBD & ABD	MSI		2007
Katartzis & Sahli	ABD	MSI		2008
Karantzalos & Paragios	SBD & ABD	GSI		2009
Salman Ahmadi et al	ABD	GSI		2010
Haala & Kada	ABD & ELD	LDD	RvC	2010
Cui et al	ABD	MSI	RsC	2011
Tack et al	SBD			2012
Mohammad Izadiand & Parvaneh Saeedi,				2012
Senaras et al				2013
Ali Ozgun Ok				2013
Ali Ozgun Ok et al				2013
Lihong Kang et al				2014
Jiaojiao Tian et al,				2014
Kovacs & Ali Ozgun Ok				2015
Yansheng Li et al		GSI		2015
Caglar Senaras & Fatos T. Yarman Vural		MSI		2016
Gregoris Liasis & Stavros Stavrou		GSI		2016
Gong Cheng & Junwei Han	ABD & SBD	GSI & MSI	RvC	2016
Ali Ozgun Ok	SBD	MSI	RsC	2016
D. Chaudhuri et al	SBD			2016
N.Chandra & J.K. Ghosh				(March) 2017
N.Chandra & J.K. Ghosh				(August) 2017
Dimitrios Konstantinidis et al				2017
N.L. Gavankar & S.K. Ghosh				2018
Masayu Norman et al				2019
N.L. Gavankar & S.K.				2019

Ghosh				
S. Shirowzhan et al	ABD	LDD		2020
Wang, X. and Li, P.				2020
Huiwei Jiang et al	SBD	MSI		2020
Meng Chen et al				2021
Khaled Moghalles et al				2021
Christian Ayala et al				2021

3 Data Set and Evaluation metrics

The previous section discuss about the methods developed for detecting buildings from satellite images. The potential of those methods has been evaluated with the available benchmark data sets. This section describes about the data set used by the several authors for qualitative and quantitative evaluation of their proposed method. The authors have used satellite and aerial images. Each data set discussed in this section include 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 data set is 33.6 MB. The building detection data set consists of 14 images obtained from two very high-resolution satellites namely Quick Bird and IKONOS-2 having a resolution of 0.60 m and 1 m respectively. Four images are obtained IKONOS-2 and ten images are obtained from Quick bird. 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 fourteen images of building detection data set ranged between 144.8494 to 157.1640 and 24.6573 to 31.8230 respectively [96]. This data set has been employed in [8],[9] to validate the developed cognitive method for building detection. In addition, this data set is also used in [96], [99], and [101] for evaluating their proposed method. **Second**, SZTAKI-INRIA benchmark dataset [136] contains the 655 rectangular footprints of the buildings in nine aerial/satellite images captured from IKONOS, QuickBird, and google earth. These images were taken from Bodensee (Germany), Manchester (United Kingdom), Normandy, Szada, Budapest, and Cot d'Azur. The methods described in the section 2.4 has used this data set for validation [100]. **Third**, Massachusetts buildings data set contains one hundred and fifty-one aerial images of the suburban and urban regions of Boston [137]-[138]. The images are of 1500x1500 pixels in size. The area covered by each image is of 2.25 kilometer square at the resolution of one meter square per pixel. The data set is divided into three categories: training data (1108 images), testing data (49 images), and validation data (14 images). This data set is used in [137]-[138] to validate the proposed machine learning based approach for building detection from aerial imagery. **Fourth,** Inria's dataset [139] was constructed to accomplish the pixel-wise labeling of the aerial images. The data set contains the aerial orthorectified colored images with geographical coverage of eight hundred ten-kilometer square with a spatial resolution of 0.3 meters. The data set is divided into two subsets: (i) training data set ($405km^2$) and (ii) testing data set ($405km^2$). The data set consists of various cities such as Chicago, Austin, Kitsap County, Western Tyrol, Vienna, Innsbruck, San Francisco, Bellingham, Eastern Tyrol, and Bloomington. Each region in training and testing data set contains thirty-six tiles (5000*5000 pixels in size) covering the area of 1500*1500 meters. The semantic labeling based method proposed in [139] utilized this data set for validation. **Fifth**, the WHU (Wuhan University) data set consists of more than two lakh twenty-two thousand manually edited aerial as well as satellite images [140]. This data set is divided into four categories: (a) the aerial images [140] includes independent buildings of Christchurch's, New Zealand with a ground resolution of 0.075 meters. The key factor of these images is that they were manually edited and is available on the official website of 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 part of the dataset is down-sampled to one lakh eighty-seven thousand buildings (0.3 meters of ground resolution) and cropped into eight thousand one hundred eighty-nine tiles with the size of 512*512 pixels. These samples are further divided into three parts: (i) training data includes the 130,500 buildings having 4736 tiles, (ii) validation data comprises of 14,500 buildings with 1036 tiles, and (iii) testing data has 42,000 buildings with 2416 tiles, (b) satellite dataset I [140] contains two hundred and four images of global cities (New York, Milan, Venice, Cordoba, Santiago, Cairo, Wuhan, Taiwan Los Angeles, and Ottawa) with resolution ranging from 0.3 meters to 2.5 meters. As the images were taken from different satellite sensors (IKONOS QuickBird, Worldview series, and ZY-3) resulting in the differences in radiometric and atmospheric corrections, atmospheric conditions, multispectral and panchromatic fusion algorithms still made the dataset challenging for evaluating the effectiveness of the proposed building detection method, (c) satellite dataset II includes 29,085 images of six neighbouring satellites on East Asia with geographical coverage of 550 km^2 and has ground resolution of 2.7 meters [140]. The training data set contains 21,556 buildings with 13,662 tiles whereas the testing data set includes 7529 buildings with 3726 tiles. This data set is primarily prepared to test the potential of deep learning methods, building change detection dataset [140] includes 12,796 buildings that were rebuilt due to the damage caused by an earthquake (6.3-magnitude) in February 2011. These aerial images were captured in April 2012 covering the area of 20.5km^2. In 2016, this database was updated by adding 3281 buildings. The WHU data set is used in [140] in validate the developed deep learning based method for building detection from aerial and satellite imagery. **Sixth**, the high-resolution aerial image of Graz city mainly consists of the buildings [141]. The size of this data set is 512*511 pixels. It is obtained from UltraCamD from Microsoft Photogrammetry having three color channels i.e., red, green, and blue. The camera UltraCamD is capable of delivering the image of size 3680*7500 pixels along the track and across-track respectively. The high-resolution aerial image is used in [141] for testing the performance of their model (Hierarchical pseudo-Conditional Random Field) for building detection from aerial imagery. The Graz data set is used in [73] to compare the labeling methods. There are several other data set which includes the multiple classes along with the buildings. The details of such types of data set are mentioned in [142]-[145], however the list of benchmark available data set is given in table 2.

Table 2: List of available data set

Building detection data set	**Source**
SZTAKI-INRIA	http://mplab.sztaki.hu/remotesensing/building_benchmark.html
Massachusetts buildings dataset	https://www.cs.toronto.edu/~vmnih/data/
Inria building detection	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/detection-and-reconstruction/
SpaceNet 2	https://spacenet.ai/spacenet-buildings-dataset-v2/

The methods described in section has used the data set discussed in section. These methods are needed to be validated to prove the potential of their proposed methods. Several methods are described in the literature for determining the accuracy of the results. Therefore, he methods which can be used for qualitative evaluation has been discussed. In general, evaluation and assessment of the results are performed by comparing the results of the method with 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 [3], [96], and [111] given in equations 1, 2, and 3 respectively are best suited. Precision also is known as a positive predictive value represents that how many instances which are selected are relevant whereas recall also known as sensitivity represents that 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 objects, which are detected corresponds to the reference dataset. The precision determines the false positive in the algorithm, however; recall determines the correctly detected objects by the algorithm.

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

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

$$f-score=\frac{(2\times precision\times recall)}{(precision+recall)} \tag{3}$$

During the assessment, all the pixels of the image are classified into three different classes namely, True Positive (TP), False Positive (FP), and False Negative (FN) [3]. Firstly, TP indicates a pixel that is labeled as a building by the proposed method and represents to building in the ground truth dataset. Secondly, FP signifies a pixel that does not represent any of the pixels labeled as buildings in the ground truth dataset, and thirdly, FN represents a pixel that is labeled as building in the ground truth dataset but it is not available in the proposed method. In equations 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 authors [8], [9], [98], [99], and [100] have used these metrics for validating their research.

Some of the authors define the terminologies stated in (1), (2), and (3) as shape accuracy, completeness, and correctness [67] and [132]. The shape accuracy is calculated using equation 4 where X_1 represents the true building region and X_2 denotes the corresponding detected values. The completeness defines the ratio of the detected buildings to the total number of buildings available in the imagery. Further, correctness is the ration of truly detected buildings to the total number of buildings. Correctness represents accuracy based on boundary extraction.

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

The other evaluation terminologies used in building detection approach are branching factor (equation 5), miss factor (equation 6), percentage of building detection (equation 7) and quality percentage (equation 8) [133] and Kappa coefficient (k) (equation 9) chance agreement (equation 10) [40] and [134].

$$Branchingfactor = \frac{FP}{TP} \quad (5)$$

$$Missfactor = \frac{FN}{TP} \quad (6)$$

$$Percentageofbuildingdetection = 100 \times \frac{TP}{(TP+FN)} \quad (7)$$

$$Qualitypercentage = 100 \times \frac{TP}{(TP+FP+FN)} \quad (8)$$

$$KappaCoefficient(k) = \frac{(TP+TN).(TP+TN+FP+FN) - chanceagreement}{(TP+TN+FP+FN)^2 T - chanceagreement}$$

(9)

$$ChanceAgreement = ¿\,(TP+FP).(TP+FN).\,(TN+FN).(TN+FP) \quad (10)$$

The Receiver Operating Characteristics (ROC) curve analysis is a well-known pixel-based evaluation technique [42]. The ROC generally represents the relationship of true positive pixels against false positive pixels. The area under the ROC curve is utilized for evaluating the quality of each change index and the produced output.

4 Conclusion

This publication categorizes and summarizes the algorithms/ methods /techniques of building detection from remote sensing images researched in the last few years. Section 1 provides the introduction, importance, and application of building detection methods. Section 2 describes the various building detection methods along with their algorithm, principle, data, advantages, and limitations.

The low level features based methods utilize the descriptors (such as color, geometrical properties, texture, contrast, shape, regularity) of and image to detect the objects. These descriptors are the basic features which include the 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. These models proves to be the best fit in the condition where the approximate/estimated shape of the boundaries are known. Snake models can be used to monitor dynamic objects while the the process of convergence defines the accuracy. The graph-based methods generates the hypothesis and topological relationships which simplifies 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 chromacity). These methods are robust however computationally costly. Recently proposed cognitive based methods explores the human cognitive capabilities such as reasoning, and perception to detect the objects from remote sensing images. These models determines and monitors the human brain activity involved in the process of object detection. In future, these models will mimic the process of object detection as humans does. On the other hand deep learning based methods learns directly from the data (image, sound, and text) to produce output. These models are flexible which can be used to find solution to complex problems in future. The same model can be applied in range of applications. Moreover, high volume of data set is required to produce the better accuracy when compared with the traditional methods. The training is highly expensive due to the hundreds of machines. Section 3 describes the details of the benchmark data set used by different authors for detecting buildings from remote sensing images. It also contains the description of the evaluation metrics used in previous studies for quantitative evaluation of the results. From the review, it is noticed that the developed methods are robust and efficient however they exhibit some limitations which are needed to be improved. Therefore, the researcher may focus on the following in the future:

i Improvement in detecting the boundaries of the buildings.

ii Some of the methods are unable to detect the building regions clearly due to the variation in shape, size, density, and color of the buildings.

iii The cognitive-based approach is needed to be tested on the images covering the large geographical area.

iv Few techniques fail to extract the buildings having mixed rooftops (texture or shade).

v The impact of haze or snow in building detection is needed to be studied.

vi Improvement in partially detected buildings due to noise.

vii To improve the accuracy of the algorithm the quality if the training data is needed to be improved.

viii The developed models perform well for an urban region where buildings are densely populated whereas they fail to produce satisfactory results in rural regions where the buildings are sparsely located. Therefore a generalized model is needed to be developed which can deliver effective results for urban as well as rural regions.

Hence to conclude, 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.

References

1. Chaudhuri, D., Kushwaha, N. K., Samal, A., & Agarwal, R. C. (2015). Automatic building detection from high-resolution satellite images based on morphology and internal gray variance. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*(5), 1767-1779.

2. Sirmacek, B., & Unsalan, C. (2009). Urban-area and building detection using SIFT keypoints and graph theory. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(4), 1156-1167.

3. Ok, A. O. (2013). Automated detection of buildings from single VHR multispectral images using shadow information and graph cuts. *ISPRS journal of photogrammetry and remote sensing*, *86*, 21-40.

4. Sirmacek, B., & Unsalan, C. (2010). A probabilistic framework to detect buildings in aerial and satellite images. *IEEE Transactions on Geoscience and Remote Sensing*, *49*(1), 211-221.

5. Konstantinidis, D., Stathaki, T., Argyriou, V., & Grammalidis, N. (2016). Building detection using enhanced HOG–LBP features and region refinement processes. *IEEE Journal of Selected topics in applied Earth observations and Remote Sensing*, *10*(3), 888-905.

6 Cheng, G., & Han, J. (2016). A survey on object detection in optical remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, *117*, 11-28.

7 Liasis, G., & Stavrou, S. (2016). Building extraction in satellite images using active contours and colour features. *International Journal of Remote Sensing*, *37*(5), 1127-1153.

8 Chandra, N., & Ghosh, J. K. (2017). A cognitive method for building detection from high-resolution satellite images. *Current Science (00113891)*, *112*(5).

9 Chandra, N., & Ghosh, J. K. (2018). A cognitive viewpoint on building detection from remotely sensed multispectral images. *IETE Journal of Research*, *64*(2), 165-175.

10 Gavankar, N. L., & Ghosh, S. K. (2018). Automatic building footprint extraction from high-resolution satellite image using mathematical morphology. *European Journal of Remote Sensing*, *51*(1), 182-193.

11 Gavankar, N. L., & Ghosh, S. K. (2019). Object based building footprint detection from high resolution multispectral satellite image using K-means clustering algorithm and shape parameters. *Geocarto International*, *34*(6), 626-643.

12 Jiang, X., He, Y., Li, G., Liu, Y., & Zhang, X. P. (2019). Building Damage Detection via Superpixel-Based Belief Fusion of Space-Borne SAR and Optical Images. *IEEE Sensors Journal*, *20*(4), 2008-2022.

13 Mayer, H. (1999). Automatic object extraction from aerial imagery—a survey focusing on buildings. *Computer vision and image understanding*, *74*(2), 138-149.

14 Baltsavias, E. P. (2004). Object extraction and revision by image analysis using existing geodata and knowledge: current status and steps towards operational systems. *ISPRS journal of photogrammetry and remote sensing*, *58*(3-4), 129-151.

15 Brenner, C. (2005). Building reconstruction from images and laser scanning. *International Journal of Applied Earth Observation and Geoinformation*, *6*(3-4), 187-198.

16 Gruen, A., & Dan, H. (1997). TOBAGO—a topology builder for the automated generation of building models. In *automatic extraction of man-made objects from aerial and space images (II)* (pp. 149-160). Birkhauser, Basel.

17 Gruen, A., & Wang, X. (1998). CC-Modeler: a topology generator for 3-D city models. *ISPRS Journal of Photogrammetry and Remote Sensing*, *53*(5), 286-295.

18 Gulch, E., Müller, H., Labe, T., & Ragia, L. (1998). On the performance of semi-automatic building extraction. *International Archives of Photogrammetry and Remote Sensing*, *32*, 331-338.

19 Gulch, E., Muller, H., & Labe, T. (1999). Integration of automatic processes into semi-automatic building extraction. *International Archives of Photogrammetry and Remote Sensing*, *32*(3; SECT 2W5), 177-186.

20 Haala, N. (1996). *Building reconstruction by combining picture and height data* (p. 12). Bavarian Academy of Sciences.

21 Henricsson, O., & Baltsavias, E. (1997). 3-D building reconstruction with ARUBA: a qualitative and quantitative evaluation. In *Automatic extraction of man-made objects from aerial and space images (II)* (pp. 65-76). Birkhauser, Basel.

22 Fischer, A., Kolbe, T. H., Lang, F., Cremers, A. B., Forstner, W., Plumer, L., & Steinhage, V. (1998). Extracting buildings from aerial images using hierarchical aggregation in 2D and 3D. *Computer Vision and Image Understanding*, *72*(2), 185-203.

23 Baillard, C., & Zisserman, A. (1999, June). Automatic reconstruction of piecewise planar models from multiple views. In *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)* (Vol. 2, pp. 559-565). IEEE.

24 Suveg, I., & Vosselman, G. (2004). Reconstruction of 3D building models from aerial images and maps. *ISPRS Journal of Photogrammetry and remote sensing*, *58*(3-4), 202-224.

25 Brenner, C., & Haala, N. (1998). Rapid acquisition of virtual reality city models from multiple data sources. *International Archives of Photogrammetry and Remote Sensing*, *32*, 323-330.

26 Brenner, C. (2000). Towards fully automatic generation of city models. *International Archives of Photogrammetry and Remote Sensing*, *33*(B3/1; PART 3), 84-92.

27 Vosselman, G. (1999). Building reconstruction using planar faces in very high density height data. *International Archives of Photogrammetry and Remote Sensing*, *32*(3; SECT 2W5), 87-94.

28 Vosselman, G., & Dijkman, S. (2001). 3D building model reconstruction from point clouds and ground plans. *International archives of photogrammetry remote sensing and spatial information sciences*, *34*(3/W4), 37-44.

29 Rottensteiner, F., & Briese, C. (2003). Automatic generation of building models from LIDAR data and the integration of aerial images.

30 Unsalan, C., & Boyer, K. L. (2005). A system to detect houses and residential street networks in multispectral satellite images. *Computer Vision and Image Understanding*, *98*(3), 423-461.

31 Haala, N., & Kada, M. (2010). An update on automatic 3D building reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing*, *65*(6), 570-580.

32 Shirowzhan, S., Lim, S., Trinder, J., Li, H., & Sepasgozar, S. M. E. (2020). Data mining for recognition of spatial distribution patterns of building heights using airborne lidar data. *Advanced Engineering Informatics*, *43*, 101033.

33 Shufelt, J. A. (1996). Exploiting photogrammetric methods for building extraction in aerial images. *International Archives of Photogrammetry and Remote Sensing*, *31*(B6), 74-79.

34 Zhang, Y. (1999). Optimisation of building detection in satellite images by combining multispectral classification and texture filtering. *ISPRS journal of photogrammetry and remote sensing*, *54*(1), 50-60.

35 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.

36 Wang, X., & Li, P. (2020). Extraction of urban building damage using spectral, height and corner information from VHR satellite images and airborne LiDAR data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *159*, 322-336.

37 Shackelford, A. K., & Davis, C. H. (2003). A combined fuzzy pixel-based and object-based approach for classification of high-resolution multispectral data over urban areas. *IEEE Transactions on GeoScience and Remote sensing*, *41*(10), 2354-2363.

38 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.

39 Inglada, J. (2007). Automatic recognition of man-made objects in high resolution optical remote sensing images by SVM classification of geometric image features. *ISPRS journal of photogrammetry and remote sensing*, *62*(3), 236-248.

40 Sumer, E., & Turker, M. (2013). An adaptive fuzzy-genetic algorithm approach for building detection using high-resolution satellite images. *Computers, Environment and Urban Systems*, *39*, 48-62.

41 Senaras, C., Ozay, M., & Vural, F. T. Y. (2013). Building detection with decision fusion. *IEEE journal of selected topics in applied earth observations and remote sensing*, *6*(3), 1295-1304.

42 Tian, J., Cui, S., & Reinartz, P. (2013). Building change detection based on satellite stereo imagery and digital surface models. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(1), 406-417.

43 Zhang, Q., Huang, X., & Zhang, G. (2016). A morphological building detection framework for high-resolution optical imagery over urban areas. *IEEE Geoscience and Remote Sensing Letters*, *13*(9), 1388-1392.

44 Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(1), 1-22.

45 Norman, M., Mohd Shafri, H. Z., Idrees, M. O., Mansor, S., & Yusuf, B. (2019). Spatio-statistical optimization of image segmentation process for building footprint extraction using very high-resolution WorldView 3 satellite data. *Geocarto International*, 1-24.

46 Kass, M., Witkin, A., & Terzopoulos, D. (1988). Snakes: Active contour models. *International journal of computer vision*, *1*(4), 321-331.

47 Li, C., Liu, J., & Fox, M. D. (2005). Segmentation of external force field for automatic initialization and splitting of snakes. *Pattern recognition*, *38*(11), 1947-1960.

48 Chan, T. F., & Vese, L. (2001). Active contours without edges. IEEE Transactions on Image Processing, Vol. 10, No. 2.

49 Chan, T. F., Sandberg, B. Y., & Vese, L. A. (2000). Active contours without edges for vector-valued images. *Journal of Visual Communication and Image Representation*, *11*(2), 130-141.

50 Lie, J., Lysaker, M., & Tai, X. C. (2006). A binary level set model and some applications to Mumford-Shah image segmentation. *IEEE transactions on image processing*, *15*(5), 1171-1181.

51 Brox, T., & Weickert, J. (2004, August). Level set based image segmentation with multiple regions. In *Joint Pattern Recognition Symposium* (pp. 415-423). Springer, Berlin, Heidelberg.

52 Chen, L., Zhou, Y., Wang, Y., & Yang, J. (2006). GACV: geodesic-aided C–V method. *Pattern Recognition*, *39*(7), 1391-1395.

53 Caselles, V., Kimmel, R., & Sapiro, G. (1995, June). Geodesic active contours. In *Proceedings of IEEE international conference on computer vision* (pp. 694-699). IEEE.

54 Caselles, V., Kimmel, R., & Sapiro, G. (1997). Geodesic active contours. *International journal of computer vision*, *22*(1), 61-79.

55 Yezzi, A., Kichenassamy, S., Kumar, A., Olver, P., & Tannenbaum, A. (1997). A geometric snake model for segmentation of medical imagery. *IEEE Transactions on medical imaging*, *16*(2), 199-209.

56 Siddiqi, K., Lauziere, Y. B., Tannenbaum, A., & Zucker, S. W. (1998). Area and length minimizing flows for shape segmentation. *IEEE Transactions on Image Processing*, *7*(3), 433-443.

57 Vasilevskiy, A., & Siddiqi, K. (2002). Flux maximizing geometric flows. *IEEE transactions on pattern analysis and machine intelligence*, *24*(12), 1565-1578.

58 Pi, L., Fan, J., & Shen, C. (2007). Color image segmentation for objects of interest with modified geodesic active contour method. *Journal of Mathematical Imaging and Vision*, *27*(1), 51-57.

59 Zheng, Y., Li, G., Sun, X., & Zhou, X. (2009). A geometric active contour model without re-initialization for color images. *Image and Vision Computing*, *27*(9), 1411-1417.

60 Hou, Z., & Han, C. (2005). Force field analysis snake: an improved parametric active contour model. *Pattern recognition letters*, *26*(5), 513-526.

61 Yan, P., & Kassim, A. A. (2006). Segmentation of volumetric MRA images by using capillary active contour. *Medical Image Analysis*, *10*(3), 317-329.

62 Guo, T., & Yasuoka, Y. (2002, November). 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* (pp. 25-29).

63 Mayunga, S. D., Zhang, Y., & Coleman, D. J. (2005, August). Semi-automatic building extraction utilizing Quickbird imagery. In *Proc. ISPRS Workshop CMRT* (Vol. 13, pp. 1-136).

64 Peng, J., Zhang, D., & Liu, Y. (2005). An improved snake model for building detection from urban aerial images. *Pattern Recognition Letters*, *26*(5), 587-595.

65 Cao, G., & Yang, X. (2007). Man-made object detection in aerial images using multi-stage level set evolution. *International Journal of Remote Sensing*, *28*(8), 1747-1757.

66 Karantzalos, K., & Paragios, N. (2009). Recognition-driven two-dimensional competing priors toward automatic and accurate building detection. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(1), 133-144.

67 Ahmadi, S., Zoej, M. V., Ebadi, H., Moghaddam, H. A., & Mohammadzadeh, A. (2010). Automatic urban building boundary extraction from high resolution aerial images using an innovative model of active contours. *International Journal of Applied Earth Observation and Geoinformation*, *12*(3), 150-157.

68 Krishnamachari, S., & Chellappa, R. (1996). Delineating buildings by grouping lines with MRFs. *IEEE Transactions on image processing*, *5*(1), 164-168.

69 Kim, T., & Muller, J. P. (1999). Development of a graph-based approach for building detection. *Image and Vision Computing*, *17*(1), 3-14.

70 Croitoru, A., & Doytsher, Y. (2003). Monocular right-angle building hypothesis generation in regularized urban areas by pose clustering. *Photogrammetric Engineering & Remote Sensing*, *69*(2), 151-169.

71 Katartzis, A., & Sahli, H. (2007). A stochastic framework for the identification of building rooftops using a single remote sensing image. *IEEE Transactions on Geoscience and Remote Sensing*, *46*(1), 259-271.

72 Cui, S., Yan, Q., Reinartz, P., & Mansour, N. (2011). Graph search and its application in building extraction from high resolution remote sensing imagery. *Search Algorithms and Applications*.

73 Schindler, K. (2012). An overview and comparison of smooth labeling methods for land-cover classification. *IEEE transactions on geoscience and remote sensing*, *50*(11), 4534-4545.

74 Li, Y., Tan, Y., Li, Y., Qi, S., & Tian, J. (2015). Built-up area detection from satellite images using multikernel learning, multifield integrating, and multihypothesis voting. *IEEE Geoscience and Remote Sensing Letters*, *12*(6), 1190-1194.

75 Smits, P. C., & Annoni, A. (1999). Updating land-cover maps by using texture information from very high-resolution space-borne imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *37*(3), 1244-1254.

76 Weizman, L., & Goldberger, J. (2009). Urban-area segmentation using visual words. *IEEE Geoscience and Remote Sensing Letters*, *6*(3), 388-392.

77 Tao, C., Tan, Y., Yu, J. G., & Tian, J. (2012, July). Urban area detection using multiple kernel learning and graph cut. In *2012 IEEE International Geoscience and Remote Sensing Symposium* (pp. 83-86). IEEE.

78 Huertas, A., & Nevatia, R. (1988). Detecting buildings in aerial images. *Computer vision, graphics, and image processing*, *41*(2), 131-152.

79 Irvin, R. B., & McKeown, D. M. (1989). Methods for exploiting the relationship between buildings and their shadows in aerial imagery. *IEEE Transactions on Systems, Man, and Cybernetics*, *19*(6), 1564-1575.

80 Mohan, R., & Nevatia, R. (1989). Using perceptual organization to extract 3D structures. *IEEE transactions on pattern analysis and machine intelligence*, *11*(11), 1121-1139.

81 Liow, Y. T., & Pavlidis, T. (1990). Use of shadows for extracting buildings in aerial images. In *Structural Pattern Analysis* (pp. 165-180).

82 Shufelt, J., & McKeown, D. M. (1993). Fusion of monocular cues to detect man-made structures in aerial imagery. CVGIP: Image understanding, 57(3), 307-330.

83 McGlone, J. C., & Shufelt, J. A. (1994). *Projective and object space geometry for monocular building extraction* (No. CMU-CS-94-118). carnegie-mellon univ pittsburgh pa dept of computer science.

84 Lin, C., & Nevatia, R. (1998). Building detection and description from a single intensity image. *Computer vision and image understanding*, *72*(2), 101-121.

85 Gevers, T., & Smeulders, A. W. (1999). Color-based object recognition. *Pattern recognition*, *32*(3), 453-464.

86 Stassopoulou, A., & Caelli, T. (2000). Building detection using Bayesian networks. *International journal of pattern recognition and artificial intelligence*, *14*(06), 715-733.

87 Polidorio, A. M., Flores, F. C., Imai, N. N., Tommaselli, A. M., & Franco, C. (2003, October). Automatic shadow segmentation in aerial color images. In *16th brazilian symposium on computer graphics and image processing (SIBGRAPI 2003)* (pp. 270-277). IEEE.

88 Turker, M. U. S. T. A. F. A., & San, B. T. (2004). Detection of collapsed buildings caused by the 1999 Izmit, Turkey earthquake through digital analysis of post-event aerial photographs. *International Journal of Remote Sensing*, *25*(21), 4701-4714.

89 Sarabandi, P., Yamazaki, F., Matsuoka, M., & Kiremidjian, A. (2004, September). Shadow detection and radiometric restoration in satellite high resolution images. In *IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium* (Vol. 6, pp. 3744-3747). IEEE.

90 Peng, J., & Liu, Y. C. (2005). Model and context-driven building extraction in dense urban aerial images. *International Journal of Remote Sensing*, *26*(7), 1289-1307.

91 Tsai, V. J. (2006). A comparative study on shadow compensation of color aerial images in invariant color models. *IEEE transactions on geoscience and remote sensing*, *44*(6), 1661-1671.

92 Chung, K. L., Lin, Y. R., & Huang, Y. H. (2008). Efficient shadow detection of color aerial images based on successive thresholding scheme. *IEEE transactions on geoscience and remote sensing*, *47*(2), 671-682.

93 Akçay, H. G., & Aksoy, S. (2010, July). Building detection using directional spatial constraints. In *2010 IEEE International Geoscience and Remote Sensing Symposium* (pp. 1932-1935). IEEE.

94 Teke, M., Başeski, E., Ok, A. Ö., Yuksel, B., & Şenaras, Ç. (2011, October). Multi-spectral false color shadow detection. In *ISPRS Conference on Photogrammetric Image Analysis* (pp. 109-119). Springer, Berlin, Heidelberg.

95 Izadi, M., & Saeedi, P. (2011). Three-dimensional polygonal building model estimation from single satellite images. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(6), 2254-2272.

96 Ok, A. O., Senaras, C., & Yuksel, B. (2012). Automated detection of arbitrarily shaped buildings in complex environments from monocular VHR optical satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *51*(3), 1701-1717.

97 Ghaffarian, S. (2014). Automatic building detection based on supervised classification using high resolution Google Earth images. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *40*(3), 101.

98 Senaras, C., & Vural, F. T. Y. (2015). A self-supervised decision fusion framework for building detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*(5), 1780-1791.

99 Manno-Kovacs, A., & Ok, A. O. (2015). Building detection from monocular VHR images by integrated urban area knowledge. *IEEE Geoscience and Remote Sensing Letters*, *12*(10), 2140-2144.

100 Benedek, C., Descombes, X., & Zerubia, J. (2010, August). Building detection in a single remotely sensed image with a point process of rectangles. In *2010 20th International Conference on Pattern Recognition* (pp. 1417-1420). IEEE.

101 Ok, A. O. (2016). Automated detection of buildings and roads in urban areas from VHR satellite images. *Journal of Geodesy and Geoinformation*, *3*(1).

102 Liu, W., & Prinet, V. (2005, July). Building detection from high-resolution satellite image using probability model. In *Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS'05.* (Vol. 6, pp. 3888-3891). IEEE.

103 Zhang, C., Wang, T., Liu, X., & Zhang, S. (2010, August). Cognitive model based method for earthquake damage assessment from high-resolution satellite images: A study following the WenChuan earthquake. In *2010 Sixth International Conference on Natural Computation* (Vol. 4, pp. 2079-2083). IEEE.

104 Chandra, N., Sharma, A., & Ghosh, J. K. (2016, April). A cognitive method for object detection from aerial image. In *2016 International Conference on Computing, Communication and Automation (ICCCA)* (pp. 327-330). IEEE.

105 Chandra, N., Ghosh, J. K., & Sharma, A. (2016, April). A cognitive based approach for building detection from high resolution satellite images. In *2016 International Conference on Advances in Computing, Communication, & Automation (ICACCA)(Spring)* (pp. 1-5). IEEE.

106 Chandra, N., Ghosh, J. K., & Sharma, A. (2019). A cognitive framework for road detection from high-resolution satellite images. *Geocarto International*, *34*(8), 909-924.

107 Chandra, N., & Ghosh, J. K. (2016, October). A cognitive perspective on road network extraction from high resolution satellite images. In *2016 2nd International Conference on Next Generation Computing Technologies (NGCT)* (pp. 772-776). IEEE.

108 Das, S., Mirnalinee, T. T., & Varghese, K. (2011). Use of salient features for the design of a multistage framework to extract roads from high-resolution multispectral satellite images. *IEEE transactions on Geoscience and Remote sensing*, *49*(10), 3906-3931.

109 Dong, W., Liao, H., Roth, R. E., & Wang, S. (2014). Eye tracking to explore the potential of enhanced imagery basemaps in web mapping. *The Cartographic Journal*, *51*(4), 313-329.

110 Sharma, A., Kumar Ghosh, J., & Kolay, S. (2019). Fixation data analysis for complex high-resolution satellite images. *Geocarto International*, 1-22.

111 Sharma, A., Ghosh, J., & Kolay, S. (2019). Reference Data Preparation for Complex Satellite Image Segmentation. *IET Image Processing*.

112 Bianchetti, R. A. (2013). Considering visual perception and cognition in the analysis of remotely sensed images. In *Conference on Spatial Information Theory*.

113 White, R. A., Coltekin, A., & Hoffman, R. R. (Eds.). (2018). *Remote sensing and cognition: human factors in image interpretation*. CRC Press.

114 Montello, D. R., & Freundschuh, S. (2005). Cognition of geographic information. *A research agenda for geographic information science*, 61-91.

115 Bianchetti, R. A., & MacEachren, A. M. (2015). Cognitive themes emerging from air photo interpretation texts published to 1960. *ISPRS International Journal of Geo-Information*, *4*(2), 551-571.

116 Bianchetti, R. A. (2014). Looking back to inform the future: The role of cognition in forest disturbance characterization from remote sensing imagery.

117 Barkowsky, T., & Freksa, C. (1997, October). Cognitive requirements on making and interpreting maps. In *international conference on spatial information theory* (pp. 347-361). Springer, Berlin, Heidelberg.

118 Battersby, S. E., Hodgson, M. E., & Wang, J. (2012). Spatial Resolution Imagery Requirements for Identifying Structure Damage in a Hurricane Disaster. *Photogrammetric engineering & remote sensing*, *78*(6), 625-635.

119 Kang, L., Zou, B., Zhang, Y., & Zhang, L. (2014, July). Building detection based on human visual cognition mechanism using PolSAR images. In *2014 IEEE Geoscience and Remote Sensing Symposium* (pp. 2742-2745). IEEE.

120 Zou, B., Zhang, Y., Wang, C., & Cheng, Y. (2015, July). Building cognition method based on human images cognition mechanism in high resolution PolSAR images. In *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 3223-3226). IEEE.

121 Zou, B., Xu, X., Zhang, L., & Song, C. (2018). High-resolution PolSAR image interpretation based on human image cognition mechanism. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *11*(11), 4259-4269.

122 Ayala, C., Sesma, R., Aranda, C., & Galar, M. (2021). A Deep Learning Approach to an Enhanced Building Footprint and Road Detection in High-Resolution Satellite Imagery. *Remote Sensing*, *13*(16), 3135.

123 Ji, S., Wei, S., & Lu, M. (2018). Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(1), 574-586.

124 Vakalopoulou, M., Karantzalos, K., Komodakis, N., & Paragios, N. (2015, July). Building detection in very high resolution multispectral data with deep learning features.

In *2015 IEEE international geoscience and remote sensing symposium (IGARSS)* (pp. 1873-1876). IEEE.

125 Ishii, T., Simo-Serra, E., Iizuka, S., Mochizuki, Y., Sugimoto, A., Ishikawa, H., & Nakamura, R. (2016, December). Detection by classification of buildings in multispectral satellite imagery. In *2016 23rd international conference on pattern recognition (ICPR)* (pp. 3344-3349). IEEE.

126 Miyazaki, H., Kuwata, K., Ohira, W., Guo, Z., Shao, X., Xu, Y., & Shibasaki, R. (2016, July). Development of an automated system for building detection from high-resolution satellite images. In *2016 4th International Workshop on Earth Observation and Remote Sensing Applications (EORSA)* (pp. 245-249). IEEE.

127 Xu, Z., Wang, R., Zhang, H., Li, N., & Zhang, L. (2017). Building extraction from high-resolution SAR imagery based on deep neural networks. *Remote Sensing Letters*, *8*(9), 888-896.

128 Vakalopoulou, M., Bus, N., Karantzalos, K., & Paragios, N. (2017, July). Integrating edge/boundary priors with classification scores for building detection in very high resolution data. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 3309-3312). IEEE.

129 Xu, Y., Wu, L., Xie, Z., & Chen, Z. (2018). Building extraction in very high resolution remote sensing imagery using deep learning and guided filters. *Remote Sensing*, *10*(1), 144.

130 Prathap, G., & Afanasyev, I. (2018, September). Deep learning approach for building detection in satellite multispectral imagery. In *2018 International Conference on Intelligent Systems (IS)* (pp. 461-465). IEEE.

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

132 Emek, R. A., & Demir, N. (2020). Building detection from sar images using unet deep learning method. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *44*, 215-218.

133 Zheng, L., Ai, P., & Wu, Y. (2020). Building Recognition of UAV Remote Sensing Images by Deep Learning. In *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium* (pp. 1185-1188). IEEE.

134 Chen, M., Wu, J., Liu, L., Zhao, W., Tian, F., Shen, Q., ... & Du, R. (2021). DR-Net: An improved network for building extraction from high resolution remote sensing image. *Remote Sensing*, *13*(2), 294.

135 Moghalles, K., Li, H. C., Al-Huda, Z., & Hezzam, E. A. (2021, March). Multi-Task Deep Network for Semantic Segmentation of Building in Very High Resolution Imagery. In *2021 International Conference of Technology, Science and Administration (ICTSA)* (pp. 1-6). IEEE.

136 **Benedek, C., Descombes, X., & Zerubia, J. (2011). Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*(1), 33-50.**

137 Mnih, V. (2013). *Machine learning for aerial image labeling*. University of Toronto (Canada).

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

139 Maggiori, E., Tarabalka, Y., Charpiat, G., & Alliez, P. (2017, July). Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 3226-3229). IEEE.

140 Ji, S., Wei, S., & Lu, M. (2018). Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(1), 574-586.

141 Thuy, N. T. (2009). Object Detection from Aerial Image. *Doctoral dissertation*, *Graz University of Technology*.

142 Yang, Y., & Newsam, S. (2010, November). 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* (pp. 270-279).

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

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

145 Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., ... & Raskar, R. D. (2018). A challenge to parse the earth through satellite images. arXiv 2018. *arXiv preprint arXiv:1805.06561*.