

## **Building Extraction from Satellite Images**

A.S. Bhadauria, H.S. Bhadauria, Anuj Kumar

*Computer Science and Engineering Department  
Govind Ballabh Pant Engineering College ( Pauri garhwal), Uttarakhand*

---

**Abstract:** *A method for detecting buildings from satellite/aerial images is proposed in this paper. The aim is to extract rectilinear buildings by using hypothesis. Hypothesis generation is accomplished by using edge detection and line generation methods. Hypothesis verification is carried out by using information obtained both from the color segmentation of HSV representation of the image. Satellite images is firstly filtered to sharpen the edges. The edges are extracted using Canny edge detection algorithm. These edges are the input for the Hough Transform stage which will produce line segments according to these extracted edges. Then, extracted line segments are used to generate building hypothesis. Verification of these hypotheses makes use of the outputs of the HSV color segmentation. In this study, color segmentation is processed on the HSV representation of the satellite/aerial images which are less sensitive to the illumination. Finally, Buildings are extracted from the urban areas.*

**Keywords:** *Building Extraction, Satellite images, Edge detection, Segmentation.*

---

### **I. Introduction**

Object detection in satellite images has been an important research topic in computer vision for many years. Some useful applications of this subject are; updating of geographic information system (GIS) databases, urban city planning and land use analysis. The fundamental challenges that drive much of the research in this field are the edge or line extraction problems and segmentation problem, which is finding a desired object and separating it from the background in the presence of distractions caused by other features such as surface markings, vegetation, shadows, and highlights. A large variety of building detection techniques and algorithms have been reported in the literature, but most detection algorithms rely on edge-based techniques that consist of linear feature detection, grouping for parallelogram structure extraction, and building polygons verification using knowledge such as geometric structure, shadow, and so forth. In order to solve this complex problem, integrating the power of multiple algorithms, cues, and available data sources is also implemented recently to improve the reliability and robustness of the extraction results. The recent availability of commercial high-resolution satellite imaging sensors such as IKONOS provides a new data source for building detection. The high spatial resolution of the imagery specifies very fine details in urban areas and facilitates the classification and detection of urban-related features such as roads and buildings. Since manual extraction of buildings from imagery is very slow, automated methods have been proposed to improve the speed and utility for urban map's production. Most of the recent work on building extraction from high resolution satellite images is based on supervised techniques. These techniques either require a classification based on initial training data to provide hypotheses for the positions and sizes of the candidate building features, or they use training sets or model databases to classify or match the buildings.

The difficult problem in building detection is performing automatic and accurate matching of the buildings in satellite/aerial imagery. The motivation behind this work is to use shadow and HSV (hue, saturation, intensity) color information gathered from the satellite/aerial image to verify the building hypotheses generated by using edge-based techniques. The approach has to be robust and accurate to be used in a system.

Building detection has attracted many researchers attention in computer vision for many years. Automatic building detection from satellite/aerial imagery involves several different problems related to computer vision since in urban areas there are many other subjects in close proximity such as trees, power lines vehicles, and parking lots and these subjects may occlude the building's rooftops. Besides, the rooftops of the buildings may be composed of different surface materials with differing reflectance properties. The above mentioned problems make the automatic building detection is a challenging problem in the computer vision area. In [1] researchers used snake-based approach to extract 2D building outlines from high resolution IKONOS satellite images and height data captured by airborne laser scanning system. A semi-automated approach is used in [2] based on active contour model (snakes) and the dynamic programming optimization technique. The method requires a digital surface model and an ortho-image. This approach can be more effective if applied after a human

operator has manually determined seed points near the boundary of a desired feature. Peng et al. [3] Extracts the principal contours of buildings in dense urban areas according to the radiometric behavior of buildings. A different approach [4] from the standard/traditional building extraction methods is using wavelet transform and image scaling which uses both the high and the low frequency components of an aerial image. In [5] two or more panchromatic images which need not to be stereo pairs are used to extract flat or symmetric gable roof rectilinear buildings and constructs 3D models. This approach generates hypotheses for rectangular roof components by grouping lines in the image hierarchically and verifies hypotheses by searching for presence of predicted walls and shadows. This algorithm leaves the verification decision at the end, which is different from the general approach hypothesize and verify sequentially. Multiple and overlapping images of the scene are used in [6] to detect complex buildings with flat or complex rooftops. Modeling the complex buildings is a major problem in the research. [7] uses both multi-view stereo and color information as cues for building detection in dense urban scenes. The significant improvement in this approach is the detection rate is achieved by running the detection procedure on different views and combining the results. Cord et al. [8] focus the problem of automatic extraction and modeling from stereoscopic pairs of high resolution (less than 10cm per ground pixel) aerial images. It is based on Digital Elevation Model (DEM) computation which involves a gradient correlation, contour-adaptive windows with a geodesic weighting, a multi-resolution coarse to fine scheme, and a two-way filtering symmetrical validation. In [9] three data sources; LiDAR, an aerial image and ground images are used to extract 3D view of the buildings. Using a hybrid model enables the researchers to collect more accurate edge information and building details compared to using only LiDAR and more detailed surface information is gathered compared to using only stereo aerial images. This method is costly and needs ground images which are not available for the ones who just have only satellite or aerial images. Approach used in [10] for Level of detail building model reconstruction is to use airborne LiDAR data and optical imagery. [11] uses probabilistic theory at its simplest level, assuming that the buildings probability distribution can be represented by a logistic function. In this approach buildings are regarded as individual objects in the image. [12] deals with the automatic extraction of building outlines using a pair of optical and synthetic aperture radar (SAR) images. This method has two main parts; first, extraction of partial potential building footprints on the SAR image, and then shapes detection on the optical one using the previously extracted primitives (lines). Main drawbacks of this approach are the SAR image could be used to validate the buildings detected in the optical image, the primitive SAR detection could be improved and large building detection algorithm be developed.

## II. METHODOLOGY

The building detection approach used in this paper is a model based method, the purpose of which is detecting rectilinear buildings. The buildings on the satellite images are 2D structures which are mostly made up of regions having borders being straight line. These straight lines are the combinations of pixels in an image. So, the building detection algorithm firstly extracts the pixels that compose line segments by means of edge detection process. Then, these extracted pixels are grouped to form straight lines which will be searched for whether they form a rectilinear building or not. Hypothesis verification is carried out by using information obtained from the color segmentation of HSV representation of the image. Edge detection process is accomplished using Canny edge detection algorithm [13]. Then, extracted edges are used to generate line segments through Hough Transform. The color HSV segmented image is produced using k-means clustering algorithm and shadow image is produced by applying a mask to the image represented in the HSV space. The approach presented in the paper utilizes the features of both canny edge detection and K-means clustering methods for the extraction of rooftop of buildings. A brief introduction of canny edge detection and K-means clustering methods is also given in sub-sections 2.1 and 2.2 respectively.

### 2.1 Canny Edge Detection Algorithm:

Edge detection is the process of finding sharp contrasts in intensities in an image. This process significantly reduces the amount of data in the image, while preserving the most important structural features of that image. An edge detection algorithm extracts the main properties from an image which are the building edges for the current study.

Canny edge detection is considered to be optimal edge detection operator. Canny defined optimal edge finding as a set of criteria that maximize the probability of detecting true edges while minimizing the probability of false edges. He found that the zero-crossings of the second directional derivative of a smoothed image were a reasonable measurement of actual edges. To smooth the image, the Canny edge detector uses Gaussian convolution, is the spread of the Gaussian and controls the degree of smoothing.

$$g(m, n) = G(m, n) * f(m, n) \tag{1}$$

$$G_{\sigma} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{m^2 + n^2}{2\sigma^2}\right] \tag{2}$$

Next, the image is convolved with a 2D first derivative operator to determine regions of sharp changes in intensities. The gradient magnitude and direction at each pixel are calculated in this step (Figure 1). Note that the maxima and minima of the first derivative gradient are the same as the zero crossings of the second directional derivative.

$$M(m, n) = \sqrt{g_m^2(m, n) + g_n^2(m, n)} \tag{3}$$

Only the maxima crossings are of interest because these pixels represent the areas of the sharpest intensity changes in the image [13]. These zero-crossings are the ridge pixels that represent the set of possible edges. All other pixels are considered non-ridge and subsequently suppressed. Finally, a two-threshold technique or hysteresis is performed along the ridge pixels to determine the final set of edges.

-1	0	1
-2	0	2
-1	0	1

1	2	1
0	0	0
-1	-2	-1

Figure 1 Kernels

Instead of using a single threshold value for filtering ridge pixels, the Canny algorithm implements a connected components analysis technique based on a hysteresis thresholding heuristic. This step uses two thresholds, t1, t2 where t1 > t2, to partition the ridge pixels into edges/non-edges. Pixels with gradient magnitudes above t1 are classified as definite edges. Pixels between t2 and t1 are classified as potential edges. Pixels under t2 are classified as non-edges. Next, all potential edges that can be traced back to a definite edge via adjacent potential edges are also marked as definite edges. The process solves some of the issues associated with edge streaking and discontinuity in the results achieved by simple detectors by identifying strong edges while accounting for comparatively weaker ones.

### 2.2 K-Means algorithm

One way of interpreting an image is to group and assign pixels to a region due to some criterion. This partitioning of the image into different regions is called segmentation. Formally, segmentation can be defined as a method to partition an image into sub-images called regions of common characteristics such that each segment is a discrete image. Cluster analysis, also called segmentation analysis or taxonomy analysis, is a way to create groups of objects, or clusters, in such a way that the profiles of objects in the same cluster are very similar and the profiles of objects in different clusters are quite distinct.

Given S a set of N points and K the number of clusters, the algorithm chooses K reference points (e.g., at random) from S. Each reference point  $R_i$  defines a cluster  $C_i$ . Then data points are partitioned into K clusters. Point p of S becomes a member of cluster  $C_i$  if p is closer in the underlying metric (e.g., the Euclidian distance) to  $R_i$  the reference point of  $C_i$  than to any other reference point. Closest means  $\min d(R_i, p)$ , where  $d(R_i, p)$  is a distance between point  $R_i$  of R and point p of S in the underlying metric. The centroid for each cluster is calculated and the centroid becomes a reference point of its cluster. During successive iterations, the centroids of each cluster are recalculated. During the iterations, the algorithm goes through all data points and determines if for point p in cluster  $C_i$ , the centroid of  $C_i$  is the nearest reference point. If so, no adjustments are made and the algorithm proceeds to the next data point. However, if the centroid of cluster  $C_j$  becomes the reference point closest to the data point p, then p is reassigned to cluster  $C_j$ , the centroids of the losing cluster  $C_i$  (minus point p) and the gaining cluster  $C_j$  (plus point p) are recomputed, and the reference points of clusters  $C_i$  and  $C_j$  are moved to their new centroids. After each iteration, every one of the K reference points is a centroid, or mean, hence the name "k-means". The iterations proceed until, for all data points, no re-allocation of points from cluster to cluster is possible. Finally, the distribution of points will correspond to the centroidal Voronoi configuration, where each data point is closer to the reference point of its cluster than to any other reference point, and each reference point is the centroid of its cluster.

### III. IMPLEMENTATION AND RESULTS

For the assessment of the proposed method, first the canny edge detection method is applied to the original image as shown in figure 2. Figure 3 shows the results of edge detection.



Figure 2 Original image



Figure 3 Canny edge detected image

#### **Hypotheses generation:**

The goal of the hypothesis generation is to find candidate buildings in the image using the line information from the previous stage. The Hough transform is used to generate the line information from the previous stage. Figure 4 shows the line information generated from the figure 3 using Haugh Transform. In this paper, only rectilinear shaped buildings are considered, so hypothesis generation stage search for rectangular structures in the image. In order to improve the building detection rate, algorithm searches for 3-sided possible rectangles instead of searching for 4-sided rectangles taking into consideration that the fourth side may be lost during edge detection stage or some kind of occlusion due to trees or other structures in the image. In addition to this 3-sided rectangle search, algorithm searches for 2-sided possible rectangles, namely any “L-shapes” again assuming the remaining 2 sides may be lost similarly. Both of these searches assume that two straight line segments are nearly perpendicular to each other and the gap between the intersection points of these two lines is below a threshold value. finally, algorithm also searches single line segments (I-shaped) that may be a part of a building and stores these selected lines as building hypotheses.

Here in this stage, algorithm generates the building hypotheses only if the line segments for each side of the candidate building are larger than the value of `min_line_length` parameter. By adjusting this parameter according to the image size, random line segments which are very small are discarded although they obey U, L or I-shaped hypotheses patterns.

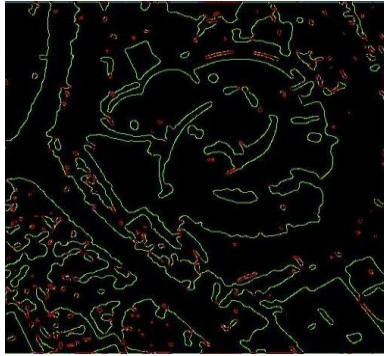


Figure 4 Lines generated by using Hough transform

The buildings are separated from each other on the basis of, none of them is occluded, cast shadows are clear, orientation of the buildings are different, and all of them but one has the same rooftop color and type. Rooftops are not uniform most of the rooftops have different colored chimneys and the background is not heavily populated; some cars, subway station and parking lots add some ambiguity to the image. Then, the lines are removed if their length is under a given threshold and the adjacent lines generated from the Hough Transform stage is merged if and only if the difference between them is under a given threshold. First, the algorithm converts the image into HSV representation and applies k-means algorithm to produce color (HSV) segmented image. It is observed that the buildings have uniform segments except that the chimneys distract the uniformity.

Since the image objects contains the spectral information compared to digital numbers. Brightness and spectral ratios of the image objects were calculated using all image layers. Textural features are estimated using standard deviations of layer values, spectral mean values of image-objects, and average spectral differences of image-objects. In the classification technique only building part is extracted from the image as shown in figure 5.



Figure 5 Final results of building extraction

#### IV. CONCLUSIONS

In this paper, a building detection algorithm from aerial or satellite imagery is proposed and implemented. Proposed algorithm is based on hypothesis generation and verification for rectilinear buildings.. Firstly, the satellite/aerial image is filtered to remove noise and make the edges stronger so that most of them can be detected by edge detection. Next, edge detection is performed using Canny edge detection algorithm. The parameters used in the Canny edge detection stage is critical since these parameters determine the number of points extracted from the image some of which constitute the building parts. These parameters are set to optimum values so as to find maximum building parts and minimum noise and/or spurious points. Then, these extracted edge points are used to generate line segments by using Hough Transform algorithm. In order to correctly generate these line segments, Hough Transform's parameters are optimized so that the algorithm does not produce spurious/false line segments. Next, building hypotheses are formed from these lines which obey the rectilinear building models. Then, edge detection stage plays a key role in the algorithm. The parameters (low threshold, high threshold, and sigma) determine the success rate of this stage. The building detection rate can be improved or degraded by adjusting these parameters, such as by lowering the threshold value more buildings can be detected in the image but a lower threshold also catches more noise in the scene which will result in false positives and an increased runtime of the algorithm. The results of the implemented

algorithm are satisfactory for both urban and sparse environments. Building detection rate is above 80% for most of the images.

### References

- [1] T. Guo and Y. Yasuoka, "Snake-based approach for building extraction from high-resolution satellite images and height data in urban areas", Proceedings of 23<sup>rd</sup> Asian Conference on Remote Sensing, 2002.
- [2] H. Rüther, H. M. Martine and E.G. Mitalo "Application of snakes and dynamic programming optimization technique in modeling of buildings in informal settlement areas" ISPRS Journal of Photogrammetric & Remote Sensing, Vol. 56, 2002.
- [3] J. Peng D. Zhang and Y. Liu "An improved snake model for building detection from urban aerial images" Pattern Recognition Letters, Vol. 26, 2005.
- [4] S. Levitt and F. Aghdasi. "An investigation into the use of wavelets and scaling for the extraction of buildings in aerial images", Proceedings of South African Symposium on Communications and Signal Processing, 1998.
- [5] S. Noronha and R. Nevatia "Detection and modeling of buildings from multiple aerial images" IEEE Transactions on pattern analysis and machine intelligence, vol.23, no.5, 2001
- [6] Z. Kim and R. Nevatia "Automatic description of complex buildings from multiple images" Computer vision and image understanding, Vol. 96, 2004.
- [7] JM. Fradkin, H. maitre and M. Roux "Building detection from multiple aerial images in dense urban areas" Computer vision and image understanding, Vol 82, 2001.
- [8] M. Cord, M. Jordan and J. Cocquerez "Accurate building structure recovery from high resolution aerial imagery" Computer vision and image understanding, Vol 82, 2001.
- [9] J.Hu, S. You and U. Neumann "Integrating LiDAR, aerial image and ground images for complete urban building modeling" Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission, 2006.
- [10] F.O. Chikomo, J.P. Mills and S.L. Barr "An integrated approach to level of detail building extraction using airborne lidar and optical imagery" Photogrammetric Image Analysis, Munich, Germany, 2007
- [11] Wei Liu, Prinet V. "Building detection from high-resolution satellite image using probability model" Geoscience and Remote Sensing Symposium, 2005.
- [12] J. Shufelt and D. McKeown, "Fusion of Monocular Cues to Detect Man-Made Structures in Aerial Imagery," Computer Vision, Graphics, and Image Processing, Vol. 57, No. 3, 1993.
- [13] Canny John, "A Computational Approach to Edge Detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-8(6), pp 679-698, 1986.