Automatic Building Detection From High-Resolution Satellite Images Based on Morphology and Internal Gray Variance

D. Chaudhuri, *Senior Member, IEEE*, N. K. Kushwaha, A. Samal, *Senior Member, IEEE*, and R. C. Agarwal

***Abstract*—Automatic building extraction remains an open research topic in digital photogrammetry and remote sensing. While many algorithms have been proposed for building extraction, none of them solve the problem completely. This is even a greater challenge in urban areas, due to high-object density and scene complexity. Standard approaches do not achieve satisfactory performance, especially with high-resolution satellite images. This paper presents a novel framework for reliable and accurate building extraction from high-resolution panchromatic images. Proposed framework exploits the domain knowledge (spatial and spectral properties) about the nature of objects in the scene, their optical interactions and their impact on the resulting image. The steps in the approach consist of 1) directional morphological enhancement; 2) multiseed-based clustering technique using internal gray variance (IGV); 3) shadow detection; 4) false alarm reduction using positional information of both building edge and shadow; and 5) adaptive threshold based segmentation technique. We have evaluated the algorithm using a variety of images from IKONOS and QuickBird satellites. The results demonstrate that the proposed algorithm is both accurate and efficient.**

***Index Terms*—Index Terms–Building detection, clustering, enhancement, feature extraction, high resolution, morphology, remote sensing, segmentation, thinning.**

I. Introduction

PHOTOGRAMMETRY and remote sensing methods generate geospatial data for large areas with acceptable accuracy. Furthermore, cost and time of producing data using these methods are lower than other production methods. Recent advancement of satellite remote sensing technology, e.g., the successful launch of IKONOS, QuickBird, and CARTOSAT-2A provides global, accurate, very high-resolution multispectral imagery to individuals, commercial organizations, and government agencies for a variety of urban development applications. Automatic detection of geographic objects such as bridges, buildings, and road crossings in satellite images is useful in many important applications including the creation and maintenance of accurate geographical databases, assessment of the extent of damages after natural disasters such as floods or earthquakes and military operations. Extraction of high-level features is also important in the development of content-based indexing schemes for retrieval of satellite images [1], [2].

Buildings form one of the most important group of manmade objects that take significant time and cost of to extract, due to their variety, complexity, and abundance in urban areas. As a result, automatic or semi-automatic extraction of this feature not only minimizes human role in producing large-scale maps but also has a dramatic impact on time and cost. Development of such an algorithm can also facilitate automatic extraction of features similar to buildings.

Many approaches to detect building features have been reported in computer vision, remote sensing, and photogrammetry literature [3]. Recent research covers extracting building information from high-resolution satellite imagery, high-quality digital elevation models (DEM), and aerial images. QuickBird and IKONOS high-resolution satellite images have been used to acquire plane-metric building information with 1-m horizontal accuracy [4], [5]. Because of irregular structure and closeness of buildings in urban areas, much of the research in the domain of automatic building extraction from high-resolution aerial and satellite images are performed with the integration of LIDAR data and images [6]. High-quality DEM such as those available from LIDAR has been used to generate three-dimensional (3-D) building models [7], [8]. Although LIDAR data provide accurate 3-D positions, it is expensive to collect and insufficient to provide surface texture.

Aerial/high-resolution satellite images are the primary source for acquisition of accurate and reliable geospatial information. Muller and Zaum [9] proposed an algorithm to detect and classify buildings from a simple aerial image using a region growing algorithm. Many automatic building extraction methods from digital surface models (DSM) or multispectral imagery are developed but they suffer from rough delineation results due to the relatively low resolution of the associated DSM and multispectral imagery. Song and Shan [10] proposed an algorithm to extract buildings from high-resolution color imagery. In their paper, they have focused on the outline of building boundary and segmentation of building roof polygons or faces. In some research, active contour models such as snakes have been employed for recognizing the boundaries of buildings [11]–[13]. Mayunga *et al.* [12] proposed a semi-automatic building extraction algorithm from QuickBird images by selecting a point on the boundary of each building. Then, the initial curves of the model are produced and the accurate boundaries of buildings are detected using an iterative procedure.

Manuscript received February 19, 2014; revised March 13, 2015; accepted April 09, 2015.

D. Chaudhuri is with the DRDO Integration Centre, Burdwan, West Bengal 713419, India (e-mail: deba_chaudhuri@yahoo.co.in)

N. K. Kushwaha and R. C. Agarwal are with the Image Analysis Center, Defence Electronics Applications Laboratory, Dehardun 248001, India.

A. Samal is with the Department of Computer Science and Engineering, University of Nebraska-Lincoln, Lincoln, NE 68588 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2015.2425655

Given the recent availability of the commercial high-resolution satellite imagery, only a few methods for building detection/extraction from less or equal than 1-m/pixel resolution imagery have been developed. The effect of resolution on the building extraction was reported in [15]. The common challenges in generating building hypotheses from less or equal than 1-m/pixel imagery include: 1) Low-signal-to-noise ratio and 2) Weak object signal.

Most of the recent work on building extraction from high-resolution satellite imagery 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 objects [5], or they use training sets or a model database to classify or match the buildings [14], [15]. Lee *et al.* [5] used multispectral IKONOS imagery to determine the approximate location and shape for buildings. An ISODATA-based clustering followed by a Hough transform is used to derive a fine extraction of the buildings. Benediktsson *et al.* [14] used a morphology-based approach to extract structural information from satellite images. Features generated by a differential morphological profile (DMP) were selected by discriminant analysis. Buildings and other land use categories were then classified using a neural network. Segl and Kaufmann [15] combined supervised shape classification with unsupervised image segmentation for detection of small buildings in suburban areas. A series of image segmentation results were generated by selecting a series of thresholds. The buildings were then classified by matching their shapes with the shapes stored in a model database. Objects with the number of correct shape matches greater than an optimal threshold were classified as buildings. Liu and Prinet [16] proposed a probability model to extract buildings from high-resolution panchromatic images in dense urban areas. They extract original features that characterize buildings and recognize buildings by computing a probability measure on each individual building. Wei *et al.* [17] proposed an algorithm to extract buildings from high-resolution panchromatic QuickBird images using clustering and edge detection. However, the main drawback of these approaches is that they are not fully automated.

In this paper, an automated building-extraction strategy for high-resolution satellite imagery is proposed that utilizes structural, contextual, and spectral information. The system runs automatically without preclassification or any training sets, although some initial algorithm parameters must be set by the user. The proposed method exploits the properties of buildings and other objects in the scene and their interactions in the resulting images. The steps in the method are: 1) local directional morphological operations to enhance the building structures; 2) multiseed-based clustering of internal gray variance (IGV) features of enhanced image; 3) identify edges of manmade structures using bimodality detection; 4) shadow extraction from the enhanced image; 5) false alarm reduction using positional information of building edge and shadow; and 6) adaptive threshold-based segmentation technique. The algorithm is extensively evaluated using a variety of high-resolution satellite imagery and the results demonstrate that our approach is effective in extracting of building structures.

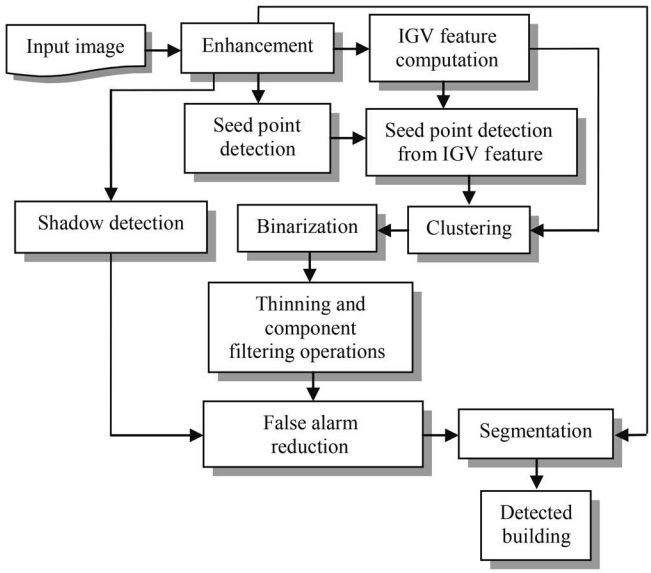


Fig. 1. Schematic of the proposed algorithm.

It should be noted that the resolution of panchromatic images (≤ 1 m) from satellite imagery is greater than that for multispectral images (2.5 m for QuickBird and 4 m for IKONOS). Hence, the panchromatic images carry more details and lend themselves to recognizing buildings more accurately, since buildings are small structures and do not carry significant color information. Therefore, we use only panchromatic images for this research.

II. Building Detection Algorithm

The focus of this paper is the detection of buildings from high-resolution panchromatic images. Our approach is designed to effectively extract the features using a multistep process as shown in Fig. 1. The details of each step are presented below.

A. Image Enhancement

The goal of the enhancement step is to improve the visual effect in the image to facilitate geographic image interpretation and to improve the contrast between the target and nontarget for high-level processing. A good enhancement operator will significantly reduce (or increase) the brightness of dark (or bright) building structures in the original image but has no effect on nontarget pixels. A key observation that is exploited here is that manmade objects tend to be more homogeneous that natural objects. Therefore, morphological operators are useful to delineate them. Since opening suppresses bright regions smaller than the structuring element, and closing suppresses dark regions, they are often used in combination as *morphological filters* for image smoothing and noise removal. The proposed enhancement technique is a combination of opening and closing operations using the same structuring element along the homogeneous direction with respect to the image and a building template.

Chaudhuri *et al.* [18] proposed an enhancement technique, which is based on the local morphological operations [19] that are directional in nature. To enhance the road structures in the image, we used a customized 5×5 structuring element to serve as a road seed template [18]. The enhancement technique is more effective with a smaller window. Using a larger window would pose difficulty in this step at the boundary of the buildings. In general, the structuring element should be small to 1) not remove details and 2) keep the computational cost low. In the case of satellite images, where the image resolution is relatively low, this is particularly important. However, since the buildings are larger structures, it is unlikely that a building can be as small as 3×3. Therefore, we believe that a 5×5 structuring element is optimal for the current generation of satellite images. If the resolution of the images is higher, it is possible to use a 3×3 structuring element. In the worst case, one has to experiment with only structuring elements of two sizes: 3×3 or 5×5. In the present paper, we use the same enhancement technique [18] and interested reader can go through our previous paper [18].

B. IGV Feature Extraction

Enhancement operation described in the previous section results in the internal regions of the manmade structures being more homogeneous than the border region. Since the variation of the internal pixels of a manmade structure is low, the corresponding IGV will be small. At the same time, the IGV will be high at the border pixels of the manmade structure. Furthermore, the variations of the internal pixels and border pixels of the nonmanmade structure (natural structure) are very less. Therefore, the IGV for internal as well as border pixels of nonmanmade structures will be almost the same. This is the motivation for our IGV feature extraction step described next.

Application of standard edge operators to find the boundaries of buildings at this stage is unlikely to be successful since the manmade objects are not well separated from the natural objects. To further enhance the differences between them, we propose the use of the IGV feature. The enhancement technique reduces (or increases) the brightness of dark (or bright) building structures and at the same time it blurs the nonmanmade regions (like trees). As a result, the variance of each pixel within an object is low and is high at the boundary of the objects. Deriving this feature will further simplify the identification of boundary points of manmade structures. Since nonmanmade regions are not sharply different from the background in the enhanced image, the boundaries of these regions are not enhanced by the IGV feature. So, both enhancement and IGV feature extraction techniques are critical to downstream processing steps to extract the buildings accurately. The IGV feature extraction steps are described below.

The IGV is computed by first computing the gray level mean within a window $w \times w$ ($w = 5$). Then, the candidate pixel value is generated by computing the sum of deviation from each pixel from the mean gray value within the window. After the application of the enhancement technique, the manmade objects appear more prominent in the image than the nonmanmade objects. In such cases, the sum of deviations at the boundary of manmade objects is significantly more than that of points inside the manmade objects. In contrast, the sum of deviations at the boundary of the nonmanmade (natural) objects are not as high as manmade objects.

Take a window of size $w \times w$.
For each pixel $(x,\ y)$ in the image h (enhanced image)
{

1. Compute the *mean gray* value at $(x,\ y)$ within the window as,

$$\mu(x,y) = \frac{1}{w \times w} \sum_{i=-w/2}^{w/2} \sum_{j=-w/2}^{w/2} h(x+i, y+j)$$

2. Compute *IGE* at $(x,\ y)$ within the window as,

$$IGV(x,y) = \sum_{i=-w/2}^{w/2} \sum_{j=-w/2}^{w/2} [h(x+i, y+j) - \mu(x,y)]^2$$

}

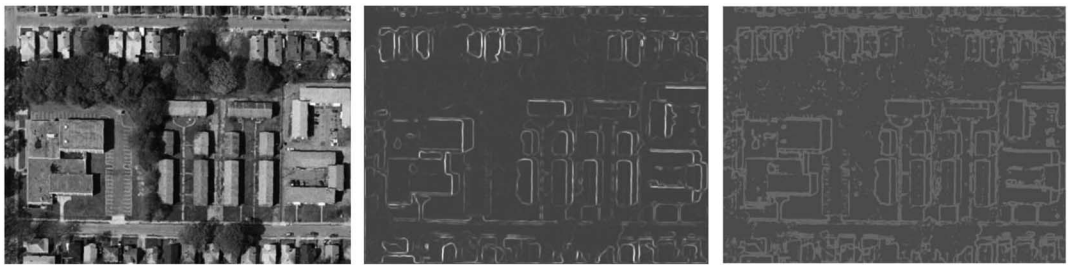
Fig. 2 shows the effect of enhancement technique along with the interior gray variance (IGV) for building images. Fig. 2(a) shows a scene with a set of buildings and Fig. 2(b) and (c) shows the result of IGV feature clustered image with and without the enhancement technique. Fig. 2(c) shows that many nonmanmade objects appear in the image, whereas in Fig. 2(b) fewer nonmanmade objects are clustered, which is the output of IGV feature clustered image after enhancement of original image. This shows the effectiveness of the enhancement technique.

It should also be noted that roads are also manmade objects but since they are flat, shadow information is extracted adjacent to the thick road edges. They are separated from buildings using a separate step later on (removing false alarms).

C. Seed Point Detection Technique

Various manmade structures such as roads, buildings, bridges, and oil tanks are typically present in the scene. The materials of the panchromatic remote sensing satellite image with limited spectral resolution have three characteristics. They are: 1) the number of material classes are unknown (nonparametric); 2) the brightness values of the classes overlap; and 3) the variation of a particular material class may not be unimodal (non-Gaussian). Due to these properties, it is difficult to correctly segment the original panchromatic image, especially manmade and nonmanmade classes in different segments. To overcome this, we have transformed the enhanced image into IGV feature space. We then use a clustering-based segmentation technique to isolate manmade structures from the background. Since seed point detection is an essential part of any clustering technique, this is also described next.

Seed-based clustering methods (like k-means) start with some initial seed points and grow clusters around them. There are two fundamental problems with all seed-based techniques.



(a) (b) (c)

Fig. 2. Effects of enhancement technique in IGE feature domain. (a) Original panchromatic image of IKONOS satellite. (b) IGV feature space clustered image after using proposed enhancement technique of (a). (c) IGV feature space clustered image of original image (a).

One is the lack of a well-defined method to select initial seed points. The other problem is that these techniques are effective for spherically shaped clusters. For elongated or more complex clusters, these methods do not always work well. To overcome this deficiency, we have used a multiseed technique-based clustering technique [20], [21].

In the proposed seed point detection technique, we have used both the information from the enhanced image and IGV feature values. First, seed points are detected using a multiseed technique of enhanced image (Algorithm A). Then, the final seed points of IGV feature values are detected by using the seed points of enhanced image (Algorithm B). These final seed points are called as variance *Seeds* (*VS*). *VS* of IGV feature values are used to cluster the IGV feature space, which is described in Section II-D. IGV feature values are real numbers and they are very close to each other. So, the between class variance in the enhanced image is higher than that in the IGV feature values. In such situations, multiseed technique will work effectively in the enhanced image than in the IGV feature space. Also, the range of IGV feature space is larger than that of the enhanced image. So, if we apply the multiseed technique directly for seed point selection in the IGV feature space, it is computationally expensive due to the fact that we have a large data volume and the seed points are very close to each other due to small variation within the IGV feature space. As a result, the clusters are not well formed and not separated enough for accurate building extraction. Therefore, we chose to select the seed points from the enhanced image instead of the IGV feature image.

Let $g_i,\ i = 0, 1, \ldots, L-1$ be the gray level value of the enhanced image h and $fr_i,\ i = 0, 1, \ldots, L-1$ be the corresponding frequency (total number of pixel) of $g_i,\ i = 0, 1, \ldots, L-1$ in the enhanced image. The seed point is defined as the *mode*, which is the gray value of highest frequency point, of the region in a particular iteration. Here, the measure of homogeneity is defined as the standard deviation (*Sd*) with respect to mode. If $Sd > T_1$ (where T_1 is the predefined threshold), then the region is a nonhomogeneous and vice versa. If the region is heterogeneous in a particular iteration, the region is split into two regions. 1) The left region whose pixel gray values are less than the mode. 2) The right region whose pixel gray values are greater than or equal to the mode. Again both the regions are recursively split if they do not satisfy the homogeneity criterion and find the modes as the seed points for the left and the right regions. Thus, the seed point detection process is applied recursively until there is no region that can be split or the region is too small to split.

Algorithm A:

Step 1) Find the minimum $g_{\min}$ and maximum $g_{\max}$ gray values of the current region (initially, the current region is the input enhanced image). Let $g_i,\ i = 0, 1, \ldots, L-1$ be the gray value of the current region and corresponding frequency (total number of pixel) $fr_i,\ i = 0, 1, \ldots, L-1$.

Step 2) Find *mode* of the current region, i.e., $m = g_i$ for which $fr_i = \max_{j=g_{\min}}^{g_{\max}} \{fr_j\}$.

Step 3) Compute standard deviation (*Sd*) with respect to mode (m) for the current region. That is,

$$Sd = \left[\frac{1}{\sum_{i=g_{\min}}^{g_{\max}} fr_i} \sum_{i=g_{\min}}^{g_{\max}} (m - g_i)^2 fr_i \right]^{1/2}.$$

Step 4) If $Sd > T_1$ (where T_1 is the predefined threshold) then go to Step 5). Otherwise go to Step 7). T_1 is called *homogeneity factor*.

Step 5) Collect the pixels whose gray values satisfy the inequality $m - v \times Sd \leq g_i \leq m + v \times Sd$, where v is a predefined threshold value (integer). The parameter v is called as *Gaussian multiplier*. Mode (m) is the seed point of the current region.

Step 6) Remove the pixel gray value which belongs to $[m - v \times Sd, m + v \times Sd]$ from the current region. Remaining pixel gray values are divided into two clusters: 1) $g_i \in [g_{\min}, m - v \times Sd)$ and 2) $g_i \in (m + v \times Sd, g_{\max}]$.

Step 7) Repeat Steps 1) to 6) for both clustered regions if either region is nonhomogeneous or large enough to split.

Let $m_1, m_2, \ldots, m_k$ be the k seed points extracted by Algorithm A. Each of these seed points $m_i, i = 1, 2, \ldots, k$ is the mode of a particular homogeneous cluster. We use these

seed points in the enhanced image to determine the seeds in the IGV feature space as shown below in Algorithm B.

Algorithm B:

1) For each seed point m_i, $i = 1, 2, \ldots, k$: find the set of all pixels of enhanced image whose gray values are m_i, i.e., $PX_i = \{(x_i, y_i) : g(x_i, y_i) = m_i\}$, $x = 1, 2, \ldots, M, y = 1, 2, \ldots, N$ and for all $i = 1, 2, \ldots, k$, where M and N are the rows and columns of the image.
2) Find the values of *IGV* at those pixels (x_i, y_i) in IGV feature space image, i.e., find $IGV(x_i, y_i)$, $i = 1, 2, \ldots, k$.
3) Calculate gray variance $GV[i] = \sum_{\forall (x_i, y_i): g(x_i, y_i) = m_i} IGV\,(x_i, y_i), \quad i = 1, 2, \ldots, k$.
4) Find variance seed $VS[i] = \frac{GV[i]}{\#PX_i}, \quad i = 1, 2, \ldots, k$, where "$\#PX_i$" is the number of points of the set PX_i.

D. Clustering Technique

If we apply multiseed clustering technique [20], [21] directly to enhanced image, we get many clustered regions and it is very difficult to identify building structures from the clustered image in unsupervised manner. To overcome this problem, we use the nearest neighbor clustering technique in IGV feature space using the detected seed points by the above seed point detection algorithms (Algorithms A and B). $ES[i], i = 1, 2, \ldots, k$ are the k seed points (cluster centers) in the IGV feature space. Each sample in the IGV feature space is grouped to the nearest cluster center. Using this approach results in the grouping of the boundary pixels of manmade structures into a single cluster distinct from the clusters made by nonmanmade features. As a result, the detection of threshold value in the binarization process is easier and gives more accurate results for manmade structures.

E. Binarization Technique

The segmented (clustered) image from the previous steps is still a gray level image. The manmade objects have natural brightness properties. In addition, the border regions of the manmade structures have higher values than its internal values. Still, it is difficult to extract the border regions of the manmade object accurately from the clustered image. To overcome this problem, we propose the used of an automatic thresholding-based binarization step.

The IGV feature space is formed into k clusters by using the nearest neighbor clustering technique described in Section II-D. In this section, our main goal is to map the points in the feature space back into the image space by binarizing the IGV clustered features. Although threshold-based binarization is simple, automatic detection of the threshold value is difficult. We propose to use a threshold on the bimodality detection [21]. It is briefly described here.

Let P be the population of points in a cluster in the IGV space. P is called a bimodal if it can be divided into two components subpopulations, say $P_{s(u)}, P_{g(u)}$ such that

1) $P_{s(u)}$ contains all the pixels with cluster value $\leq$ some u and $P_{g(u)}$ contains all the pixels with cluster value $> u$.
2) The variances of $P_{s(u)}$ and $P_{g(u)}$ are small relative to the variance of P.

Let us assume that n and σ^2 are the total frequency and variance of P, $n_{s(u)}$ and $\sigma^2_{s(u)}$ be the total frequency and variance of $P_{s(u)}$ and similarly let $n_{g(u)}$ and $\sigma^2_{g(u)}$ be the total frequency and variance of $P_{g(u)}$. Now, we will determine the gray level value u' such that the function

$$W(u) = \frac{n_{s(u)}\sigma^2_{s(u)} + n_{g(u)}\sigma^2_{g(u)}}{n\sigma^2}, \quad cl_{\min} \leq u \leq cl_{\max}$$

is minimized, where $cl_{\min}$ and $cl_{\max}$ are the minimum and maximum gray level values of P. The gray level value u' is called *bimodality parameter*.

Now, the cluster feature (cl_f) space is binarized and a binary image is generated by using the bimodality parameter u'. That is, for each pixel $(x,\ y)$,

$$V(x, y) = \begin{cases} 0, & \text{if } cl_f(x, y) < u' \\ 1, & \text{if } cl_f(x, y) \geq u'. \end{cases}$$

F. Thinning and Component Filtering Operations

Many false alarms present in the image after the binarization step can be removed by using some domain knowledge. First, a region can be removed if there are no building edges and shadows in the image. In addition, very small regions are unlikely to be buildings and hence can be removed. In both cases, thinning is an essential step since thick edges make it difficult to accurately determine the relative position of buildings and their shadows. The perimeter and length of an object are important shape factors and they are difficult to estimate accurately with an accurate representation of the object and its border.

The binary image formed after the steps described before highlights the boundaries of the manmade structures. But these boundaries are thicker. To improve the results, we use a thinning step to derive the linear/curvilinear representation of the boundaries. In our implementation, we have used a well-known thinning algorithm [22]. After thinning, many small regions remain due to noise. These small regions are removed using a simple filtering step based on the length. Thus, all thinned regions whose lengths are less than a predefined threshold (L_T) are eliminated from further processing.

G. Shadow Detection Technique

Remotely sensed images with very high-spatial resolution can clearly distinguish detailed features in the scene. In many images, shadows are also visible in the image. In high-resolution satellite images, shadows are usually cast by elevated objects such as buildings, bridges, and towers, especially in urban areas. Shadows provide geometric and semantic information in an image, including cues about the shape and relative position of objects.

Paul [23] proposed an approach to automatically detect and remove shadows in an image. Wang and Wang [24] presented

a method of shadow detection and compensation in high-resolution satellite image based on the principal component analysis and luminance-based multiscale retinex algorithm. Liu *et al.* [25] have described a SAFS-based shadow detection algorithm, which can adaptively select correct features and parameters for shadow detection through an analysis of the type of input image and the chromatic properties of the shadow samples.

In high-resolution images, shadow information has been frequently used to detect buildings. In this paper, we use shadow information to confirm that a detected object is a building. One property that provides a clue about the presence of shadows is lower luminance, which is a primary property of shadows. We use the enhanced image for the detection of shadows since they manifest as clear and darker regions in the image. Therefore, thresholding is a potentially useful method of shadow detection [24]. The proposed shadow detection technique for panchromatic satellite images is a modified version of the threshold-based technique proposed by Otsu's method [28]. Before description of the proposed shadow detection algorithm, some mathematical definitions are described below.

Definition 1: Left region R_l is the set of gray values, which are less than or equal to a threshold value T.

$R_l(T) = \{g : g \leq T \, \forall \mathrm{g} \in I\}$ where I is the set of all pixel values in the image.

Definition 2: Right region R_r is the set of gray values, which are greater than a threshold value T.

$R_r(T) = \{g : g > T \, \forall g \in I\}$ where I is the set of all pixel values in the image.

Definition 3: Left frequency $f_l(T)$: Let $R_l(T)$ be the left region for a given threshold value T and $\mathrm{g} \in R_l(T)$ be a gray value. Let the corresponding frequency of g be $h_g^l(T)$. The left frequency is denoted as $f_l(T)$ and defined by $f_l(T) = \sum_{\forall g \in R_l} h_g^l(T)$.

Definition 4: Right frequency $f_r(T)$: Let $R_r(T)$ be the right region for a given threshold value T and $\mathrm{g} \in R_r(T)$ be a gray value. Corresponding frequency of g be $h_g^r(T)$. The right frequency is denoted as $f_r(T)$and defined by $f_r(T) = \sum_{\forall g \in R_r} h_g^r(T)$.

Definition 5: Relative frequency, $f_v(T)$ is defined as the ratio of the left frequency ($f_l(T)$) to total frequency $f_v(T) = \frac{f_l(T)}{f_t(T)}$ where $f_t(T) = f_l(T) + f_r(T)$.

The two quantities $f_l(T)$ and $f_r(T)$ are used to compute the left and right (relative) frequencies. To compute them, we first find the left region and right region using Definitions 1 and 2. Then, using Definition 3, we find the gray values which are less than equal to T and then find the frequencies of all those gray values within the whole image. The sum of all these frequencies is the value of $f_l(T)$. Similarly, we can compute $f_r(T)$.

Note that, we consider relative frequency for left region of an image because we are primarily focused on shadow features that are darker and the challenge is to find the right threshold that will accurately segment them. So, in each iteration, we focus only on adjusting the threshold to refine the left region. The shadow detection procedure is given below.

1. Procedure Shadow-Detection (I:Image){
2. //Initialization
   $\gamma = 0$; //number of iterations
   $t^\gamma = 256$; //threshold vale for 8-bit data
   $f_r(t^\gamma) = 0$; //right frequency
   $f_l(t^\gamma) = M \times N$; //left frequency, M and N are rows and columns
   $f_c(t^\gamma) = M \times N$; //current frequency or total frequency
   $f_v(t^\gamma) = 1$; //relative frequency
   $R_l(t^\gamma) = I$; // I is the whole image region
3. Compute the left region mean gray value of the enhanced image (I)

$$\mu = \frac{1}{f_l(t^\gamma)} \sum_{\forall g \in R_l(t^\gamma)} g \times h_g^l(t^\gamma)$$

4. Find $R_l(\mu)$, $R_r(\mu)$, $f_l(\mu)$, $f_r(\mu)$, $f_t(\mu)$ and $f_v(\mu)$// Using Definitions 1–5
5. If $(f_v(\mu) < \varepsilon)$ // ε is the threshold value
   {
   $\gamma = \gamma + 1$;
   $t^\gamma = OtsuThreshold(R_l(t^{\gamma-1}))$;
   }
   Go to Step 7.
6. If $(f_v(\mu) \geq \varepsilon)$
   {
   $\gamma = \gamma + 1$;
   $t^\gamma = OtsuThreshold(R_l(t^{\gamma-1}))$;
   Find $R_l(t^\gamma)$, $R_r(t^\gamma)$, $f_l(t^\gamma)$, $f_r(t^\gamma)$, $f_t(t^\gamma)$ and $f_v(t^\gamma)$
   }
   Go to Step 3.
7. Segment the image (I) using the final threshold value $S_T = t^\gamma$ (Shadow detection threshold) as
   If $h(x, y) \leq S_T$ then the pixel $(x,\ y)$ is shadow pixel otherwise non-shadow.

}

Fig. 3(a) shows a QuickBird satellite gray tone image, which we use to compare our approach with other methods. The original image and other shadow detection results of different methods have been considered from Liu *et al.* [25]. Fig. 3(a) shows a QuickBird satellite grayscale image of size 674×1024 with resolution $0.61\,\mathrm{m} \times 0.61\,\mathrm{m}$ of a typical downtown scene, and it is of lower quality due to contrast that is too high and chromatic characteristics that are clearly unnatural [25]. A close observation indicates that most of the image is covered by the building structures and shadows. Fig. 3(b) shows the enhanced image and it is clear that the shadows are darker than the original image. Fig. 3(c) shows the detected shadows (bright regions) by the proposed shadow detection technique. Results using the approaches proposed in Tsai [26], Chung *et al.* [27], and Liu *et al.* [25] are shown in Fig. 3(d)–(f). Fig. 3(g) shows the shadows derived manually [25] and 3(h) shows the shadow and nonshadow binary mask in the image by the proposed algorithm. We have inserted the "ideal shadow marks" image Fig. 3(g) from Liu *et al.* [25]. The

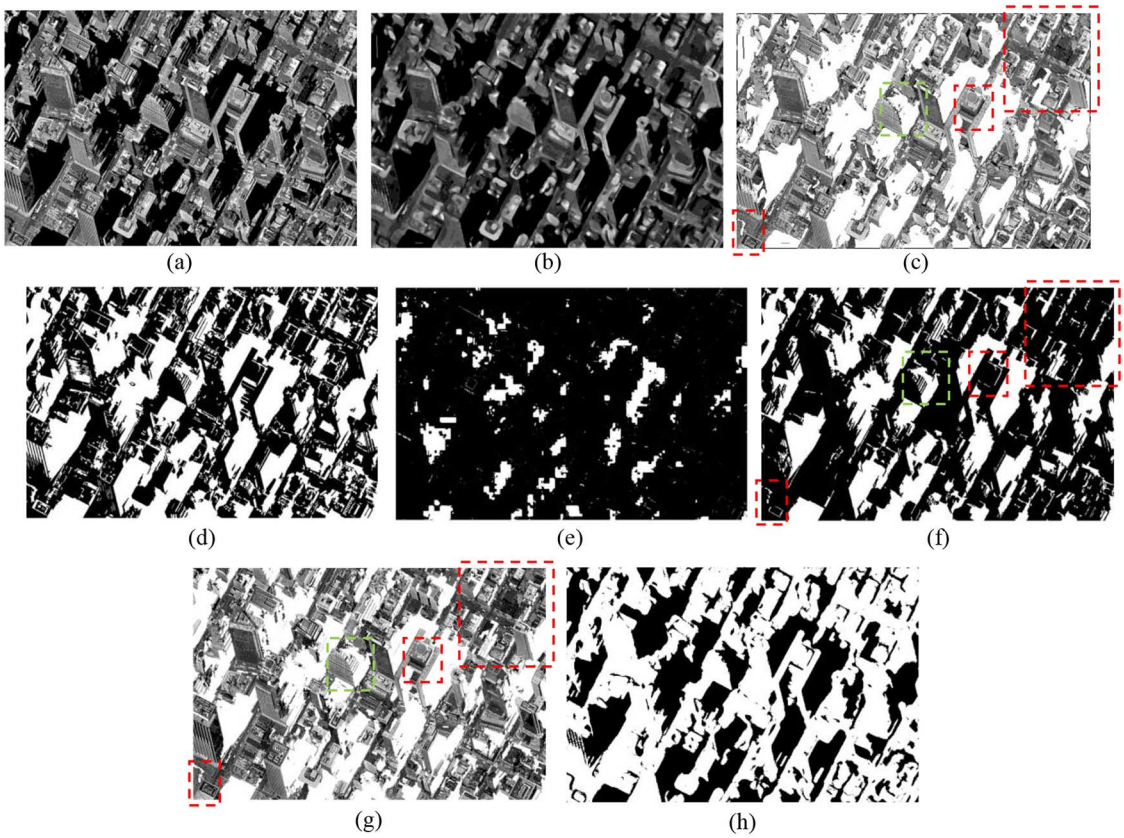


Fig. 3. QuickBird satellite grayscale image of resolution $0.61\,\mathrm{m} \times 0.61\,\mathrm{m}$. (a) Original grayscale image. (b) Enhanced image. (c) Detected shadow by the proposed shadow detection (bright region). (d) Shadows detected by Tsai's method [26]. (e) Shadows detected by algorithm by Chung *et al.* [27]. (f) Shadows detected by Liu *et al.* method [25]. (g) Ideal shadow marks. (h) Shadow and nonshadow binary mask by the proposed algorithm (shown in white).

detected results using our proposed method [Fig. 3(c)], Tsai's algorithm [26] [Fig. 3(d)] and Liu *et al.* [25] [Fig. 3(f)] are better than the results from the algorithm proposed by Chung *et al.* [27] [Fig. 3(e)] and they are similar to the manually interpreted shadow map [Fig. 3(g)]. Comparing Fig. 3(c) and (d) with Fig. 3(g), we also note that the shadow detection results from the proposed method are better than the results from Tsai's algorithm. Furthermore, a closer examination of the results show that both the proposed method and the one by Liu *et al.* [25] have some places that are detected as per ideal shadow marks while some places are not detected accurately. The shadows within red dotted regions are detected better by the proposed algorithm as per ideal shadow marks than Liu *et al.* [25] method for those areas. In contrast, the shadow within green regions is detected better by Liu *et al.* [25] method than the proposed method as per ideal shadow marks.

H. False Alarm Reduction Technique

This step is used for reduction of false alarms from the output of thinning and component filtering (Section II-F) step (edge map image) by using the positional information of building edge and shadow. The location of the shadows with respect to buildings depends on the look angle of the satellite sensor and the position of the sun. Furthermore, the shadow positions are relatively the same for all building in a particular image, i.e., the shadow position is either to the left/right/above/below/diagonal with respect to the buildings in a particular image. Relative position of shadow is extracted from the image header information like look angle of the sensor and sun position (acquire time). Using this information, we identify the objects that have a high probability of being buildings using the following steps.

1) Remove isolated shadows, i.e., if there is no edge in the edge map image in an appropriate position with respect to shadows then remove those shadows.
2) Identify at least one nonzero pixel of edge map image as the "building seed."
3) If any building seed exists then eight-neighbor connected component of the edge map image for the seed is kept as it is; otherwise, the edge component from the edge map image is deleted.

I. Segmentation Technique

When the targets and the background in the image are well contrasted, thresholding with a suitable value can result in effective segmentation. In our algorithm, we use the approach proposed by Otsu [28]. The algorithm chooses the intensity value that maximizes a measure of class separability computed as the ratio of the between class variance to the local variance. Here, we use the enhanced image for final building detection step by local adaptive thresholding-based segmentation procedure. Any local adaptive thresholding-based operation depends on the region size, i.e., the area of the image in which the

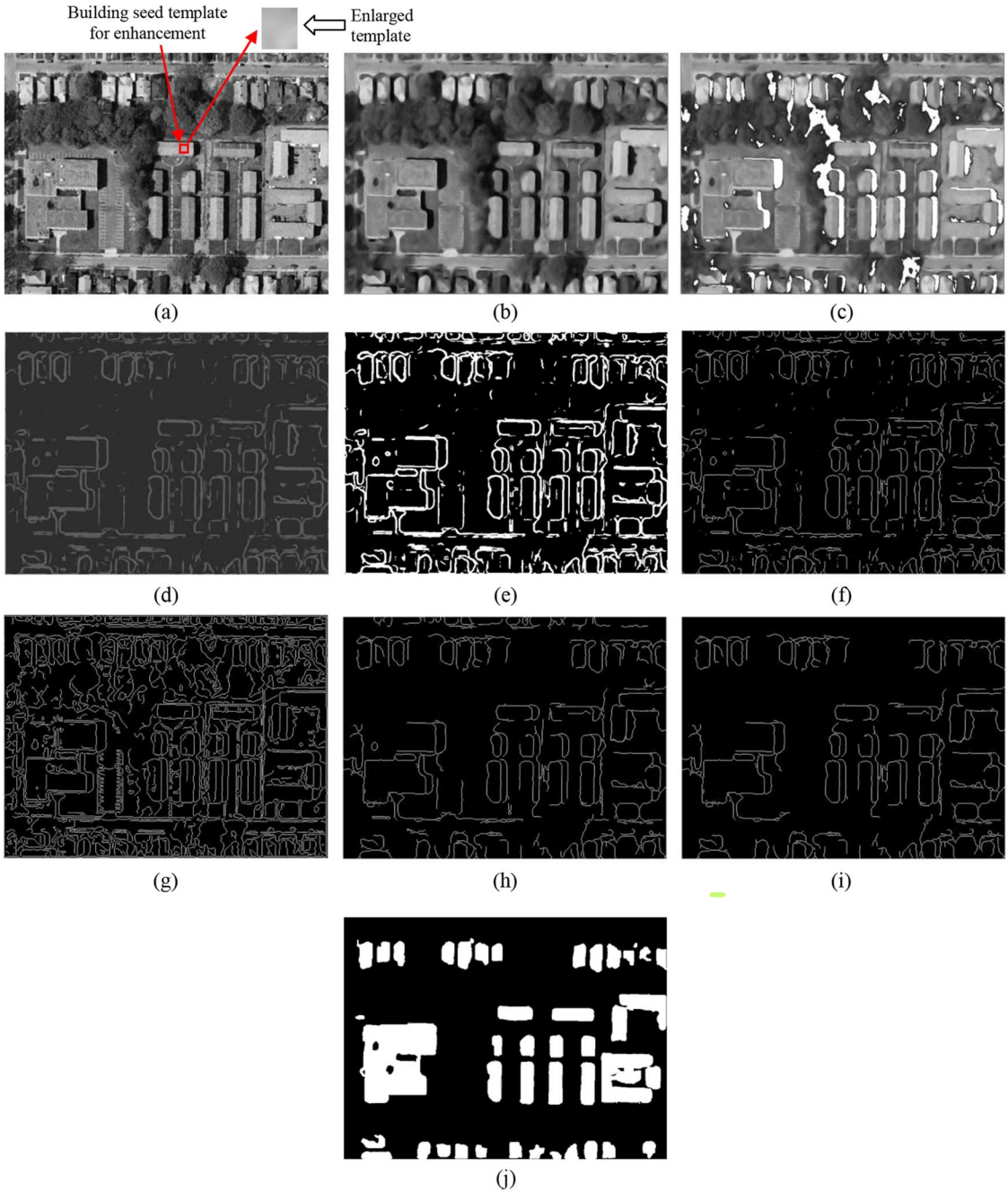


Fig. 4. IKONOS satellite panchromatic image of resolution 1 m × 1 m. (a) Original image. (b) Enhanced image. (c) Shadow overlade image. (d) Clustered image of the IGV features. (e) Binary image. (f) Thinned image. (g) Edge image by canny operator. (h) Small component filtered image. (i) Building edge image. (j) Final detected building structures.

operation will be implemented for segmentation. We use the probable building edges (i.e., the output of false alarm reduction technique module). Each connected edge region of the final (high probable) building edge image is considered as a candidate for region thresholding. The proposed adaptive threshold-based segmentation procedure using the final building edge and enhanced images is described as below.

1) First, find four points, i.e., left-most, right-most, top-most, and bottom-most points of a particular connected edge region from the final building edge image. These four points give the bounding rectangle region and this rectangle is the region of interest for segmentation.
2) Find the corresponding rectangular region of the enhanced image.
3) Apply Otsu's algorithm within the rectangular region of the enhanced image and find the threshold value (O_T).
4) Segment the rectangular region of the enhanced image as

$$g(x,y) = \begin{cases} 0, & \text{if } h(x,y) < O_T \\ 1, & \text{if } h(x,y) \geq O_T \end{cases}$$

where the pixel $(x,\ y)$ belongs to the rectangular region.

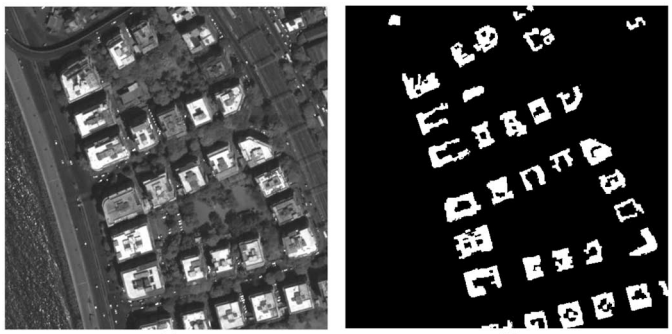
5) Repeat the above steps for all connected edge regions of the final building edge image.

The result of this step is the set of building structures extracted in the image. It should be noted that a classical mean-shift-based edge detection/segmentation method using properly set mean-shift parameters may give better results but it has many drawbacks as well. The algorithm does not work well with images in which the edges are ill defined or there are too many edges. Furthermore, it is less immune to noise than other techniques, e.g., thresholding and clustering. Finally, this procedure is time consuming, an important consideration for processing large images.

III. Experimental Results and Discussions

In this paper, we have evaluated the performance of our approach with scenes obtained from IKONOS and QuickBird panchromatic imagery. We have also compared our results with other algorithms in literature that have used IKONOS and QuickBird images. In the proposed building detection algorithm, there are three heuristic parameters: 1) homogeneity factor T_1; 2) Gaussian multiplier, l for seed point detection algorithm (Algorithm A); and 3) length threshold L_T (in Section II-F). Through extensive experimentation, we have determined that for best results the values of T_1, v, and L_T to be 2.8, 3, and 20, respectively,.

Fig. 4(a) shows an original IKONOS panchromatic image of size 800×600. The 5×5 building seed template used for enhancement and its enlarged version are also shown in Fig. 4(a). The enhanced image using the proposed enhancement technique is shown in Fig. 4(b). It is clear that the building structures in Fig. 4(b) are enhanced and smooth inside the individual building structures in comparison to the original building structures in Fig. 4(a). At the same time, nonbuilding structures (like trees) are blurred and merged with the background. As a result, it is more effective to isolate building structures using enhanced image from the background and other nonbuilding structures. The detected shadows by the proposed shadow detection technique are shown as white color overlaid on the original image [Fig. 4(c)]. Fig. 4(d) shows the clustered image by the proposed clustering technique. We have observed that the IGV feature values are low inside the structures of a building and high at the boundaries of the buildings, so most of the low IGV feature values are formed as a single cluster. Also at the same time, high IGV feature values are formed in different clusters as the variance intensity. Fig. 4(e) shows the image using the proposed binarization process and it can be seen that most of the boundaries of manmade structures are detected properly. At the same time, there are no boundaries of the nonmanmade objects are preserved. Fig. 4(f) shows the thinned image. The result shows the presence of many small isolated linear/curvilinear features. Such features are removed from the result by component filtering procedure and Fig. 4(h) shows the image after their removal. We show the edge image of the original image by using Canny edge operator [29] in Fig. 4(g). It can be seen from Fig. 4(g) that small unwanted edges of nonmanmade structures are present and hence it is difficult to isolate the building edges from the nonbuilding edges in many places. Thus, a qualitative analysis of the results shows the proposed combined algorithm extracted the building edge features better than other edge operators. Fig. 4(i) shows the final building edge image after false alarm reduction. Finally, the building structures after segmentation step are shown in Fig. 4(j).



(a) (b)

Fig. 5. QuickBird satellite panchromatic image of ground resolution $0.6\,\mathrm{m} \times 0.6\,\mathrm{m}$. (a) Original image. (b) Final detected building structures.

Results of our approach on a QuickBird image are shown in Fig. 5. The original QuickBird image of size 800×800 is shown in Fig. 5(a). The final detected building structures by the proposed algorithm are shown in Fig. 5(b).

Fig. 6(a) shows a QuickBird panchromatic image of Beijing area, acquired in 2002. The resolution of the image is $0.6\,\mathrm{m} \times 0.6\,\mathrm{m}$. The detected buildings by Liu and Prinet method [16] and the proposed method are shown in Fig. 6(b) and (c), respectively. It can be seen that a few buildings in the bottom left corner and middle of the image right (marked in red) are not detected by the algorithm proposed by Liu and Prinet. Also, a small portion of the building in the middle of the image is not detected by Liu and Prinet method. But, those buildings are clearly detected by the proposed algorithm.

Fig. 7(a) shows another QuickBird panchromatic image of Beijing area, acquired in 2002. The results from the Liu and Prinet method [16] and the proposed method are shown in Fig. 7(b) and (c), respectively. Here also, we can see that a few buildings in the bottom left corner (marked) are not detected by Liu and Prinet method, whereas those buildings are detected by the proposed algorithm.

Fig. 8 shows the comparison of our approach with that of Theng [4]. We use a QuickBird image of nonurban scene from Burlington city. Fig. 8(b) and (c) shows the detected buildings by Theng [4] and the proposed method, respectively. We can see that both the methods extracted buildings correctly.

Fig. 9(a) and (d) shows IKONOS satellite panchromatic images with 1-m resolution of size 222×222. We use both images to compare our algorithm with the algorithm proposed by Sun *et al.* [30]. Fig. 9(b) and (e) shows the detected buildings by Sun *et al.* [30] method. The detected buildings by the proposed method are shown in Fig. 9(c) and (f). Both the algorithms detected buildings almost correctly. We can see that a few buildings in the bottom left corner (marked in red) of the image [Fig. 9(d)] are not detected by Sun *et al.* [30] method. But, the proposed algorithm detects those buildings. Also Sun *et al.* [30] method depends on many heuristics parameters like L (levels), α (scale factor), λ (segmentation factor), ε (area

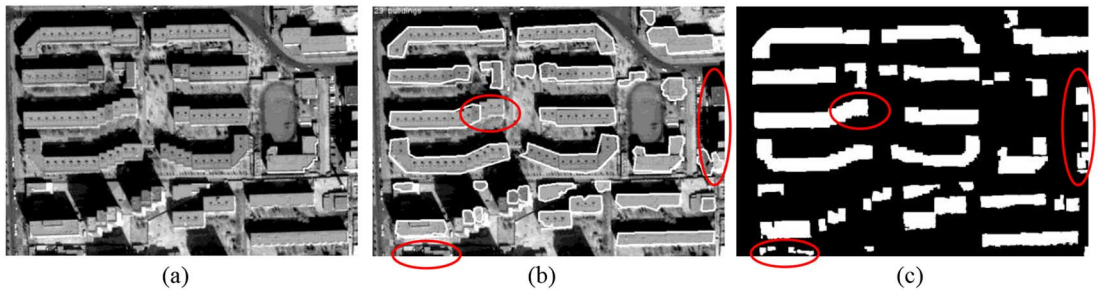


Fig. 6. QuickBird satellite PAN image. (a) Original image of size 574×408. (b) Detected building by Liu and Prinet [16] method. (c) Detected building by the proposed algorithm.

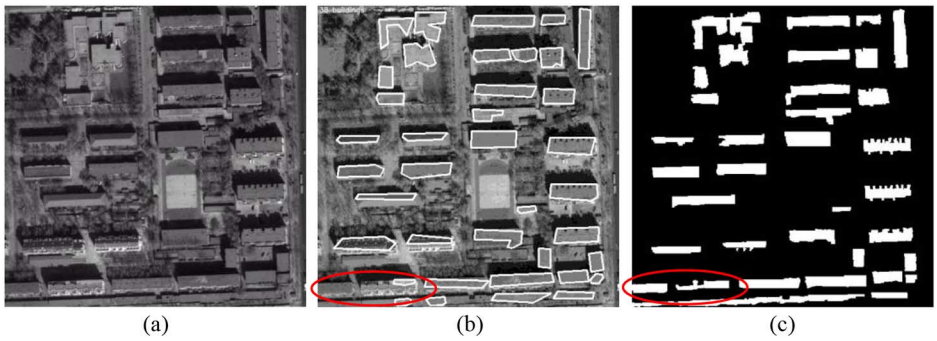


Fig. 7. QuickBird satellite PAN image. (a) Original image of size 512×512. (b) Detected building by Liu and Prinet [16] method. (c) Detected building by the proposed algorithm.

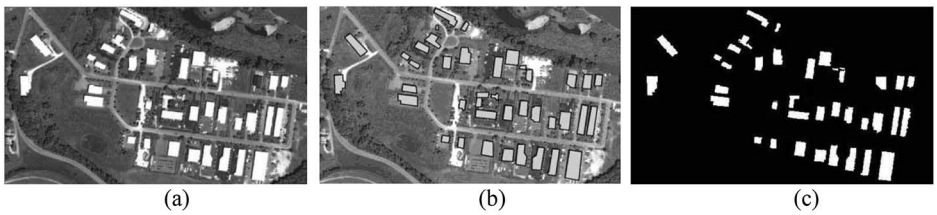


Fig. 8. QuickBird satellite PAN image. (a) Original image of size 514×300. (b) Detected building by Theng [4] method. (c) Detected building by the proposed algorithm.

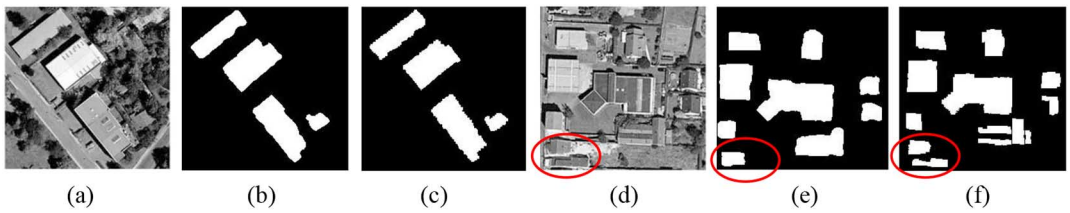


Fig. 9. IKONOS satellite PAN image. (a) Original image of size 222×222. (b) Detected building by Sun *et al.* [30] method. (c) Detected building by the proposed algorithm. (d) Original image of size 222×222. (e) Detected building by Sun *et al.* [30] method. (f) Detected building by the proposed algorithm.

threshold), σ (label threshold), CCS parameters c, and T. In comparison, the proposed method uses only three parameters.

Computational Speed: The time to compute the features is also an important criterion in selecting an algorithm. Table I summarizes the performance of our proposed algorithm and those of the others in literature. All algorithms are implemented and timed on an HP xw6400 workstation (Intel(R) Xeon(R), 5130 at 200 GHz, 2.00 GB RAM, Microsoft windows XP). Table I shows the runtimes for our proposed algorithm and the algorithms proposed by Liu and Prinet [16], Theng [4], and Sun *et al.* [30] for the different images used in this paper. Table I

TABLE I
COMPARISON OF COMPUTATION SPEED FOR DIFFERENT ALGORITHMS (IN MILLISECONDS)

Fig. no.	Proposed algorithm	Liu and Prinet algorithm [16]	Theng algorithm [4]	Sun *et al.* algorithm [30]
Fig. 6	309	387		
Fig. 7	312	395		
Fig. 8	298		378	
Fig. 9(a)	118			169
Fig. 9(b)	117			170

TABLE II
EVALUATION RESULTS OF FOUR IMAGES BY DIFFERENT METHODS

Image	Proposed method					Liu and Prinet method [16]					Sun *et al.* method [30]				
	TP	FP	TN	DP (%)	BF (%)	TP	FP	TN	DP (%)	BF (%)	TP	FP	TN	DP (%)	BF (%)
Fig. 6	35	2	4	89.7	5.4	29	2	10	74.4	6.5					
Fig. 7	36	1	5	87.8	2.7	34	4	6	85	10.5					
Fig. 9(a)	4	1	2	66.7	20						4	1	2	66.7	20
Fig. 9(b)	13	1	1	92.8	7.1						9	1	4	64.3	10
Mean	DP = 84.25% BF = 8.8% (overall)					DP = 79.7% BF = 8.5%					DP = 65.5% BF = 15%				

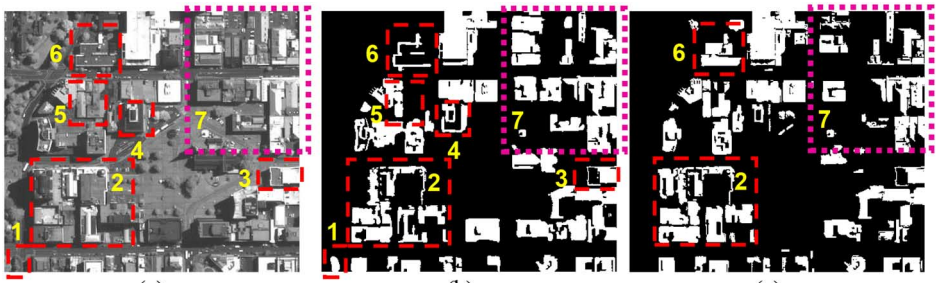


Fig. 10. SkySat-1 satellite panchromatic image. (a) Original image of size 404×358. (b) Detected building by the proposed algorithm. (c) Detected building by Sun *et al.* [30] method.

shows that our proposed algorithm is significantly faster than others proposed in literature.

Finally, we compare the performance of our algorithm and others in literature using metrics proposed in Lin and Nevatia [31]. Here, we compare the results with the actual ground truth derived manually. The two metrics used in our research are

$$\text{Detection percentage } (DP) = \frac{100 \times TP}{TP + TN}$$

$$\text{Branch Factor } (BF) = \frac{100 \times FP}{TP + FP}$$

where *TP* (true positive) is a building detected by both a person and the algorithm, *FP* (false positive) is a building detected by the automatic approach but not a person, and *TN* (true negative) is a building detected by a person but not by the algorithm.

A building is rated as detected, if at least a small part of it is detected by the automatic approach. The two metrics are calculated by comparing the ground truth (building information) and the buildings detected by the algorithms. *DP* describes how many of the existing buildings in the scene are found by the automatic approach and the *BF* indicates how many buildings are found erroneously. The *DP* is 100% if the whole image is classified as class building; but in that case, *BF* would also be very large. The goal is to maximize the *DP* while keeping the *BF* low.

Table II summarizes the performance of the algorithms on a set of four images of different sizes. The number of buildings of images in Figs. 6, 7, 9(a) and (b) are 39, 41, 6, 14, respectively. We have compared the performance of our algorithm with those of Liu and Prinet [16] and Sun *et al.* [30]. Table II also presents the mean values for the metrics over all images. We can see that the detection percentage for the proposed algorithm is higher than the other methods. Simultaneously, the branch factor is smaller than or comparable to the other methods.

A. Limitation

Overall, our algorithm outperforms other algorithms in literature in the dataset we examined. It should be noted that our segmentation approach is, at the core, threshold-based and relies on the homogeneity of the building tops. If the top of the building is partially bright and partially dark, then this technique will detect only a part of the building. In general, the algorithm fails to accurately detect buildings with mixed (shade or texture) rooftops. This is due to the drawback of the proposed segmentation algorithm, which is threshold-based within a specific region of interest. An advanced segmentation algorithm based on the structural, contextual, and spectral information within the region of interest will be able to address this challenge. Furthermore, our approach relies on the shadows of the buildings to accurately location the buildings. In nadir-view images and urban images with low-rise buildings, shadows may not be clear or may not exist. In such situations, the buildings are not accurately detected.

An example of the failure of our algorithm can be seen in Fig. 5, where the rooftops of the buildings have different shades and texture. As expected, the proposed technique is unable to fully detect the buildings.

Fig. 10(a) shows an original SkySat-1 satellite panchromatic image of size 404×358 with resolution of $85\,\text{cm} \times 85\,\text{cm}$. The detected buildings by the proposed method and using the

method by Sun *et al.* [30] method are shown in Fig. 10(b) and (c), respectively. A close look at the results show that both methods show mixed success. Sevaral buildings marked in red (#2, #3, #4, and #6) in Fig. 10(a) are not detected accurately by the proposed method [Fig. 10(b)] due to different shades and textures on the rooftops of those regions. Some of these buildings within Region #2 are detected more accurately by Sun *et al.* [30]. However, the method fails to detect some complete buildings within Regions #2 and #7. The proposed algorithm detects buildings better than Sun *et al.* [30] within Region #7.

It can be seen that there are no shadows for some buildings in Region #5 (on the right) and our algorithm fails to detect these buildings. The object within Region #1 (bottom left corner) is a tree but the proposed algorithm detects it as a building because its features like texture and shadow information of that object are similar to building structures. Use of structural information for building detection will be able to overcome such problems.

IV. Summary

Satellite imagery has significant information and form an important source of knowledge in many applications. Therefore, extraction of salient features from satellite imagery is a critical task. However, isolating and identifying features from such imagery is complex and the many difficult challenges need to be addressed before the features can be efficiently and accurately identified. Buildings form an important class of objects in satellite imagery that carry important information about human activities. Therefore, their efficient extraction will have many significant applications. There are many civilian, commercial, and military applications of such information, including the determination of existing of buildings after a natural or manmade disaster.

State-of-the art in automated extraction of buildings from satellite imagery is in its early stages. Two of the most difficult problems in developing accurate algorithms for this task include low signal-to-noise ratio and weak object signal in the images. Most of the algorithms in literature are not fully automated. In this paper, we have presented a fully automated method to accurately extract buildings from high-resolution panchromatic remotely sensed imagery. The approach exploits both the spectral and spatial properties of buildings using a multistep approach. The main steps in our algorithm are: manmade structure enhancement, IGV feature extraction, seed point detection using multiseed technique, multiseed-based clustering technique using IGV feature, shadow detection, false alarm reduction using positional information of both building edge, and shadow and adaptive threshold based segmentation. The proposed algorithm was extensively evaluated using a variety of images from IKONOS and QuickBird satellites and compared with existing algorithms. The results demonstrate that the algorithm is both more accurate and efficient in comparison to other algorithms in literature.

References

[1] D. Chaudhuri and A. Samal, "An automatic bridge detection technique for multi-spectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 9, pp. 2720–2727, Sep. 2008.

[2] N. Lomenie, J. Barbeau, and R. Trias-Sanz, "Integrating texture and geometric information for an automatic bridge detection systems," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Toulouse, France, Jul. 2003, pp. 3952–3954.

[3] K. Khoshelham, C. Nardinocchi, E. Frontoni, A. Mancini, and P. Zingaretti, "Performance evaluation of automated approaches to building detection in multi-source aerial data," *ISPRS J. Photogramm. Remote Sens.*, vol. 65, pp. 123–133, 2010.

[4] L. B. Theng, "Semi-automatic building extraction from satellite imagery," *J. Eng. Lett.*, vol. 13, no. 3, pp. EL_13_3_5, 2006.

[5] D. S. Lee, J. Shan, and J. S. Bethel, "Class guided building extraction from IKONOS imagery," *Photogramm. Eng. Remote Sens.*, vol. 69, no. 2, pp. 143–150, 2003.

[6] F. Rottensteiner, J. Trinder, S. Clode, and K. Kubik, "Using the Dempster-Shater method for the fusion of LIDAR data and multi-spectral images for building detection," *Inform. Fusion*, vol. 6, no. 4, pp. 283–300, 2005.

[7] R. O. C. Tse, C. M. Gold, and D. Kidner, "A new approach to urban modeling based on LIDAR," in *Proc. Winter Sch. Comput. Sci. 14th Int. Cof. Cent. Eur. Comput. Graphics Visual. Comput. Vision*, Plzen, Czech Republic, Jan./Feb. 2006, pp. 279–286.

[8] F. Tarsha-Kurdi, T. Landes, P. Grussenmever, and E. Smigiel, "New approach for automatic detection of buildings in airborne laser scanner data using first echo only," in *Proc. ISPRS Symp. Photogramm. Comput. Vision*, Bonn, Germany, Sep. 2006, pp. 1–6.

[9] S. Muller and D. W. Zaum, "Robust building detection in aerial images," in *Proc. Joint Workshop ISPRS Ger. Assoc. Pattern Recog. Object Extr. 3-D City Models Road Databases Traffic Monit. Concepts Algorithms Eval.*, Vienna, Austria, 2005, vol. 36 (3/W 24), pp. 143–148.

[10] Y. Song and J. Shan, "Building extraction from high-resolution color imagery based on edge flow driven active contour and JSEG," in *Proc. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Beijing, China, 2008, vol. XXXVII, pp. 185–190, Part B 3a.

[11] J. Peng, D. Zhang, and Y. Liu, "An improved snake model for building detection from urban aerial images," *Pattern Recog. Lett.*, vol. 26, no. 5, pp. 587–595, 2005.

[12] S. D. Mayunga, Y. Zhang, and D. J. Coleman, "Semi-automatic building extraction utilizing QuickBird imagery," *IAPRS*, vol. XXXVI, pp. 29–30, 2005, Part-3/W24.

[13] S. Ahmady, H. Ebadi, M. J. V. Zouj, and H. A. Moghaddam, "Automatic building extraction from high-resolution aerial images using active contour model," in *Proc. Int. Archi. Photogramm. Remote Sens. Spatial Inf. Sci.*, Beijing, China, 2008, vol. XXXVII, pp. 453–456, Part-B3b.

[14] J. A. Benediktsson, M. Pesaresi, and K. Arnason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1940–1949, Sep. 2003.

[15] K. Segl and H. Kaufmann, "Detection of small objects from high-resolution panchromatic satellite images based on supervised image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 9, pp. 2080–2083, Sep. 2001.

[16] W. Liu and V. Prinet, "Building detection from high-resolution satellite image using probability model," in *Proc. IEEE Geosci. Remote Sens. Symp. (IGARSS)*, 2005, vol. 6, pp. 3888–3891.

[17] Y. Wei, Z. Zhao, and J. Song, "Urban building extraction from high-resolution satellite panchromatic image using clustering and edge detection," in *Proc. IEEE Geosci. Remote Sens. Symp. (IGARSS)*, 2004, vol. 7, pp. 2008–2010.

[18] D. Chaudhuri, N. K. Kushwaha, and A. Samal, "Semi-automated road detection from high-resolution satellite images by directional morphological enhancement and segmentation techniques," in *Proc. IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 5, pp. 1538–1544, Oct. 2012.

[19] R. M. Haralick, S. R. Sternberg, and X. Zhuang, "Image analysis using mathematical morphology," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 9, no. 4, pp. 532–550, Jul. 1987.

[20] D. Chaudhuri and B. B. Chaudhuri, "A novel multi-seed non-hierarchical data clustering technique," *IEEE Trans. Syst. Man Cybern., B*, vol. 27, no. 5, pp. 871–877, Sep. 1997.

[21] D. Chaudhuri and A. Agrawal, "Split-and-merge procedure for image segmentation using bimodality detection approach," *Defence Sci. J.*, vol. 60, no. 3, pp. 290–301, 2010.

[22] A. Rosenfeld and A. Kak, *Digital Picture Processing*. New York, NY, USA: Academic, 1982.

[23] D. M. Paul, "Shadow analysis in high-resolution satellite imagery of urban area," *Photogramm. Eng. Remote Sens.*, vol. 71, no. 2, pp. 169–177, 2005.

[24] S. Wang and Y. Wang, "Shadow detection and compensation in high-resolution satellite images based on retinex," in *Proc. 5th Int. Conf. Image Graph.*, 2009, pp. 209–212.
[25] J. Liu, T. Fang, and D. Li, "Shadow detection in remotely sensed images based on self-adaptive feature selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 5092–5103, Dec. 2011.
[26] V. J. D. Tsai, "A comparative study on shadow compensation of color aerial images in invariant color models," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1661–1671, Jun. 2006.
[27] K. L. Chung, Y. R. Lin, and Y. H. Huang, "Efficient shadow detection of color aerial images based on successive thresholding scheme," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 2, pp. 671–682, Feb. 2009.
[28] N. Otsu, "A threshold selection method from gray level histograms," *IEEE Trans. Syst. Man Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.
[29] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 6, pp. 679–698, Nov. 1986.
[30] X. Sun *et al.*, "Contextual models for automatic building extraction in high resolution remote sensing image using object based boosting method," in *Proc. Int. IEEE Geosci. Remote Sens. Symp. (IGARSS)*, Boston, MA, USA, 2008, pp. II-437–II-440.
[31] C. Lin and R. Nevatia, "Building detection and description from a single intensity image," *Comput. Vision Image Understanding*, vol. 72, no. 2, pp. 101–121, 1998.

D. Chaudhuri (M'11–SM'14) received the M.Sc. Degree in applied mathematics from Jadavpur University, Kolkata, India in 1985 and the Ph.D. degree in image processing and pattern recognition from Indian statistical Institute, Kolkata, India, in 1994.

He is currently a Senior Scientist and DGM with the DRDO Integration Centre, Panagarh, India. He was visited many other countries as a Visiting Scientist or Visiting Professor. He has authored over 55 papers in international journals and conferences. His research interests include image processing, pattern recognition, computer vision, remote sensing, and target detection from various sensors images.

Dr. Chaudhuri is a reviewer and Associate Editor of many international journals. He is a senior member of IEEE and fellow of IETE.

N. K. Kushwaha received the B.Tech degree in computer science and engineering from Indian Institute of Technology, Kanpur, India, in 2009.

Currently, he is a Scientist with Defence Electronics Applications Laboratory, Dehradun, India. His research interests include image processing, pattern recognition, computer vision, and remote sensing.

A. Samal (S'85–M'88–SM'10) received the B.Tech degree in computer science from Indian Institute of Technology, Kanpur, India, in 1983 and the Ph.D. degree in computer science from University of Utah, Salt Lake City, UT, USA, in 1988.

Since 1988, he has been with the Department of Computer Science and Engineering with the University of Nebraska-Lincoln, Lincoln, NE, USA, where he is currently a Professor. He has published over 70 papers in these areas in international journals and conferences. His research interests include image understanding, document analysis, geospatial computing, and distributed computation.

R. C. Agarwal received the B.Tech in electronics and communication engineering from Indian Institute of Technology, Roorkee, India, and M.Tech degree in communication and radar from Indian Institute of Technology, Delhi, India, in 1985.

Currently, he is a Director of Defence Electronics Applications Laboratory, Dehradun, India.