

Automated Building Extraction from High-Resolution Satellite Imagery in Urban Areas Using Structural, Contextual, and Spectral Information

Xiaoying Jin

*Department of Electrical and Computer Engineering, University of Missouri-Columbia, Columbia, MO 65211, USA
Email: xje4e@mizzou.edu*

Curt H. Davis

*Department of Electrical and Computer Engineering, University of Missouri-Columbia, Columbia, MO 65211, USA
Email: davisch@missouri.edu*

Received 1 January 2004; Revised 17 August 2004

High-resolution satellite imagery provides an important new data source for building extraction. We demonstrate an integrated strategy for identifying buildings in 1-meter resolution satellite imagery of urban areas. Buildings are extracted using structural, contextual, and spectral information. First, a series of geodesic opening and closing operations are used to build a differential morphological profile (DMP) that provides image structural information. Building hypotheses are generated and verified through shape analysis applied to the DMP. Second, shadows are extracted using the DMP to provide reliable contextual information to hypothesize position and size of adjacent buildings. Seed building rectangles are verified and grown on a finely segmented image. Next, bright buildings are extracted using spectral information. The extraction results from the different information sources are combined after independent extraction. Performance evaluation of the building extraction on an urban test site using IKONOS satellite imagery of the City of Columbia, Missouri, is reported. With the combination of structural, contextual, and spectral information, 72.7% of the building areas are extracted with a quality percentage 58.8%.

Keywords and phrases: building extraction, high-resolution satellite imagery, mathematical morphology, shadow, hypothesis and verification, information fusion.

1. INTRODUCTION

Monocular building extraction has been an active research topic in photogrammetry and computer vision for many years. Some useful applications are automation in cartographic mapping and updating of geographic information system (GIS) databases. Early research on building extraction was often done using aerial imagery due to its high spatial resolution of 1 meter or less. A wide range of techniques and algorithms have been proposed for automatically constructing 2D or 3D building models from aerial imagery. Comprehensive surveys of research in this area can be found in [1, 2, 3]. Considering both radiometry and geometry, a large population of these algorithms are edge-based techniques [4, 5, 6] that consist of linear feature detection, grouping for parallelogram structure hypotheses extraction, and

building polygons verification using knowledge such as geometric structure [5, 6], shadow [5, 7], illuminating angles [5], and so forth. In order to cope with the high complexity of real scenes, integrating the power of multiple algorithms, cues, and available data sources is needed to improve the reliability and robustness of the extraction results [1, 8].

The recent availability of commercial high-resolution satellite imaging sensors such as IKONOS and QuickBird provide a new data source for building extraction. The high spatial resolution of the imagery reveals very fine details in urban areas and greatly facilitates the classification and extraction of urban-related features such as roads [9, 10, 11, 12] and buildings [12, 13, 14, 15, 16, 17, 18, 19]. Launched in September 1999, IKONOS was the first commercial high-resolution satellite. IKONOS collects 1-m panchromatic (PAN) and 4-m multispectral (MS) imagery. With its high geometric accuracy and spatial resolution, it is possible to identify fine-scale features such as individual roads and buildings in the urban environment and also provide very accurate geodetic coordinates.

Since manual extraction of buildings from imagery is very time-consuming, automated methods have the potential to improve the speed and utility for cartographic mapping and are therefore highly desirable.

Given the recent availability of the commercial high-resolution satellite imagery, only a few methods for building detection/extraction from 1-meter resolution imagery have been developed. The effect of resolution on the building extraction was reported in [14, 17, 18]. The following difficulties commonly arose when generating building hypotheses from 1-meter imagery.

- (1) Low signal-to-noise ratio disturbs the extraction of low-level geometric primitives such as edges. Therefore, the minimum cue density required for medium-level perceptual grouping cannot always be obtained. Some edges are broken and cannot form reliable cues for building extraction.
- (2) Compared to submeter spatial resolution aerial imagery, 1-meter satellite imagery has weaker object resolving power because the same object is represented with relatively fewer pixels. In addition, higher object density in image space makes it more difficult to separate a single object from surrounding ones and pixel mixing becomes more serious.
- (3) As reported in [17], 1-meter high-resolution satellite imagery also leads to certain interpretation restrictions. About 15% of the building areas measured in aerial images could not be adequately modeled in the satellite imagery.

To address resolution limitations, several systems have been developed to detect buildings in high-resolution satellite imagery where the use of perceptual cues is minimized. Park et al. [13] used a rectangular building model to search and find missing lines using information from detected lines. They created pairs of antiparallel lines over the roof using a line-rolling algorithm. Sohn and Dowman [14] used a local Fourier analysis to analyze the dominant orientation angle in a building cluster. A building unit shape (BUS) space was generated by recursive partitioning of regions using a hyperline in 2D image space. A seeded BUS then searches for its neighbors and is grown when predefined homogeneous criterion are satisfied. Lee et al. [15] applied classification to multispectral IKONOS imagery to provide approximate position and shape for candidate building objects. Fine extraction was then carried out in the corresponding panchromatic image through ISODATA segmentation and squaring based on the Hough transform. Segl and Kaufmann [16] 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 thresholds within a certain range. The buildings were classified by shape matching with a model database. Objects with the number of correct shape classifications higher than an optimal threshold were detected as buildings. Shackelford and Davis [19] used a pixel-based hierarchical classification to develop a preliminary estimate of potential buildings as well as other impervious surfaces. Buildings were then cat-

egorized as a distinct object class using a fuzzy logic analysis of a segmented image that incorporated spectral, spatial, and contextual (e.g., shadow) information. Benediktsson et al. [12] used mathematical morphological operations to extract structural information from the image. Features generated by a differential morphological profile (DMP) were selected by discriminant analysis and decision boundary feature extraction. Buildings and other land use categories were then classified using a neural network.

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 [15, 19], or they use training sets or a model database to classify or match the buildings [12, 16]. Thus, these approaches 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. First, a series of geodesic opening and closing operations of different sizes are used to build a differential morphological profile (DMP) to provide image structural information. Building hypotheses are generated and verified through shape analysis on the DMP. Second, shadows extracted from the DMP provide reliable contextual information to hypothesize building position and size. Seed building rectangles are then verified and grown on a finely segmented image. Third, bright buildings are extracted using spectral information. The final building-extraction results are then obtained by combining the extraction results from the three information sources. Among these, the shadow information is a part of the scene model that provides an information source independent of the properties of building object itself. The integrated building-extraction strategy is tested on an urban area using IKONOS imagery of the City of Columbia, Missouri. Performance evaluations of the different extraction combinations from multiple sources are reported and analyzed. With the integration of structural, contextual, and spectral information, the detection percentage and quality of the building extraction are greatly improved.

2. METHODS

In this paper, IKONOS satellite imagery is used to test our integrated building-extraction strategy. High-resolution IKONOS satellite images consist of 1-m panchromatic (PAN) and 4-m multispectral (MS) bands. Both the PAN and MS data have 11-bit information content. The 4-m MS data contain four individual bands: red (R), green (G), blue (B), and near infrared (NIR). To exploit the high spatial resolution of the PAN data and high spectral resolution of the MS data, the PAN data were fused with the MS data using a color normalization method [20] implemented in a commercial software package (ENVI 3.5) to generate a four-band pan-sharpened multispectral (PS-MS) image with 1-m resolution.

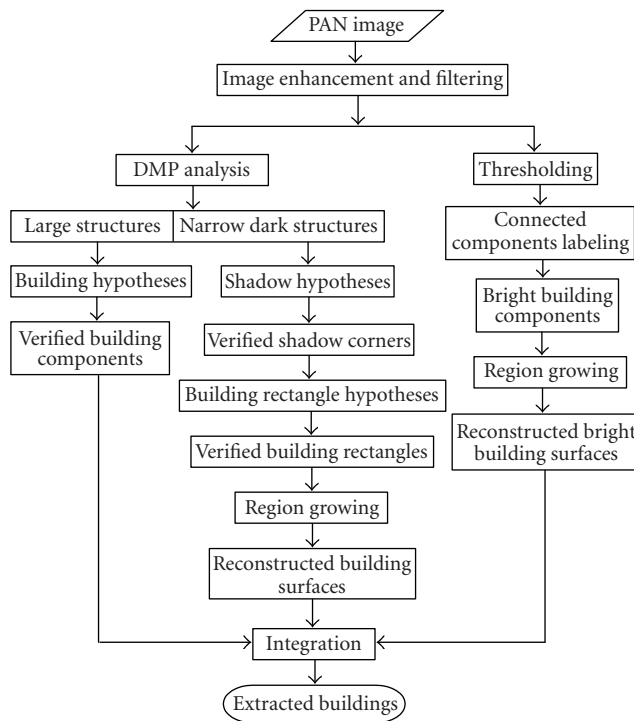


FIGURE 1: Flowchart of the integrated multidetector building-extraction strategy.

Here we concentrate on urban areas because of the high density and regularity of the buildings in these areas. In urban areas, roads are often characterized by a series of parallel and orthogonal straight lines grouped to form a grid structure [9, 10, 21, 22]. Buildings are modeled as mostly rectangular and homogeneous objects with their sides parallel or perpendicular to the road grid. The buildings are extracted by integrating structural, contextual (shadow), and spectral information. A flowchart illustrating the integrated multidetector approach is shown in Figure 1.

In the proposed integrated building-extraction strategy, a multiscale DMP is used extensively for both building and shadow hypotheses. Three building detectors are applied to a preprocessed PAN image. Two of the detectors are based on DMP analysis of the preprocessed PAN image. The first detector is mainly based on structural information of the building itself, where buildings hypotheses of relatively large scale are generated from the DMP. Then the hypothesized building components are verified through shape information of the components. The second detector is primarily based on contextual (shadow) information of the buildings. Shadow hypotheses are generated from narrow dark structures identified in the DMP. Shadow components are verified using spectral characteristics and image collection geometry, and then shadow corners are generated by projection analysis. The enclosed rectangles of shadow corners are used as building hypotheses and verified using spectral analysis of each rectangle individually. Seed building rectangles are then grown on a finely segmented image. The third build-

ing detector is primarily based on the spectral information of building itself. The purpose is to extract bright buildings, especially small ones, that are ignored by the other two detectors. After thresholding the preprocessed PAN image, the bright building components are labeled and grown to reconstruct the complete building surfaces. After independent extraction, the results from three detectors are integrated to generate the final solution. The three individual detectors operate on the input PAN image while the PS-MS image is used primarily for building and shadow verification based on spectral information. Also, a watershed segmentation algorithm is employed on the preprocessed PS-MS image to generate a finely segmented image used in the region-growing step. A detailed description of each step is provided in the following sections.

2.1. Preprocessing

Raw IKONOS satellite imagery typically have a low local contrast due to a wide radiometric dynamic range of the scene content and possible atmospheric disturbances. A linear stretch with a 2% clip on both ends of the data is used to enhance the image contrast.

From empirical observation, cartographic features such as roads and buildings have a certain range of scale. The width of most road segments in the 1-m IKONOS imagery is usually between 8–30 m [10]. And most buildings in the urban areas have a length of 10–100 m and a width of 5–50 m. A morphological opening operation by reconstruction followed by a closing by reconstruction [23, 24] is applied to each channel of the PS-MS and PAN image data to smooth out small disturbances such as cars on roads and chimneys on buildings. The structuring element (SE) was chosen to be a disc with radius $r = 2$. With an SE at this scale, roads and buildings will not be adversely affected. After morphological smoothing, a median filter with a 5×5 kernel is then used to further smooth the spectral response within the local neighborhood.

Next, an edge-based watershed segmentation method [25] is used to separate image content into different homogeneous regions. The edge information from the segmentation is exploited later in building growth from the segmented image. In this segmentation approach, the Sobel edge operator is first utilized on *each* channel of PS-MS image. The edge magnitudes for each channel are then combined by the “MAX” operation to obtain a single edge magnitude for each pixel. The watershed segmentation algorithm works to detect catchment basins as regions and crest lines as boundaries for these regions. Over-segmentation is a well-known phenomenon in watershed segmentation. One solution is to modify the image to remove regional minima that are too shallow. Here we ignore edges with a magnitude less than a chosen threshold in the watershed segmentation so that the number of segments is as small as possible while still retaining the edges of most buildings in the image. This process generates a finely segmented image. A subset of the IKONOS image in the urban area of the City of Columbia, Missouri, and its watershed segmentation are shown in Figure 2.



FIGURE 2: (a) A subset of PS-MS IKONOS satellite image in dense urban area of the City of Columbia, Missouri (only R, G, and B channels are shown). (b) Watershed segmentation result.

The two primary perpendicular directions of the road network can be detected using a spatial signature weighted Hough transform (SSWHT) [10]. This was done and the image was rotated by an angle less than 45° so that the primary directions of roads and buildings are horizontal and vertical in the image space. This is important for the shadow-supported building extraction described in Section 2.4. For the urban area of the IKONOS image used in this study, the directions of roads are nearly horizontal and vertical, so no rotation was needed.

2.2. The differential morphological profile

Mathematical morphology has been applied to a wide variety of practical problems such as noise filtering, image segmentation, shape detection and decomposition, and pattern recognition, to name but a few [12, 23, 24, 25, 26, 27]. Mathematical morphology differs from many other image processing techniques because it is a nonlinear approach, usually dealing with discrete data in terms of sets and set operations.

The *morphological profile* and the *differential morphological profile* (DMP) are new concepts first introduced by Pesaresi and Benediktsson in 2001 [26]. Both are based on the use of opening and closing by reconstruction with different structuring element (SE) sizes. Here we briefly review these concepts.

Let $\gamma_{\lambda_i}^*$ be a morphological opening operator by reconstruction using structuring element $SE = \lambda_i$. λ_0 is the SE with only one element, and the size of λ_i increases with increasing $i \in [0, n]$, where n is the total number of iterations. The *opening profile* $\Pi\gamma(x)$ at the point x of the image I is defined as a vector

$$\Pi\gamma(x) = \{\Pi\gamma_i : \Pi\gamma_i = \gamma_{\lambda_i}^*(x), \forall i \in [0, n]\}. \quad (1)$$

Also, let $\phi_{\lambda_i}^*$ be a morphological closing operator by reconstruction using structuring element $SE = \lambda_i$. The *closing pro-*

file $\Pi\phi(x)$ at the point x of the image I is defined as the vector

$$\Pi\phi(x) = \{\Pi\phi_i : \Pi\phi_i = \phi_{\lambda_i}^*(x), \forall i \in [0, n]\}. \quad (2)$$

In the above, $\Pi\gamma_0(x) = \Pi\phi_0(x) = I(x)$ by the definition of opening and closing by reconstruction.

The *derivative of the opening profile* $\Delta\gamma(x)$ is defined by the vector

$$\Delta\gamma(x) = \{\Delta\gamma_i : \Delta\gamma_i = |\Pi\gamma_i - \Pi\gamma_{i-1}|, \forall i \in [1, n]\}. \quad (3)$$

By duality, the *derivative of the closing profile* $\Delta\phi(x)$ is defined by the vector

$$\Delta\phi(x) = \{\Delta\phi_i : \Delta\phi_i = |\Pi\phi_i - \Pi\phi_{i-1}|, \forall i \in [1, n]\}. \quad (4)$$

In general, the differential morphological profile (DMP) $\Delta(x)$ can be written as the vector

$$\Delta(x) = \left\{ \Delta_c : \begin{array}{l} \Delta_c = \Delta\phi_{n-c+1}, \forall c \in [1, n] \\ \Delta_c = \Delta\gamma_{c-n}, \forall c \in [n+1, 2n] \end{array} \right\} \quad (5)$$

with $c = 1, \dots, 2n$. The response for the derivative calculated using small SEs is near the central position of the DMP vector, while the response for the greatest SEs in the closing and opening profile are recorded at the beginning ($c = 1$) and at the end ($c = 2n$), respectively. The signal recorded in the DMP gives information about the size and the type of the structures in the image. Small structures will have high response near the center of the DMP while large structures will have high response near the two ends of the DMP. Structures darker than the surrounding background will have high response near the beginning of the DMP while brighter structures will have high response near the end of the DMP.

By observing the position of the greatest response in the DMP, Pesaresi and Benediktsson [26] defined the morphological multiscale characteristic and used it for image segmentation of high-resolution satellite imagery. In [12], Benediktsson et al. used the DMP and panchromatic intensity value to form a feature vector for each pixel in the image. Then a neural network was employed on the reduced feature vector to classify the pixels into six information classes using training sets. To date, the DMP has not been applied for automated feature extraction research.

2.3. Building hypothesis and verification by DMP

In urban areas, buildings typically have a length of 10–100 m and a width of 5–50 m. Buildings seldom have a length longer than the distance of a typical city block. Buildings may be made of different materials, such as bitumen, concrete, metal, synthetic materials, tiles, and so forth. The spectrum of buildings can have significant overlap with parking lot and road surfaces since they may be made of the same materials. Therefore, structural information provides a complementary way to discriminate buildings from other land cover types in addition to the spectral signature of individual pixels. Generally, buildings in most urban scenes will have a wide variety

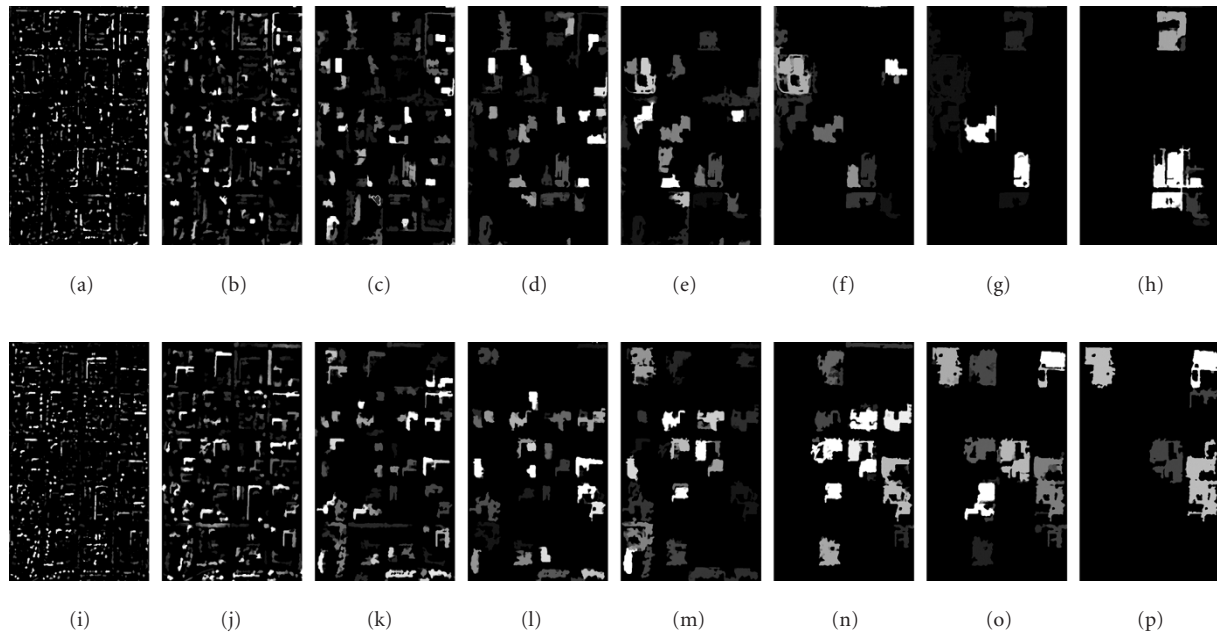


FIGURE 3: Structural decomposition of the image in Figure 2a using the differential morphological profile. The images have been visually enhanced. The derivative has been calculated relative to a series generated by 8 iterations of the SE with radius from 3–24 m. The upper plots show the derivative of the opening profile with $r =$ (a) 3, (b) 6, (c) 9, (d) 12, (e) 15, (f) 18, (g) 21, (h) 24. The lower plots show the derivative of the closing profile with $r =$ (i) 3, (j) 6, (k) 9, (l) 12, (m) 15, (n) 18, (o) 21, (p) 24.

of sizes, so an SE with a fixed size cannot aid in the discrimination of buildings of variable size. Hence, the DMP with SEs of variable size is used here to extract structural information from the image.

Considering the scale of buildings in the image, a 16D DMP was created ($n = 8$). Disc-shaped morphological SEs with radius r increasing from 3 to 24 m (step size is equal to 3 m) were used. In the DMP, structures with a scale (width) at the same level of the scale (diameter) of a specific SE will have high response at the position of that SE in the DMP. From observation of the DMP, we found that when $r \leq 6$ the DMP results for building extraction are not reliable since a lot of small structures that are darker (isolated trees) or brighter (substructures) than the surroundings will be confused with small buildings. Thus, we only utilized SEs with r from 9 to 24 m to detect buildings. Figure 3 demonstrates how the DMP decomposes an image based on structural information. As we will explain later, the small bright buildings will be detected by spectral analysis and the derivative of the closing profile with small SEs will be used for shadow detection.

To discriminate structures at a certain SE scale, a threshold should be set on each dimension of the DMP. The structures with DMP values higher than the threshold are hypothesized as candidate buildings. Bright buildings generally have high contrast with the surroundings so a relatively high threshold was set to 20, while dark buildings have relatively low contrast with the surroundings so the threshold was set to 15. After thresholding, connected components are labeled as separate candidate buildings.

To verify the hypothesized connected components as buildings, the following shape and size criteria must be satisfied.

- (1) The connected components should have an area compatible to the current SE scale. Components with areas less than half of the area covered by the current SE are rejected.
- (2) The minimum enclosing rectangle (MER) [27] of the current connected components is found. If the length of MER is longer than the distance of a typical city block, the corresponding connected component is rejected.
- (3) If the rectangular fit is lower than a threshold, the connected component is rejected since a majority of 2D building shapes in urban centers are rectangular. The rectangular fit is calculated as the area of the component divided by the area of its MER.

After verification based on the shape characteristic of the component, we observed that some parking lots were detected as well if they had a rectangular shape. So for those structures larger than a certain scale, we need to use contextual information to further verify the candidate connected component since buildings will cast shadows on the ground while parking lots will not. The position of shadows relative to buildings is known a priori based on satellite viewing geometry as it relates to sun azimuth and elevation. In our implementation, if a dark area (shadow) was detected on the “shadow” side of a connected component, then this component was finally verified as a building.

2.4. Shadow-supported building extraction

Context plays an important role in the identification and extraction of urban features from airborne and satellite imagery [28, 29, 30]. Here shadows were used as a strong supporting context for building extraction since a common feature of buildings is that they have some height above the ground and will therefore cast shadows.

The amount of shadows present in the high-resolution imagery will vary depending on the sun azimuth and elevation angles and the sensor azimuth angle. As the sun elevation angle decreases from 90° , the length of the shadow buildings cast on the ground grows, and as the sun and sensor azimuth angles get further apart, the amount of shadows visible to the sensor will increase. Also, the sun azimuth angle determines the position of the shadow relative to the buildings that cast it. For the IKONOS imagery in this study, the sun elevation angle was 61° , the sun azimuth angle was 139° , and the sensor azimuth angle was 352° . As a result, building shadows are prominent as seen in the image. Also, because of the sun azimuth angle, shadows are cast on the west and north sides of the buildings, and the shadows cast along the north direction are a little wider than the shadows along the west direction. So in the image, the shadows on the north side of the buildings are more readily seen than the shadows on the west side of the buildings.

2.4.1. Shadow delineation by DMP

In [3, 7] shadows were extracted from aerial imagery using simple gray-level thresholding. However, because of the insufficient spatial resolution of the high-resolution satellite imagery, pixel mixing is a significant problem for narrowly cast shadows. In this paper, we extracted shadows by analyzing the derivative of the closing profile (DCP) for small SEs with radius r from 3 to 12 m since the shadows are generally darker than their surroundings and the width of the shadows are typically within the scales of those SEs.

By using the same threshold for the dark building hypothesis in Section 2.3 on the DCP, the connected components are labeled. Since the shadows have low reflectance values in the NIR channel and low intensities in the PAN image, we set a relaxed threshold to filter out other structures that are obviously not shadows such as some trees. For the IKONOS image in this study, shadows have NIR reflectance values less than 350 and PAN intensities less than 400. In addition, we incorporate other restrictions so that most detected shadows are reliable. First, at least one of the dimensions along the potential shadow's north and west directions should be longer than 15 m to form reliable structures. Second, for some dark buildings also detected by DMP, we use an elongation measure to filter these structures out. The elongation measure was defined as the area of the component divided by the area of a square with the side length equal to the maximum distance of the pixels from the connected component to its boundary. Generally shadows will have an elongation measure much larger than 1.

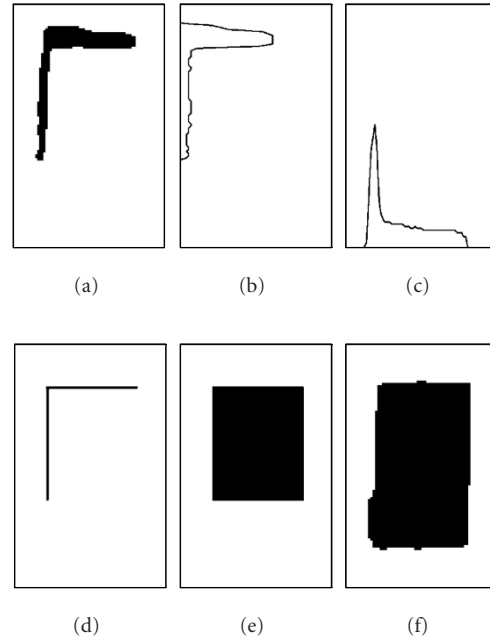


FIGURE 4: Illustration of shadow building context. (a) Detected shadow. (b) Its vertical projection. (c) Its horizontal projection. (d) Determined shadow corner. (e) Seed building rectangle. (f) Detected building after growing.

Since the position of the shadows relative to the building is known a priori from the collection geometry of the satellite imagery, we can detect the sun side of the shadow, which is the boundary between a building and its cast shadow. By assuming the building is rectilinear, we project the shadow-connected components onto the two perpendicular main directions of buildings. Here we project the shadow along horizontal and vertical directions since the image was rotated based on the primary road network directions in Section 2.1. The number of shadow pixels projected along each direction was accumulated. The upper/lower bound of shadows cast on the north side of the building was detected by thresholding the vertical projection. The left/right bound of shadows cast on the west of the building was detected by thresholding the horizontal projection. Because the only corners of interest are those whose concave sides are oriented toward the sun, we retain only the sun-side bound of the shadows and reject all corners whose bisecting vectors do not fall within the same quadrant as the sun direction. The image of one detected shadow and its projections are shown in Figure 4. If a shadow was detected along only one direction, the length of the shadow must be larger than a threshold to be a reliable detected shadow. We add a line with length equal to 10 m along another direction to form a corner, where 10 m is the minimum building dimension to form a reliable shadow. Through the projection analyses, a list of shadow corners were obtained where each record contained the positions of the corners and the dimensions of the shadow along two directions.



FIGURE 5: Detected shadow corners highlighted in red overlain on the urban area of IKONOS image of City of Columbia, Missouri.

The detected shadow corners by DMP were overlain on the original IKONOS image and are highlighted in red in Figure 5. From the image, we can see that many (but not all) obvious shadows are correctly delineated and are consistent with upper-left corners of the corresponding buildings.

2.4.2. Shadow-supported building verification and growing

The rectangular regions enclosed by the detected shadow corners in the preceding subsection are selected to form the hypothesized building seeds. The hypotheses are verified by analyzing the homogeneity and vegetation content of the enclosed rectangular regions. Homogeneity is measured by the maximum variance on each channel of the PS-MS imagery. The buildings have maximum variance less than 12 000. Vegetation content is measured using the normalized difference vegetation index (NDVI) that is defined as

$$\text{NDVI} = \frac{\text{NIR} - \text{R}}{\text{NIR} + \text{R}}, \quad (6)$$

where NIR is the reflectance value in the near-infrared channel, and R is the reflectance value in the red channel. Candidate rectangular regions are considered to be vegetation and the building hypotheses are rejected when the mean NDVI value in the region is higher than 0.06.

Typically, regions adjacent to shadows do not represent an entire building due to the incomplete detection of the shadow. To reconstruct the 2D building shapes, regions verified by the above steps are grown on the watershed segmented image. Because most of the building roofs are homogeneous, we grow the seed regions to neighboring segments when a homogeneity criterion is satisfied. In this paper, a heterogeneity measure is defined as

$$H = \max_{j=1, \dots, \text{nb}} \frac{|\mu_{0j} - \mu_{1j}|}{\sigma_{0j}}, \quad (7)$$

where nb is the number of channels which is 4 for PS-MS IKONOS imagery, μ_{0j} and σ_{0j} are the mean and standard deviation, respectively, of the intensity level of the j th band of a seed region, and μ_{1j} is the mean intensity level of the j th band of the neighboring segment. The seed region is grown iteratively into the neighboring segments when the heterogeneity measure is less than a predefined threshold.

Region growing applied to the edge-based watershed segmentation image greatly improves the speed of the growing algorithm and this satisfies the homogeneity criterion as well since edges are only generated on heterogeneous surfaces. In addition, region growing applied to the watershed segmentation will position the grown building boundary on the edges detected in the image.

2.5. Bright building extraction

Small buildings are not detected by the extraction based on the DMP and shadows. In the DMP analysis, the radius of SEs is chosen in the range of 9–24 m so that confusion between small buildings and other small substructures will be eliminated. In the shadow-supported building detection, we only consider reliable shadows longer than a certain threshold since the shadows of small buildings are often confused with other dark linear structures in the image.

From observation of the image, it is evident that a large number of the buildings in urban areas have a high intensity in the PAN image due to its concrete material and smooth surface. Thus, we use spectral information to aid in the detection of these buildings. The majority of small buildings can be detected in this way. After setting a threshold for the PAN intensity, we labeled the connected components in the binary image to obtain bright building seed areas. After deleting areas less than the minimum building dimension, which is $5 \times 10 = 50$ pixels for the IKONOS image, the seed areas are then grown on the watershed segmentation image to reconstruct the complete building shape. The detail of the growing step was described in Section 2.4.2.

2.6. Extraction integration

After building extraction by the three independent automated extraction strategies, the extraction results are combined using the logic “OR” operation. An evaluation of the extraction results for different combinations of the strategies was conducted and the results are reported in the next section. In addition, we analyze the interrelationship between the different strategies.

3. RESULTS

The automated building-extraction strategy using structural, contextual, and spectral information is tested using the IKONOS image of the City of Columbia, Missouri, acquired on April 30, 2000. The test site is a subsample of 1000×800 pixels that covers the urban center of the City of Columbia. The panchromatic image of this urban area is shown in Figure 6a. The copresence of buildings of different sizes and

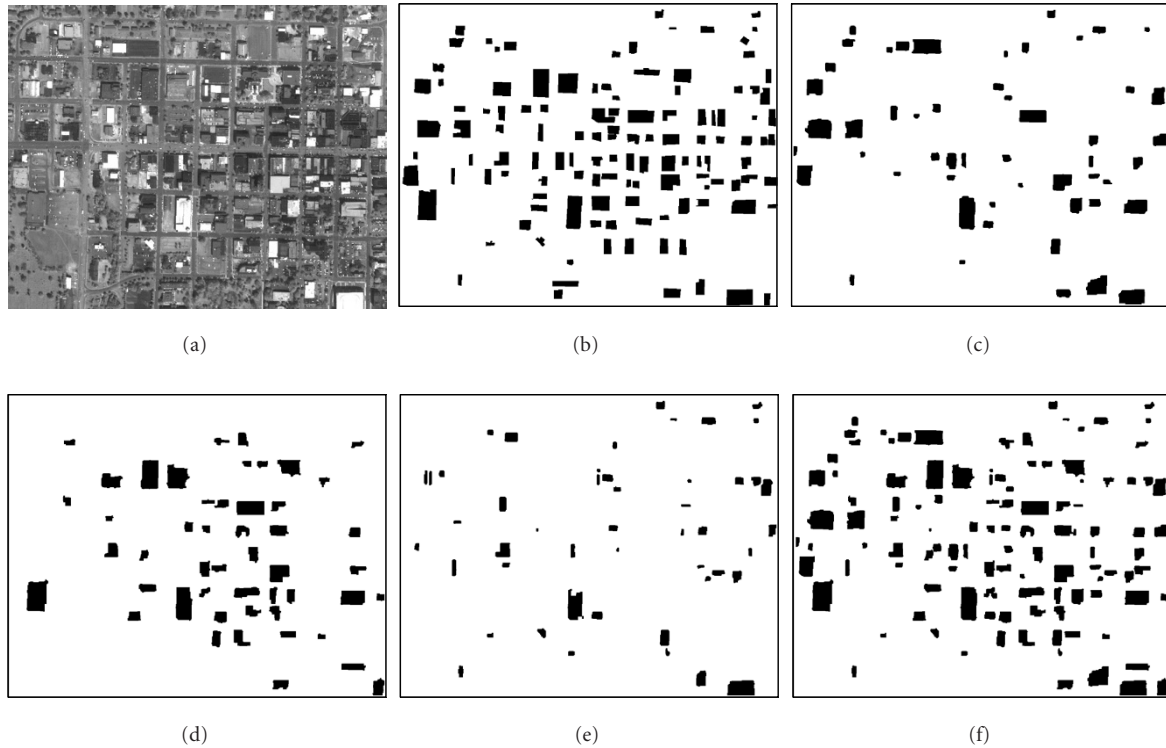


FIGURE 6: Building-extraction results. (a) Panchromatic IKONOS image of the test urban area. (b) Manually labeled buildings as ground truth. (c) Building-extraction result based primarily on structural (DMP) information. (d) Building-extraction result based primarily on contextual (shadow) information. (e) Building-extraction result based primarily on spectral information. (f) Final integrated building-extraction result.

spectral responses is obvious in the scene. Using the SSWHT on the urban area [10], the two primary directions for the road network were detected as 1° and 91° with respect to the horizontal direction in image space. Thus, most buildings and roads are oriented in north-south and east-west directions and no rotation of the image was needed as described in Section 2.1.

Within the urban test area, 121 buildings were manually delineated. The manually delineated buildings were used as a reference building set to assess the automated building-extraction accuracy. For performance evaluation, we use the evaluation measures widely accepted for building extraction [2, 15]. The extracted buildings and the manually delineated buildings are compared pixel-by-pixel. All pixels in the image are categorized into four types.

- (1) True positive (TP). Both manual and automated methods label the pixel belonging to the buildings.
- (2) True negative (TN). Both manual and automated methods label the pixel belonging to the background.
- (3) False positive (FP). The automated method incorrectly labels the pixel as belonging to a building.
- (4) False negative (FN). The automated method does not correctly label a pixel truly belonging to a building.

To evaluate performance, the number of pixels that fall into each of the four categories TP, TN, FP, FN are deter-

mined, and the following measures are computed:

$$\begin{aligned}
 \text{branching factor} &= \frac{\text{FP}}{\text{TP}}, \\
 \text{miss factor} &= \frac{\text{FN}}{\text{TP}}, \\
 \text{detection percentage} &= 100 \cdot \frac{\text{TP}}{\text{TP} + \text{FN}}, \\
 \text{quality percentage} &= 100 \cdot \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}.
 \end{aligned} \tag{8}$$

The interpretation of the above measures is as follows. The *detection percentage* denotes the percentage of building pixels correctly labeled by the automated process. The *branching factor* is a measure of the commission error where the system incorrectly labels background pixels as building. The *miss factor* measures the omission error where the system incorrectly labels building pixels as background. The miss factor can be derived from the detection percentage. Among these statistics, the *quality percentage* measures the absolute quality of the extraction and is the most stringent measure. To obtain 100% quality, the extraction algorithm must correctly label every object pixel ($\text{FN} = 0$) without mislabeling any background pixels ($\text{FP} = 0$).

The three extraction algorithms described in Sections 2.3, 2.4, and 2.5 were run independently on the test IKONOS image. The DMP was generated on the PAN image using

TABLE 1: Performance statistics of building extraction from different combinations of the multisource extraction strategies.

Detection source information	Branching factor	Miss factor	Detection percentage	Quality percentage
Structural	0.25	1.61	38.4	35
Contextual	0.31	1.4	41.7	37
Spectral	0.15	3.85	20.6	20
Structural and contextual	0.32	0.46	68.6	56.3
Structural and spectral	0.27	1.3	43.5	39
Spectral and contextual	0.27	0.71	58.6	50.7
Structural and contextual and spectral	0.33	0.38	72.7	58.8

disc-shaped morphological SEs with increasing radius r from 3 to 24 m (step size is equal to 3 m). The knowledge primarily used by each algorithm is structural, contextual, and spectral information, respectively. After independent extraction, the extraction results are combined.

As pointed out in the introduction, the building-extraction algorithms run automatically without preclassification or utilization of training sets, although some initial algorithm parameters must be set by the user. The chosen parameter values for our data set were provided in Section 2. In general, the algorithm parameters can be categorized into three groups: object scales, shape constraints, and spectral parameters. Object scale parameters include the distance of a typical city block (10 m), road width range (8–30 m), building length range (10–100 m), and width range (5–50 m). These parameters have special meaning and can be easily changed into image pixel units when applied to images with different spatial resolutions. The shape constraints include rectangular fit threshold (0.8) for building verification and elongation threshold (1.2) for shadow verification. These parameters were set as the default values and do not need to be adapted when used with different images. The spectral parameters include DMP thresholds, NIR and PAN intensity thresholds for shadows, homogeneity and NDVI thresholds for building rectangles. The default parameter values are determined based on the observation of spectral characteristics of individual features and need to be tuned with different images.

Performance evaluations for different combinations of the three extraction strategies are shown in Table 1. The building-extraction results are shown in Figure 6. From the results, we can see that extractions primarily based on the structural and contextual information have the largest contribution to the final detection percentage. The extraction using spectral information only detects bright buildings and therefore it has a relatively lower detection percentage. However, since small buildings cannot be detected by the extractions based on the structural or contextual information, it provides a reliable way to detect small bright buildings and contributes to the final detection percentage.

Among the three extraction methods, the two methods based primarily on structural and spectral information utilize properties of the building itself. The extraction results from these two strategies have significant overlap. In contrast, the extraction primarily based on contextual informa-

tion models shadows as a part of the scene and provides an information source independent of the properties of building object itself. Since this extraction uses a contextual information model, the result is relatively independent of the extractions from the two other information sources.

From the results in Table 1, we can see that no single extraction strategy can detect more than half of the building areas. This is because it is impossible to use a single model or information source to extract the wide variety of building types present in most urban areas where there are large variations in the shape, size, contrast, and spectral content. For the extraction based on DMP, buildings having low contrast with the surrounding background in PAN image cannot be detected. For example, a lot of buildings with a dark brown color shown in Figure 5 cannot be detected by DMP due to their low contrast with the surrounding areas in the panchromatic image shown in Figure 6a. However, for these kinds of buildings, they often have obvious shadow supporting information. Most of these building types are detected by the context-based building-extraction method. Conversely, buildings that have no reliable corresponding shadows can usually be detected by structural analysis from the DMP as long as they have a high contrast with their surroundings and are larger than a certain size. With the integration of structural, contextual, and spectral information, the detection percentage and quality of the building extraction are greatly improved. Overall, 72.7% of the buildings area is extracted with a quality percentage 58.8%. By close observation of the extraction result, we found that a small population of buildings with gabled roofs were not detected. In addition, a few dark parking lots such as the one near the upper-left corner of the image shown in Figure 6a was mistakenly labeled as building.

Comparing the extraction performance with other building extraction systems from aerial imagery [2], our extraction performance measures are among the best results obtained in [2] for 83 test images even though the satellite imagery used here has a lower spatial resolution. Because of the relative recent availability of high-resolution satellite images, only a few building-extraction results have been reported in the literature. Lee et al. [15] achieved a 64.4% detection percentage and a 51.3% quality percentage based on a class-guided building-extraction strategy, and we note that this strategy required supervised input (training data) and is therefore not fully automated.

4. CONCLUSION

An automated building-extraction strategy that uses structural, contextual, and spectral information is presented in this paper and applied to high-resolution satellite imagery. The recently proposed concept of DMP in mathematical morphology is used extensively in this paper to generate both building and shadow hypotheses based on the structural information. The final building extraction is done by integrating the results of three different extraction strategies based on different information sources. First, building hypotheses are generated and verified through shape analysis of the DMP. Second, shadows are modeled as a part of the scene and provide reliable contextual information to independently hypothesize the position and the size of adjacent buildings. Third, small bright buildings that cannot be reliably detected by structural and contextual information are detected using spectral information. Building shapes were reconstructed by using verified seed building areas that are then grown on a finely segmented image using a homogeneity criterion.

From the analysis of the extraction results, we can see that urban feature extraction can be greatly improved and enriched by using multiple information sources and models. By fusion of multisource information, our integrated building-extraction strategy detected 72.7% of the building areas in a complex urban image with a quality percentage of 58.8%. This extraction performance is among the best results reported in the literature thus far. However, there are still significant misclassifications in the extraction. Future work will extend the model to include buildings with gabled roofs. Also, using other information sources and more complex decision-level fusion strategies could improve the overall extraction performance.

ACKNOWLEDGMENT

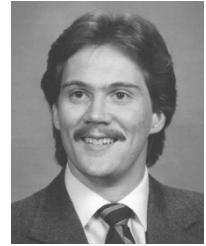
This research was supported by the Raytheon Synergy program under subcontract 3000615 from NASA.

REFERENCES

- [1] H. Mayer, "Automatic object extraction from aerial imagery—a survey focusing on buildings," *Computer Vision and Image Understanding*, vol. 74, no. 2, pp. 138–149, 1999.
- [2] J. A. Shufelt, "Performance evaluation and analysis of monocular building extraction from aerial imagery," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, no. 4, pp. 311–326, 1999.
- [3] B. Guindon, "Computer-based aerial image understanding: a review and assessment of its application to planimetric information extraction from very high resolution satellite images," *Canadian Journal of Remote Sensing*, vol. 23, no. 1, pp. 38–47, 1997.
- [4] Z. Kim and R. Nevatia, "Uncertain reasoning and learning for feature grouping," *Computer Vision and Image Understanding*, vol. 76, no. 3, pp. 278–288, 1999.
- [5] C. Lin and R. Nevatia, "Building detection and description from a single intensity image," *Computer Vision and Image Understanding*, vol. 72, no. 2, pp. 101–121, 1998.
- [6] C. Jaynes, F. Stolle, and R. Collins, "Task driven perceptual organization for extraction of rooftop polygons," in *Proc. IEEE Workshop on Applications of Computer Vision (WACV '94)*, pp. 152–159, Sarasota, Fla, USA, December 1994.
- [7] R. B. Irvin and D. M. McKeown, "Methods for exploiting the relationship between buildings and their shadows in aerial imagery," *IEEE Trans. Syst., Man, Cybern.*, vol. 19, no. 6, pp. 1564–1575, 1989.
- [8] P. Gamba, O. Hellwich, and P. Lombardo, "Theme issue on algorithms and techniques for multi-source data fusion in urban areas," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 58, no. 1–2, pp. 1–3, 2003.
- [9] D. Haverkamp and R. Poulsen, "Complementary methods for extracting road centerlines from IKONOS imagery," in *Image and Signal Processing for Remote Sensing VIII*, vol. 4885 of *Proceedings of SPIE*, pp. 501–511, Crete, Greece, September 2003.
- [10] X. Jin and C. H. Davis, "An integrated system for automatic road mapping from high-resolution multispectral satellite imagery by information fusion," *International Journal on Information Fusion*, vol. 6, no. 4, pp. 257–273, 2005, Special issue on fusion of urban remotely sensed features.
- [11] B. Guindon, "A framework for the development and assessment of object recognition modules from high-resolution satellite images," *Canadian Journal of Remote Sensing*, vol. 26, no. 4, pp. 334–348, 2000.
- [12] 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 Sensing*, vol. 41, no. 9, pp. 1940–1949, 2003.
- [13] W. Park, S. Kwak, and T. Kim, "Line-rolling algorithm for automated building extraction from 1-meter resolution satellite images," in *Proc. International Symposium on Remote Sensing*, pp. 31–36, Kyung-Ju, Korea, November 2000.
- [14] G. Sohn and I. J. Dowman, "Extraction of buildings from high-resolution satellite data," in *Automatic Extraction of Man-Made Objects from Aerial and Space Images (III)*, pp. 345–355, Swets & Zeitlinger B.V., Lisse, The Netherlands, 2001.
- [15] D. S. Lee, J. Shan, and J. S. Bethel, "Class-guided building extraction from Ikonos imagery," *Photogrammetric Engineering and Remote Sensing*, vol. 69, no. 2, pp. 143–150, 2003.
- [16] K. Segl and H. Kaufmann, "Detection of small objects from high-resolution panchromatic satellite imagery based on supervised image segmentation," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, no. 9, pp. 2080–2083, 2001.
- [17] E. Baltsavias, M. Pateraki, and L. Zhang, "Radiometric and geometric evaluation of Ikonos geo images and their use for 3D building modelling," in *Proc. Joint ISPRS Workshop on High Resolution Mapping from Space*, Hannover, Germany, September 2001.
- [18] J. P. Muller, C. Ourzik, T. Kim, and I. J. Dowman, "Assessment of the effects of resolution on automated DEM and building extraction," in *Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*, pp. 245–256, Birkhauser Verlag, Basel, Switzerland, 1997.
- [19] A. K. Shackelford and C. H. Davis, "A combined fuzzy pixel-based and object-based approach for classification of high-resolution multispectral data over urban areas," *IEEE Trans. Geosci. Remote Sensing*, vol. 41, no. 10, pp. 2354–2363, 2003.
- [20] J. Vrabel, "Multispectral imagery advanced band sharpening study," *Photogrammetric Engineering and Remote Sensing*, vol. 66, no. 1, pp. 73–79, 2000.
- [21] F. Dell'Acqua and P. Gamba, "Detection of urban structures in SAR images by robust fuzzy clustering algorithms: the example of street tracking," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, no. 10, pp. 2287–2297, 2001.

- [22] K. Price, "Road grid extraction and verification," in *International Archives of Photogrammetry and Remote Sensing*, vol. 32, pp. 101–106, Valladolid, Spain, June 1999.
- [23] L. Vincent, "Morphological gray scale reconstruction in image analysis: applications and efficient algorithms," *IEEE Trans. Image Processing*, vol. 2, no. 2, pp. 176–201, 1993.
- [24] P. Soille and M. Pesaresi, "Advances in mathematical morphology applied to geoscience and remote sensing," *IEEE Trans. Geosci. Remote Sensing*, vol. 40, no. 9, pp. 2042–2055, 2002.
- [25] L. Vincent and P. Soille, "Watersheds in digital spaces: an efficient algorithm based on immersion simulations," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, no. 6, pp. 583–598, 1991.
- [26] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, no. 2, pp. 309–320, 2001.
- [27] L. da F. Costa and R. M. Cesar Jr., *Shape Analysis and Classification: Theory and Practice*, CRC Press, Boca Raton, Fla, USA, 2001.
- [28] H. Mayer, I. Laptev, A. Baumgartner, and C. Steger, "Automatic road extraction based on multi-scale modeling, context, and snakes," in *International Archives of Photogrammetry and Remote Sensing*, pp. 106–113, 1997.
- [29] P. Burlina, V. Parameswaran, and R. Chellappa, "Sensitivity analysis and learning strategies for context-based vehicle detection algorithms," in *Proc. DARPA Image Understanding Workshop*, pp. 577–584, New Orleans, La, USA, May 1997.
- [30] S. Hinz, A. Baumgartner, and H. Ebner, "Modeling contextual knowledge for controlling road extraction in urban areas," in *Proc. IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*, pp. 40–44, Rome, Italy, November 2001.

Curt H. Davis received the B.S. and Ph.D. degrees in electrical engineering from the University of Kansas, Lawrence, Kan, in 1988 and 1992, respectively. He is currently the Croft Distinguished Professor of Electrical and Computer Engineering and Director of the Center for Geospatial Intelligence (geoint.missouri.edu) at the University of Missouri-Columbia. Dr. Davis is a former Chairman of the Instrumentation/Future Technologies Committee of the IEEE Geoscience and Remote Sensing Society. In 1996, Dr. Davis was selected by the International Union of Radio Science for their Young Scientist Award. In 1998, Dr. Davis was elected to the IEEE Senior Member grade and in 2002 he received the American Society of Photogrammetry and Remote Sensing (ASPRS) President's Award for Practical Papers for his publication on high-resolution digital elevation models for urban applications. Dr. Davis recently served as the Technical Program Cochairman of the 2004 IEEE Geoscience and Remote Sensing Symposium held in Anchorage, Alaska, in September 2004. His research interests are in the areas of radar systems, RF and microwave signal propagation, wireless communication systems, satellite and airborne remote sensing systems, satellite altimetry, high-resolution Earth image processing, ice sheet mapping and change detection, high-resolution digital elevation models, urban mapping, and feature extraction.



Xiaoying Jin obtained her B.S. and M.S. degrees in communication and information systems from the Department of Electrical Engineering, Wuhan University, China, in 1996 and 1999, respectively. During her study toward the M.S. degree, she worked as a Research Assistant in the Multimedia and Image Processing Laboratory, Wuhan University. She was awarded a Guizhiting Scholarship in 1995 and 1998. She worked as a software engineer in Huawei Technologies Co. Ltd., China, from 1999 to 2000. She is now working towards the Ph.D. degree in electrical and computer engineering from the University of Missouri-Columbia. She was awarded an Outstanding Academic Achievement Award by the university in 2002. Since August 2000, she has been a Research Assistant in the Remote Sensing Laboratory, University of Missouri-Columbia. Her current research interests include feature extraction from high-resolution satellite imagery, pattern recognition, computer vision, information fusion, and signal and image processing. Ms. Jin is a Member of Tau Beta Pi, IEEE, and SPIE Society.

