

Automatic Building Detection Using the Dempster-Shafer Algorithm

Yi Hui Lu, John C. Trinder, and Kurt Kubik

Abstract

An approach and strategy for automatic detection of buildings from aerial images using combined image analysis and interpretation techniques is described in this paper. It is undertaken in several steps. A dense DSM is obtained by stereo image matching and then the results of multi-band classification, the DSM, and Normalized Difference Vegetation Index (NDVI) are used to reveal preliminary building interest areas. From these areas, a shape modeling algorithm has been used to precisely delineate their boundaries. The Dempster-Shafer data fusion technique is then applied to detect buildings from the combination of three data sources by a statistically-based classification. A number of test areas, which include buildings of different sizes, shape, and roof color have been investigated. The tests are encouraging and demonstrate that all processes in this system are important for effective building detection.

Introduction

One of the major challenges in the fields of computer vision and digital photogrammetry is the 3D reconstruction of the terrain surface from aerial images of urban or suburban areas where buildings, roads, trees and vegetation are intermingled in an intricate and complex fashion. The automatic determination of Digital Terrain Models (DTM) by stereo image matching algorithms has been one of the primary goals of digital photogrammetry for many years, particularly for the production of digital orthophotos, 3D building reconstruction, 3D city models, the application and management of 3D databases for urban and town planning, and Geographic Information Systems (GIS) modeling.

Stereo image matching determines corresponding pixels or features in overlapping images and is fundamental to digital photogrammetry for elevation determination. However, conventional image matching supplies a Digital Surface Model (DSM) or visible surface, since it determines elevations of the tops of man-made objects such as buildings, or vegetation, and hence does not represent the terrain surface (Baltsavias *et al.*, 1995; Henricsson *et al.*, 1996; Tönjes,

1996). Therefore, it is necessary to identify buildings, trees and other objects on the surface to be able to reduce the elevations to the bare earth DEM. Many automated building detection and extraction methods have been proposed by researchers. Shadow analysis-based algorithms have been used by Liow and Pavlidis (1990) and Nevatia *et al.* (1999). Information fusion-based systems have been reported by McKeown (1991), and Haala and Hahn (1995). Methods supported by DTM and orthoimages have also been reported by Baltasavias *et al.* (1995), Horiguchi *et al.* (2000), Straub and Heipke (2001), and Brunn (2001). Considering the different shapes, environments, and image intensity for different buildings, together with the occurrence of occlusions and shadow effects, the automation of building extraction is a complicated and difficult procedure (Sahar and Krupnik, 1999). In addition to developing better schemes, the inclusion of more information is an essential direction for the research. Henricsson (1998), Chen and Hsu (2000), and Niederost (2001) used color images to improve the system performance for roof determination and edge extraction. Spreeuwears *et al.* (1997) and Gabet *et al.* (1997) used multi-view images to reduce the effect of occlusions. Multi-image 3D feature and DSM extraction for building change detection were proposed by Papanoditis *et al.* (1998 and 2001). Laser scanner data were used by McIntosh *et al.* (2000), Masaharu and Hasegawa (2000), and Haala *et al.* (1998). Multi-resolution analysis of wavelets for house extraction has been proposed by Shi and Shibasaki (1995). Huertas and Nevatia (1988), Shufel and McKeown (1993), Henricsson *et al.* (1997), Lammi (1996), Henricsson and Baltasavias (1997), Jaynes *et al.* (1997a), Jaynes *et al.* (1997b), Baillard *et al.* (1998), Kim and Muller (1998), Vosselman (1999), and Lu *et al.* (2003) have also demonstrated developments in algorithms for building extraction. Despite the development by researchers of many automatic building extraction algorithms based on images, terrain laser scan data and their combination (Collin *et al.* 1998; Hanson *et al.* 2001; Henricsson, 1998; Walter, 1999; Haala and Brenner, 1999), there are no operational algorithms because each method is focused on a particular application and data sources, and is usually not transferable to different features.

The goal of this research is to define building areas occurring on overlapping aerial or satellite images over a variety of terrain types and ground cover, for the reconstruction of terrain elevations of the bare earth surface, without the input of additional data such as terrain laser scanning or GIS. The approach includes an attempt to understand and

Yi Hui Lu is with the NSW Department of Environment and Conservation, Sydney 2000 NSW, Australia (Yi.Lu@environment.nsw.gov.au) She was formerly at the School of Surveying and SIS, University of NSW, UNSW Sydney 2052 NSW, Australia.

John C. Trinder is with the School of Surveying and SIS, University of NSW, UNSW Sydney 2052 NSW Australia (j.trinder@unsw.edu.au).

Kurt Kubik is with the Department of Computer Science and Electrical Engineering, University of Queensland, QLD 4072 Australia (kubik@itee.uq.edu.au).

Photogrammetric Engineering & Remote Sensing
Vol. 72, No. 4, April 2006, pp. 395–403.

0099-1112/06/7204-0395/\$3.00/0
© 2006 American Society for Photogrammetry
and Remote Sensing

interpret the image content and the characteristics of the terrain cover, such as buildings and trees, as an integral part of the process of determining terrain elevations. A further goal is also to demonstrate the suitability of data fusion for combining evidence derived from multiple data sources for detecting buildings.

General Description of Automatic Building Detection System

Figure 1 illustrates the architecture of the automatic building detection system developed in this research, which is divided into three parts. Only the Dempster-Shafer algorithm, shown as “Fusion” in the figure, will be described in detail. The components of the system aim to include information on elevations, spectral characteristics of the images, and outlines of area features to assist in the detection of buildings. The upper section of Figure 1, in the box named “Classification”, detects several characteristics of the terrain cover. On the left side of the box, a dense set of elevation points is determined by area-based image matching on a high-resolution stereo image pair of aerial photographs. These points represent the DSM of the region on an absolute height datum, and reveal objects above the terrain surface, such as buildings and trees.

The right hand side of this box in Figure 1 shows processes undertaken on multispectral images of the area. Ideally, the images would be the same as were processed for the determination of the DSM, but they may be a lower resolution set, if high-resolution multispectral images are not available. A K-means clustering is undertaken to detect four clusters, namely buildings, trees, ground, and grass. Then, using a post-classification procedure, a segmented image is created comprising regions of connected pixels that are contained in the same class. The NDVI (Normalized Difference Vegetation Index) is used to transform the multi-spectral data into an image band representing vegetation. In these processes, the NDVI and DSM can be considered as two key parameters, which may define the differences between vegetated and non-vegetated objects. For example, an initial assumption could be made that areas with heights above some limit, are likely to be either trees or buildings. Areas with low NDVI, and heights above the general terrain surface are likely to be buildings, whereas areas with high NDVI and heights above that surface are likely to be trees. Areas with high NDVI, with heights similar to the terrain surface are likely to be grassland or cultivated areas. Hence the four information

layers shown in the “Classification” box in Figure 1, namely, the land-cover classification, the results of the K-means clustering, DSM, and NDVI, play an important role in differentiating between buildings, trees, and grassland. They are therefore overlaid using a ArcView® Map Query operation to detect preliminary “building interest areas,” which are used as input to the “Shape Modeling” process.

In the “Shape Modeling” box, boundaries of the preliminary building interest areas derived from the “Classification” section are determined from the image using a level set method for curve propagating interfaces, which was introduced by Osher and Sethian (1988). It is based on mathematical and numerical work of curve and surface motion by Sethian (1985, 1995, 1999), and offers a highly robust and accurate method for tracking interfaces moving under complex motions. A more detailed description of the level set algorithm and the examples of its operation can be found in (Lu *et al.*, 2003).

In the “Fusion” box, the Dempster-Shafer fusion theory is used to combine three data sources, the DSM, the results of the K-means clustering and the building outlines, to detect final building areas, as described in the next sections.

Data Fusion Using Dempster-Shafer Theory

During the preliminary data analysis in the Classification box in Figure 1, it was evident that no single data source provided a consistent means of interpreting the image for detection of buildings. In order to reliably detect building regions, an approach that combines the evidence from a number of data sources is needed. Region evaluation with redundant data can help reduce imprecision in the detection of features, while complementary data can provide a more complete description. A multi-source evidential reasoning based region evaluation (MEBRE) module has been developed for this task. The module uses three data sources derived in Figure 1: the K-means clustered image (referred to as *Clustered image*), region outlines derived from the level set modeling (referred to as *LevelSet*) and the DSM. Evaluation at feature level was implemented in the MEBRE module by the Dempster-Shafer method, which is a statistical-based data fusion classification algorithm, used when the data contributing to the determination of the analysis of the images is subject to uncertainty. One of the advantages of this module is that it uses the spectral and spatial characteristics of the features. This approach effectively combines and counterbalances multiple evidence from different data sources in order to define the building regions, instead of relying on scoring techniques. A further advantage of Dempster-Shafer evidential theory is that it provides estimates of imprecision and uncertainty of the information derived from different sources (Shafer, 1976; Klein, 1999; Hegarat-Masclé, 1997). The application of a data fusion technique aims to improve the quality of a decision by making use of redundant and complementary data, while decreasing its imprecision and uncertainty.

Assume a set of n propositions making up the hypothesis space as denoted by Θ , and 2^Θ are the subsets of Θ . Based on the information from the data sources, a probability mass \mathbf{m} can be assigned to any proposition or union of propositions. For $\forall A \in 2^\Theta$, \mathbf{m} is defined for every element A and the mass value $\mathbf{m}(A)$ is in the interval $[0,1]$.

The following mass equations can be obtained:

$$\begin{aligned} \mathbf{m}(\emptyset) &= 0 \\ \mathbf{m}(\Theta) &= \sum_{A \subset 2^\Theta} \mathbf{m}(A) = 1 \end{aligned} \quad (1)$$

where \emptyset is the empty set.

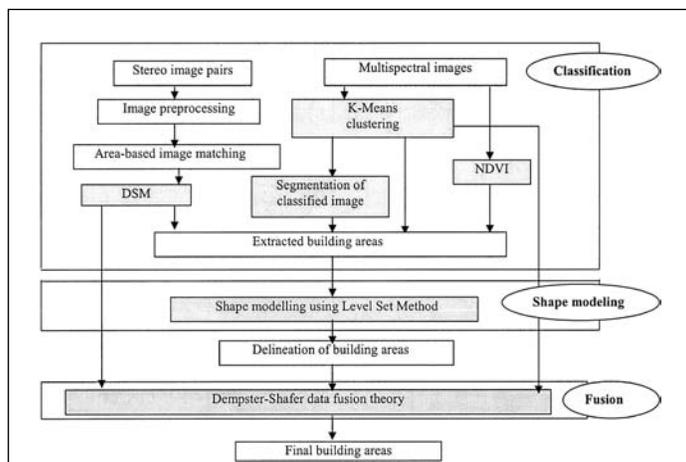


Figure 1. Architecture of the building detection system demonstrating the three modules of Classification, Shape Modeling, and Fusion.

In image classification, each pixel is assigned a class based on the content in the image, and Θ is the set of hypotheses about a pixel class. The Dempster-Shafer theory permits the consideration of any subset of Θ . Applied to image classification problems, it means that not only single classes, but also any union of classes can be represented. The number of classes (including all possible unions, but excluding the null set) is called the power of the set and is equal to $2^n - 1$. For example, if there are three propositions $n = 3$, there are $2^3 - 1 = 7$ classes, given by $C1$, $C2$, $C3$, $C1 \cup C2$, $C1 \cup C3$, $C2 \cup C3$, and $C1 \cup C2 \cup C3$ (Klein, 1999; Hegarat-Masclé, 1997; Shafer, 1976).

The Dempster-Shafer theory provides a representation of both imprecision and uncertainty through the definition of two parameters: *Support* (*Sup*) and *Plausibility* (*Pls*), which are obtained from the probability mass \mathbf{m} . *Support* for a given proposition means that all masses assigned directly by the data sources are summed. *Plausibility* for a given proposition means all masses not assigned to its negation are summed. For $\forall A \in 2^\Theta$ and $\forall B \in 2^\Theta$, the two parameters are defined respectively as follows:

$$\begin{aligned} Sup(A) &= \sum_{B \subseteq A} m(B) \\ Pls(A) &= \sum_{B \cap A \neq \emptyset} m(B) \end{aligned} \quad (2)$$

An uncertainty interval is defined by $[Sup(A), Pls(A)]$ where

$$Sup(A) \leq Pls(A)$$

$$Pls(A) = 1 - Sup(\bar{A}), \quad A \cup \bar{A} = \Theta, \quad A \cap \bar{A} = \emptyset. \quad (3)$$

The *Sup* of hypothesis A may be interpreted as the minimum uncertainty value about A . Its *Pls* may be interpreted as the maximum uncertainty value of A . For several data sources, the Dempster-Shafer method enables the combination of probability masses from these sources to obtain a single value for the probability of a proposition.

Assume there are two data sources, B_i and C_j , each comprising n object types. The object type A , is a subset of either B_i or C_j . The total probability mass committed to a subset A from the two data sources is

$$m(A) = \frac{\sum_{B_i \cap C_j = A} m_1(B_i)m_2(C_j)}{1 - K}, \quad K \neq 1$$

where

$$k = \sum_{B_i \cap C_j = \emptyset} m_1(B_i)m_2(C_j). \quad (4)$$

In the Dempster-Shafer theory, the hypotheses about single classes and unions of classes are respectively called simple hypotheses and compound hypotheses. When the probability masses of simple hypotheses are not null, a decision rule must be determined that best suits the application, such as the maximum *Sup*, which is formulated as follows:

$$\begin{aligned} \max(Sup(A)) \\ Sup(A) \geq Sup(\bar{A}). \end{aligned} \quad (5)$$

Multi-sources Region Evaluation

If there are $1, 2, \dots, p$ data sources and B_i ($i = 1 \dots n$) object types in the data sources, \mathbf{m}_n is the basic probability

mass provided by source n ($1 \leq n \leq p$, $p \geq 3$). Using Dempster-Shafer evidential theory, the combination of all the data sources is defined as follows:

$$\begin{aligned} m(A) &= \frac{\sum_{B_1 \cap B_2 \dots \cap B_p = A} \prod_{1 \leq n \leq p} m_n(B_i)}{1 - K} \\ K &= \sum_{B_1 \cap B_2 \dots \cap B_p = \emptyset} \prod_{1 \leq n \leq p} m_n(B_i). \end{aligned} \quad (6)$$

Multi-source Evidential Reasoning-based Region Evaluation Module

The multi-source evidential reasoning based region evaluation (MEBRE) method has been used to combine the three datasets of building classification (*Clustered image*), level set results (*LevelSet*), and *DSM*, to detect the final building areas as illustrated in Figure 2, in which $\mathbf{m}_n(B_i)$ is the probability mass for class B_i in Dataset n . Since regions defined by the *LevelSet* are spatial features, the corresponding feature areas in the *DSM* and *Clustered image* were determined in terms of their pixel locations. Probability masses $\mathbf{m}_n(B_i)$ were defined for each feature area in each data source, based on the Dempster-Shafer theory according to Equation 6. Building regions were then predicted from the *Sup* and *Pls* derived from Equations 2 and 5.

Dempster-Shafer theory has been used for unsupervised pixel level classification in Hegarat-Masclé *et al.* (1997). Because only non-spatial features were considered, classification errors occur at the boundaries of features. The evaluation procedure by Dempster-Shafer evidential reasoning, based on spatial features in this research aims to overcome this problem.

Clustered image, *LevelSet*, and *DSM* are assigned as Datasets 1, 2, and 3, respectively. Classes $C1$, $C2$, and $C3$ represent trees, buildings and ground, respectively. To determine the probability mass $\mathbf{m}(C1)$, Class $C1$ in Dataset 1 is combined with all the classes in Datasets 2 and 3 to find propositions which belong to class $C1$. Although there are 49 possible combinations for calculating probability mass $\mathbf{m}(C1)$, only the combinations which belong to class $C1$ should be included. Since the intersections of other classes, such as intersections between $C1 \cup C2$ in Dataset 1 and $C1$

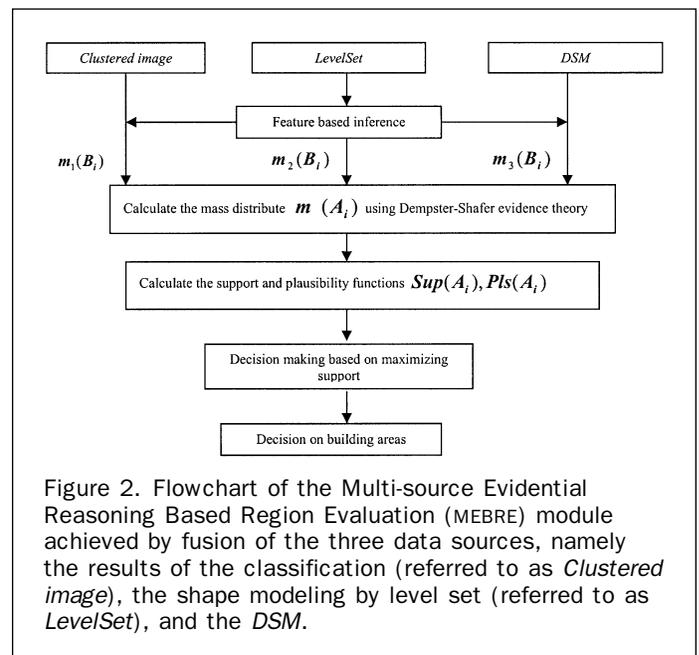


Figure 2. Flowchart of the Multi-source Evidential Reasoning Based Region Evaluation (MEBRE) module achieved by fusion of the three data sources, namely the results of the classification (referred to as *Clustered image*), the shape modeling by level set (referred to as *LevelSet*), and the *DSM*.

in Dataset 2 and 3, belong to class C1, they are also added to probability mass $m(C1)$.

Initial Probability Mass Definition for Region Evaluation

The definition of probability functions for region evaluation remains largely an unsolved problem. In image processing, the definition of the initial probability masses can be obtained at three different levels. At the most abstract or highest level, information representation is derived in a similar way to methods in artificial intelligence, where probability masses are assigned to propositions, often provided by experts (Gordon and Shortliffe, 1985). At the middle level, the definition of probability masses is derived from attributes, and may involve simple geometrical models (van Cleynenbreugel *et al.*, 1991). This definition is suitable for model-based pattern recognition, but it is difficult to use in image fusion for the classification of complex structures such as buildings in urban areas, for which no model exists. At the pixel level, probability masses are obtained from statistical pattern recognition. The most widely used approach assigns probability masses based on simple hypotheses only (Rasoulia *et al.*, 1990), but the absence of probability masses for compound hypotheses limits the power of Dempster-Shafer evidential method. This paper attempts to include both simple and compound probability hypotheses.

For MEMRE, the probability masses are assigned according to the information provided by each image. This method is more reliable and is able to take into account a larger variety of situations. For the *Clustered image*, the probability masses $m(A)$ have non-null values assigned to C1, C2, C3, and C2 \cup C3 as shown in Table 1. Since C2 for buildings and C3 for the ground may have the same spectral and texture characteristics in the image, there are ambiguities between these two classes in *Clustered image*. Hence, C2, C3, and C2 \cup C3 in Table 1 are assigned the same probability, t . Null probability masses are assigned to the other compound hypotheses. Since all the masses sum to 1, the probability mass for C1 is $1-3t$. For each detected region by *LevelSet*, the numbers of pixels representing trees, ground, grass, and buildings can be calculated, respectively. The probability t can be defined based on the number of pixels assigned to each building region:

$$\begin{aligned} \text{total} &= \text{NumofTree} + \text{NumofGround} \\ &\quad + \text{NumofGrass} + \text{NumofBuilding} \\ \text{probBuilding} &= \text{NumofBuilding}/\text{total}. \end{aligned}$$

Since $t \in \left[0, \frac{1}{3}\right]$ as shown in Table 1, the calculated probability mass of pixels being buildings from the above formula can be normalized in the t range in order to make the total probability equal to 1 for all object types in one dataset:

$$t = \frac{\text{probBuilding}}{3}. \quad (7)$$

For regions derived by *LevelSet*, since the detected building regions are derived from the processing of low-level image analysis and interpretation and level set modeling based image segmentation, they are considered to be more reliable and are assigned higher probabilities. As shown in Table 1, buildings in class C2 are assigned a probability of $1-3u$ and other non-null classes are assigned a probability of u . If the other classes are assigned lower probabilities, the class of building will have a higher probability since the sum of probabilities is 1. Hence, for all the test images, u has been assigned a value of 0.15 as shown in Table 4, which leads to an appropriately higher value of C2 of 0.55 for the building regions. In the tuning tests below on a single building, it will be shown that even when the probability value of C2 decreases to 0.45 (definition 4 of probabilities in Table 5), the building can still be detected.

For the *DSM*, since the buildings are in a suburban residential area, there are ample trees and grassed and bare ground surrounding the buildings. Buildings and most trees in the DSM will have similar heights above the ground, and hence these two classes are easily confused. The initial assignment of probabilities is given in Table 1, but the effects of tuning these assignments will be demonstrated in the following. In Table 1, class C1 representing trees, C2 buildings and C1 \cup C2 are therefore assigned the same probabilities, s , and hence the probability mass of class C3, the ground, is $1-3s$. The calculation of s is based on the mean of the DSM values. For example, if the elevation of an area, a building, or tree is high compared with the surrounding area, the probability s assigned to that area in the DSM will be higher than for other areas i.e., 0.33 or 0.34 as shown in Table 4 (0.34 is used for rounding-off purposes only to ensure the sum of the column is 1).

TABLE 1. CALCULATION OF PROBABILITY MASSES, SUPPORT AND PLAUSIBILITY, FOR DEMPSTER-SHAFFER DATA FUSION FOR THREE DATA SOURCES where $1 - k = 4(ut + st + us - 9uts)$

A	m_1 (B) <i>Clustered Image</i>	m_2 (B) <i>LevelSet</i>	m_3 (B) <i>DSM</i>	m (A)	$Sup(A)$	$Sup(\bar{A})$
C1 trees	$1 - 3t$	u	s	$\frac{4us(1 - 3t)}{1 - k}$	$\frac{4us(1 - 3t)}{1 - k}$	$\frac{4t(s + u - 6su)}{1 - k}$
C2 buildings	t	$1 - 3u$	s	$\frac{4st(1 - 3u)}{1 - k}$	$\frac{4st(1 - 3u)}{1 - k}$	$\frac{4u(s + t - 6st)}{1 - k}$
C3 ground	t	u	$1 - 3s$	$\frac{4ut(1 - 3s)}{1 - k}$	$\frac{4ut(1 - 3s)}{1 - k}$	$\frac{4s(u + t - 6ut)}{1 - k}$
C1 \cup C2	0	0	s	0	$\frac{4s(u + t - 6ut)}{1 - k}$	$\frac{4ut(1 - 3s)}{1 - k}$
C1 \cup C3	0	u	0	0	$\frac{4u(s + t - 6st)}{1 - k}$	$\frac{4st(1 - 3u)}{1 - k}$
C2 \cup C3	t	0	0	0	$\frac{4t(s + u - 6su)}{1 - k}$	$\frac{4us(1 - 3t)}{1 - k}$
C1 \cup C2 \cup C3	0	0	0	0	1	0

Experiments with Region Evaluation Module

Initially an example is given of the process of detecting the building shown in Figure 3 (which is confirmed as a building by visual interpretation of the images), first using two sets of data, namely the *Clustered image* and *LevelSet*, and then using all three sets. The purpose of this test is to demonstrate the significance of using all sources of data. Table 2 shows the numbers of pixels for 4 classes in the *Clustered image* corresponding to the building derived from *LevelSet* dataset. Based on Equation 7, the normalized probability of buildings t is 0.16, which is obtained from the classified pixel numbers in Table 2.

Table 3 shows computed probability masses, Sup and Pls values, for each simple and compound hypothesis for the model based on the two data sources, i.e., *Clustered image* and *LevelSet*. In the case where the probability masses of simple hypotheses are non-null, a decision rule that maximizes the Sup over all hypotheses will always favor compound hypotheses. For example, in Table 3, compound hypothesis $C1 \cup C2$ has $Sup(A) = 0.78$ and $Sup(\bar{A}) = 0.22$, which satisfies the decision rule, but does not provide a definitive decision on the correct feature. In

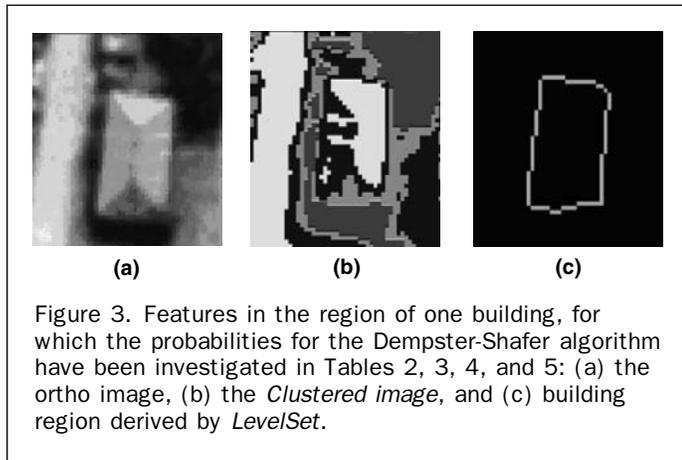


Figure 3. Features in the region of one building, for which the probabilities for the Dempster-Shafer algorithm have been investigated in Tables 2, 3, 4, and 5: (a) the ortho image, (b) the *Clustered image*, and (c) building region derived by *LevelSet*.

TABLE 2. NUMBERS OF PIXELS ASSIGNED TO EACH CLASS IN THE *CLUSTERED IMAGE* CORRESPONDING TO THE BUILDING IN FIGURE 3

Classes Number	Tree	Grass	Ground	Building
	17	90	194	276

TABLE 3. EVALUATION OF THE BUILDING SHOWN IN FIGURE 3 USING TWO DATASETS: *CLUSTERED IMAGE* AND REGIONS DERIVED BY *LEVELSET*

$t = 0.16, u = 0.15, 1 - k = 0.428$					
A	m_1 (B)		m (A)	$Sup(A)$	$Sup(\bar{A})$
	<i>Clustered Image</i>	<i>LevelSet</i>			
C1	0.52	0.15	0.36	0.36	0.64
C2	0.16	0.55	0.41	0.41	0.59
C3	0.16	0.15	0.22	0.22	0.78
$C1 \cup C2$	0	0	0	0.78	0.22
$C1 \cup C3$	0	0.15	0	0.59	0.41
$C2 \cup C3$	0.16	0	0	0.64	0.36
$C1 \cup C2 \cup C3$	0	0	0	1	0

order to avoid this situation, a decision rule involving only the simple hypotheses must be used. Based on Equation 5, there is no value of $Sup(A)$ which can be concluded as a maximum, and also $Sup(A) \geq Sup(\bar{A})$. Therefore, in this example, the detected building from two data sets is evaluated as non-building class, which would be an incorrect decision.

Table 4 shows the example of the same building detected based on three datasets. The value "0.53" (in the sixth column) is the maximum Sup value for a simple hypothesis, and at same time $sup(c2) > sup(\bar{c2})$. Therefore, the processed area has been correctly evaluated as a building.

Since choosing the initial probability masses is important, several experiments have been made to investigate how their selection affects the region evaluation. Table 5 summarizes definitions 2, 3, and 4, showing the changes in probabilities from those given for definition 1. For definition 2, the initial probabilities for Datasets 2 and 3 have been kept the same, while the probabilities for Dataset 1 have been changed as demonstrated by the arrows. The probability of the compound class $C1 \cup C3$ has been increased since grass is included in class C3 as ground, and may be confused with the class of trees. After Dempster-Shafer evaluation, the building is correctly assigned, because of the strength of the supporting probabilities. For definition 3, the probabilities for Datasets 1 and 2 remain the same as in definition 1, but the probability for the class of buildings s has been decreased and the compound probability $C2 \cup C3$ has been increased. In this case the building is also correctly detected. For definition 4 of probabilities in Table 5, the probabilities for Datasets 1 and 3 are the same as in definition 1, but the probability u in Table 1 has been decreased and the compound probabilities $C1 \cup C2$ and $C2 \cup C3$ increased, resulting also in correct detection of the class of building. This definition shows that, although the probability for the building class C2 has been decreased, a corresponding increase in the probabilities of the compound classes $C1 \cup C2$ and $C2 \cup C3$ will lead to the building still being correctly detected.

From these tests, it is noted that the building can be detected in all cases provided sufficiently high probability masses are assigned in the multiple datasets. The higher probabilities of unions of classes, which include the class buildings, such as $C1 \cup C2$ can assist in their correct detection. If the probabilities of the union of classes not involving buildings are incorrectly assigned values that are too high, the buildings will not be correctly detected. It is clearly important to ensure that a correct analysis is made of these probabilities before the Dempster-Shafer analysis is

TABLE 4. EVALUATION OF THE SAME BUILDING SHOWN IN FIGURE 3 AND TABLE 3 USING THREE DATA SETS: *CLUSTERED IMAGE*, *LEVELSET*, AND *DSM*, USING DEFINITION 1 OF PROBABILITIES

$t = 0.16, u = 0.15, s = 0.33, 1 - k = 0.22$						
A	m_1 (B)			m (A)	$Sup(A)$	$Sup(\bar{A})$
	<i>Clustered Image</i>	<i>LevelSet</i>	<i>DSM</i>			
C1 trees	0.52	0.15	0.33	0.47	0.47	0.53
C2 buildings	0.16	0.55	0.34	0.53	<u>0.53</u>	<u>0.47</u>
C3 ground	0.16	0.15	0	0.01	0	1
$C1 \cup C2$	0	0	0.33	0	1	0
$C1 \cup C3$	0	0.15	0	0	0.47	0.53
$C2 \cup C3$	0.16	0	0	0	0.53	0.47
$C1 \cup C2 \cup C3$	0	0	0	0	1	0

TABLE 5. SUMMARY OF SIGNIFICANT COLUMNS IN THE TABLES DERIVED FOR DEFINITIONS OF PROBABILITY MASSES 2, 3, AND 4, FOR THE DETECTION OF THE BUILDING SHOWN IN FIGURE 3. THE DIRECTIONS OF THE CHANGES IN THE PROBABILITIES ARE INDICATED BY ARROWS. THE RESULTING VALUES FOR $Sup(A)$ AND $Sup(\bar{A})$ ARE ALSO SHOWN

A	Definition 2 of Probabilities			Definition 3 of Probabilities			Definition 4 of Probabilities		
	$m_1(B)$ Clustered Image	$Sup(A)$	$Sup(\bar{A})$	$m_3(B)$ DSM	$Sup(A)$	$Sup(\bar{A})$	$m_2(B)$ LevelSet	$Sup(A)$	$Sup(\bar{A})$
C1 trees	0.27 ↓	0.47	0.53	0.23 ↓	0.30	0.70	0.12 ↓	0.45	0.55
C2 buildings	0.16	<u>0.53</u>	<u>0.47</u>	0.24 ↓	<u>0.57</u>	<u>0.43</u>	0.45 ↓	<u>0.55</u>	<u>0.45</u>
C3 ground	0.16	0.01	0.99	0	0.13	0.87	0.12 ↓	0	1
C1 ∪ C2	0	0.99	0.01	0.23 ↓	0.87	0.13	0.10 ↑	1	0
C1 ∪ C3	0.25 ↑	0.47	0.53	0	0.43	0.57	0.11 ↓	0.45	0.55
C2 ∪ C3	0.16	0.53	0.47	0.3 ↑	0.70	0.30	0.10 ↑	0.55	0.45
C1 ∪ C2 ∪ C3	0	1	0	0	1	0	0	1	0

undertaken. It would also be appropriate in future, to investigate the significance level of the Sup values required to ensure a reliable detection of buildings.

Tests and Results

The results of processing one area will be described in this section based on the probabilities defined typically by Tables 1 and 2. No further tuning of the probabilities was performed for these tests. A summary of results obtained for four additional areas will be given at the end of this section. Figure 4 illustrates the left image of a pair of 1:20 000 scale color aerial images with 515×521 pixels in the row and column directions, respectively, with a GSD of 0.3 meter, taken at a flying height of 3,070 meters. The scale of the images is smaller than desired, but larger scale images of the area were not available. The image contains a number of white roofed and two red roofed buildings. The majority of buildings have a distinguishable contrast against the background.

The processing steps followed were those shown in Figure 1. A dense DSM was determined using the digital

photogrammetry software Socet Set® v.4.2 to ensure that the majority of structures were revealed in the DSM. Due to the fact that no high resolution multi-spectral images were available, the color images were processed using the red and green image bands to obtain a Visible Vegetation Index (vvi) of the area, instead of the NDVI. The vvi was likely to be less effective in determining vegetation regions, but this was unavoidable. Areas with high vvi should represent the vegetation and the areas with low vvi, ground and buildings. The four information layers shown in the “Classification” box in Figure 1, namely DSM, K-means clustered image, segmented image and vvi were input into ARCVIEW®, using the Map Queries operation, and preliminary building interest areas detected. Using the region growing algorithm, small spots which did not belong to buildings were deleted from the Map Queries result. The resulting *building interest areas* are shown in Figure 5. Some road areas, wrongly assigned as buildings, were deleted and the correct building areas derived from classification have been successfully maintained. However, building areas that were not detected in the “Classification” step due to their unusual roof color, will not be recovered in subsequent steps.



Figure 4. Left image of the stereo aerial photos used in the first test.

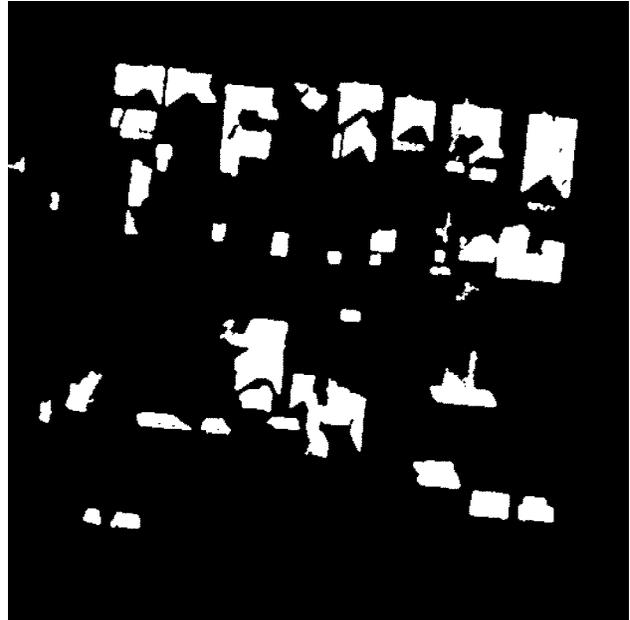


Figure 5. Detected building interest areas after Classification.

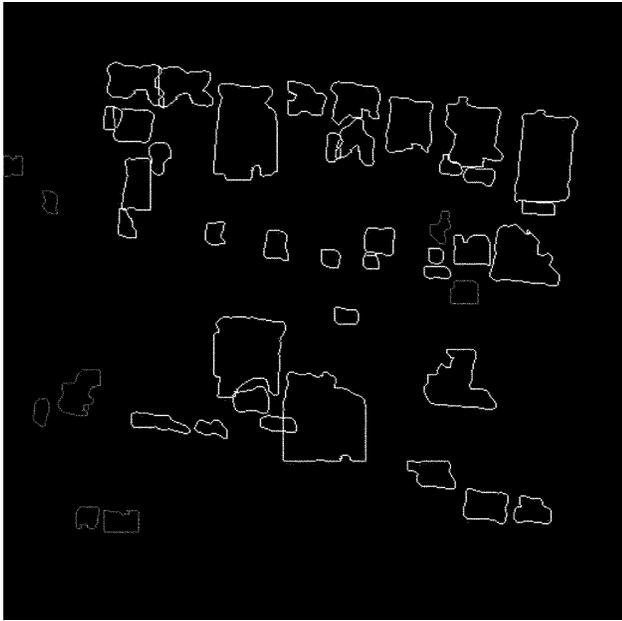


Figure 6. Detected buildings after application of data fusion by Dempster-Shafer based on the three datasets *Clustered image*, *LevelSet*, and *DSM*. Buildings shown in bright outlines are the correctly detected buildings, while those with darker outlines are those that were incorrectly detected as buildings and eliminated.



Figure 7. Detected buildings overlaid on the ortho image.

Shape modeling with the level set method was then implemented to delineate the boundaries of the buildings as revealed in Figures 6 and 7. Some roads and cars were assigned as buildings after the level set shape modeling, because the building interest areas supplied wrong information and caused the interpretation of some building regions to be unreliable. Thus, it was necessary to use Dempster-Shafer data fusion method to evaluate the regions.

Based on the decision rule given in Equation 5, the three data sources, *Clustered image*, *LevelSet*, and *DSM* were combined to produce the revised building areas shown in Figure 6. Most areas which were incorrectly assigned as buildings have been detected, as shown in darker outlines, while the bright outlines are the correctly detected buildings. The final building boundaries overlaid on the ortho-photo are shown in Figure 7.

The consequence of the data fusion is that eight incorrect building areas have been detected and deleted in the final result, as shown in Figure 7, which is a significant improvement in the detection rate. There were 32 buildings in the image, and 31 buildings were detected, as shown in the first line of Table 6. A summary of the results of five sets of images

processed for these tests, i.e., the test shown above and four further test areas given in Table 6 demonstrates that a success rate for building detection in excess of 80 percent is possible.

Conclusions

The method described in this paper combines stereo image matching, multi-spectral image analysis, shape modeling by the level set method, and Dempster-Shafer data fusion theory to locate building areas in the test images. The DSM and multispectral image analysis were used to supply approximate building areas, which were processed by the level set method to determine the outlines of the buildings. The level set method has been applied for the first time in this study for urban image analysis. The Dempster-Shafer data fusion technique provides the theoretical basis for evaluating the reliability of the detected buildings from the combination of the different data sources by a statistically-based classification. The tests described in the paper demonstrate that each step in the system is important, leading to an effective and robust detection of buildings with a success rate in excess of 80 percent.

Based on the test areas shown, the results are encouraging, but further research is needed to refine these methods, since some buildings will be missed in the early processing steps and never recovered because they have unusual roof

TABLE 6. SUMMARY OF RESULTS OF DETECTION OF BUILDINGS FOR ALL FIVE TESTS IN THIS STUDY

Tests in this Study	Total Buildings	Regions of <i>LevelSet</i>	Detected Buildings	Deleted Wrong Buildings	False Evaluation	Detection Rates
Test1	32	39	31	8	0	96.8%
Test2	96	91	85	5	1	88.5%
Test3	61	65	50	11	4	81.9%
Test4	50	43	40	3	0	80%
Test5	26	24	21	2	1	80.8%

characteristics, such as color and texture. Future improvements in the method could involve revision of the order of the processes to detect the lost buildings. In addition, further tuning of the initial probabilities should be undertaken to determine the behavior of *Support* and *Plausibility* values, and their level of significance for correctly defining buildings.

Acknowledgments

Research described in this paper has been supported by a grant from the Australian Research Council.

References

- Baillard, C., O. Dissard, O. Jamet, and H. Maitre, 1998. Extraction and textural characterization of above ground areas from aerial stereo pairs: A quality assessment, *ISPRS Journal of Photogrammetry and Remote Sensing*, 53:130–141.
- Baltsavias, E., S. Mason, and D. Stallmann, 1995. Use of DTMs/DSMs and orthoimages to support building extraction, *Automatic Extraction of Man-made Objects from Aerial and Space Images*, Birkhauser Verlag, Basel, pp. 199–210.
- Brunn, A., 2001. Statistical interpretation of DEM and image data for building extraction, *Automatic Extraction of Man-Made Objects from Aerial and Space Images (III)* (E.P. Baltsavias, A. Gruen, and L. Van Gool, editors), Balkema, Lisse, The Netherlands, pp. 171–180.
- Chen, L., and W. Hsu, 2000. Extraction of man-made buildings in multispectral stereoscopic images, *International Archives of Photogrammetry and Remote Sensing*, 33(3):169–176.
- Collin, R., C. Jaynes, Y. Cheng, X. Wang, and F. Stoller, 1998. The Ascender System: Automated site modeling from multiple aerial images, *Computer Vision and Image Understanding*, 72(2):143–162.
- Gabet, L., G. Giraudon, and L. Remouard, 1997. Automatic generation of high resolution urban zone digital elevation model, *ISPRS Journal of Photogrammetry and Remote Sensing*, 52(4):33–47.
- Gordon, T., and E. Shortliffe, 1985. A method for managing evidential reasoning in a hierarchical hypothesis space, *Artificial Intelligence*, 26:323–357.
- Haala, N., and M. Hahn, 1995. Data fusion for the detection and reconstruction of buildings, *Automatic Extraction of Man-Made Objects from Aerial and Space Image*, Birkhauser Press, pp. 211–220.
- Haala, N., C. Brenner, and K.-H. Anders, 1998. 3D urban GIS from laser altimeter and 2D map data, *International Archives of Photogrammetry and Remote Sensing*, 32(3/1):339–346.
- Haala, N., and C. Brenner, 1999. Extraction of building and trees in urban environments, *ISPRS Journal of Photogrammetry and Remote Sensing*, 54:130–137.
- Hanson, A., M. Marengoni, H. Schultz, F. Stolle, E. Riseman, and C. Jaynes, 2001. Ascender II: A framework for reconstruction of scenes from aerial images, *Automatic Extraction of Man-Made Objects from Aerial and Space Images* (E.P. Baltsavias, A. Gruen, and L. Van Gool, editors), Balkema, Lisse, The Netherlands, pp. 25–34.
- Hegar-Masclé, S., I. Bloch, and D. Vidal-Madjar, 1997. Application of Dempster-Shafer evidence theory to unsupervised classification in multisource remote sensing, *IEEE Transactions on Geoscience and Remote Sensing*, 35(4):1018–1031.
- Henricsson, O., 1998. The role of color attributes and similarity grouping in 3-D building reconstruction, *Computer Vision and Image Understanding*, 72(2):163–184.
- Henricsson, O., and E. Baltsavias, 1997. 3-D building reconstruction with ARUBA: A quantitative evaluation, *Automatic Extraction of Man-Made Objects from Aerial and Space Image (II)*, (A. Gruen, E.P. Baltsavias, and O. Henricsson, editors), Birkhäuser, Basel, pp. 65–76.
- Henricsson, O., F. Bignone, W. Willuhn, F. Ade, and O. Kuebler, 1996. Project AMOBE: Strategies, current status and future work, *International Archives of Photogrammetry and Remote Sensing*, 31(3):321–330.
- Horiguchi, S., S. Ozawa, S. Nagai, and K. Sugiyama, 2000. Reconstruction road and block from DEM in urban areas, *International Archives of Photogrammetry and Remote Sensing*, 33(3):413–420.
- Huertas, A., and R. Nevatia, 1988. Detecting buildings in aerial images, *Computer Vision, Graphics and Image Processing*, 41(2):131–152.
- Jaynes, C., A. Hanson, A. Riseman, and H. Schultz, 1997a. Building reconstruction from optical and range images, *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR'97)*, pp. 380–386.
- Jaynes, C., R. Collins, Y. Cheng, X. Wang, F. Stolle, H. Schultz, A. Hanson, and E. Riseman, 1997b. Automatic construction of three-dimensional models of buildings, *RADIUS: Image Understanding for Imagery Intelligence* (O. Firschein and T. Strat, editors), Morgan Kaufmann, San Francisco, pp. 223–236.
- Kim, T., and J.P. Muller, 1998. A technique for 3-D building reconstruction, *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(9):923–930.
- Klein, L., 1999. *Sensor and Data Fusion Concepts and Applications*, Second edition, SPIE Optical Engineering Press.
- Lammi, J., 1996. 3-D modelling of buildings from digital aerial imagery, *International Archives of Photogrammetry and Remote Sensing*, 31(2):213–217.
- Liow, Y.T., and T. Pavlidis, 1990. Use of shadows for extracting buildings in aerial images, *Computer Vision, Graphics and Image Processing*, 49:242–277.
- Lu, Y., J.C. Trinder, and K. Kubik, 2003. Automatic building extraction based on multi-source data fusion, *Proceedings of the Joint Workshop of ISPRS WG I/2, I/5 and IC WG II/IV and EARSeL Special Interest Group for 3D Remote Sensing*, University of Hannover, 06–08 October, Germany, unpaginated CD-ROM.
- Masaharu, H., and H. Hasegawa, 2000. Three-dimensional city modeling from laser scanner data by extracting building polygons using region segmentation method, *International Archives of Photogrammetry and Remote Sensing*, 33(3):556–562.
- Mcintosh, K., A. Krupnik, and T. Schenk, 2000. Improvement of automatic DSM generation over urban areas using airborne laser scanner data, *International Archives of Photogrammetry and Remote Sensing*, 33(3):563–570.
- McKeown, D.M., 1991. Information fusion in cartographic feature extraction from aerial imagery, *Digital Photogrammetric Systems*, Wichmann, Karlsruhe, pp. 103–110.
- Nevatia, R., A. Huertas, and Z. Kim, 1999. The MURI project for rapid feature extraction in urban areas, *International Archives of Photogrammetry and Remote Sensing*, 32(3–2W5):3–14.
- Niederost, M., 2001. Automated update of building information in maps using medium-scale imagery, *Automatic Extraction of Man-Made Objects from Aerial and Space Images (III)* (E.P. Baltsavias, A. Gruen, and L. Van Gool, editors), Balkema, Lisse, The Netherlands, pp. 161–170.
- Osher, S., and J.A. Sethian, 1988. Fronts propagating with curvature dependent speed: Algorithms based on Hamilton-Jacobi formulation, *Journal of Computational Physics*, 79:12–49.
- Paparoditis, N., G. Maillet, F. Taillandier, F. Jung, L. Guigues, and D. Boldo, 2001. Multi-image 3D feature and DSM extraction for change detection and building reconstruction, *Automatic Extraction of Man-Made Objects from Aerial and Space Images (III)* (E.P. Baltsavias, A. Gruen, and L. Van Gool, editors), Balkema, Lisse, The Netherlands, pp. 217–230.
- Paparoditis, N., M. Cord, M. Jordan, and J. Cocquerez, 1998. Building detection and reconstruction from mid- and high-resolution aerial imagery, *Computer Vision and Image Understanding-Special Issue on Building Detection and Reconstruction from Aerial Images* 72(2):122–142.
- Rasoulia, H., W. Thompson, L. Kazda, and R. Parra-loera, 1990. Application of the mathematical theory of evidence to the image cueing and image segmentation problem, *SPIE Signal and Image Processing Systems Performance Evaluation*, 1310:199–206.
- Sahar, L., and A. Krupnik, 1999. Semiautomatic extraction of building outlines from large-scale aerial image, *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(4):459–465.

- Sethian, J.A., 1985. Curvature and the evolution of fronts, *Communications In Mathematical Physics* 101(4):487–499.
- Sethian, J.A., 1995. Shape modelling with front propagation: A level set approach, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(2):158–175.
- Sethian, J.A., 1999. *Level Set Methods and Fast Marching Methods*, Cambridge University Press.
- Shafer, G., 1976. *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, New Jersey.
- Shi, Z., and R. Shibasaki, 1995. Automated extraction of man-made structures using region and line-based stereo matching in digital aerial images, *Proceeding of GIS AM/AF ASIA'95*, Thailand, pp. G.1.1–G.1.10.
- Shufel, J., and D. McKeown, 1993. Fusion of monocular cues to detect man-made structures in aerial imagery, *Computer Vision, Graphics, and Image Processing*, 57(3):307–330.
- Spreeuwears, L., K. Schutte, and Z. Houkes. 1997. A model driven approach to extract buildings from multi-view aerial imagery, *Automatic Extraction of Man-Made Objects from Aerial and Space Image (II)* (A. Gruen, E.P. Baltsavias, and O. Henricsson, editors), Birkhauser Verlag, Berlin, pp. 109–118.
- Straub, B., and C. Heipke, 2001. Automatic extraction of trees for 3D city models from images and height data, *Automatic Extraction of Man-Made Objects from Aerial and Space Images (III)* (E.P. Baltsavias, A. Gruen, and L. Van Gool, editors), Balkema, Lisse, The Netherlands, pp. 267–280.
- Tönjes, R., 1996. Knowledge based modelling of landscapes, *International Archives of Photogrammetry and Remote Sensing*, 31(3):868–873.
- Van Cleynenbreugel, J., S. Osinga, F. Fierene, P. Suetens, and A. Osterlinck, 1991. Road extraction from multi-temporal satellite images by an evidential reasoning approach, *Pattern Recognition Letter*, 12:371–380.
- Vosselman, G., 1999. Building reconstruction using planar faces in very high density height data. *International Archives of Photogrammetry and Remote Sensing*, 32(3–2W5):87–92.
- Walter, V., 1999. Automated GIS data collection and update, *Proceedings of Photogrammetric Week 1999*, Stuttgart, Germany, pp. 267–280.

(Received 04 October 2004; accepted 25 January 2005; revised 11 April 2005)