Semiautomatic Extraction of Building Outlines from Large-Scale Aerial Images

Liora Sahar and Amnon Krupnik

Abstract

One of the most challenging problems in digital photogrammetry today is automatic extraction of cartographic features from large-scale aerial images. Automatic detection of such features in general, and buildings in particular, will significantly improve the map compilation process as well as other tasks, such as DEM and orthophoto generation.

This paper presents a semiautomatic approach for the extraction of buildings from large-scale images. The main goal is to recover three-dimensional (3D) outlines of buildings. The extraction should be sufficiently accurate, reliable, and efficient in order to comply with the high standards set by the photogrammetric community. In order to obtain a 3D description, the approach comprises stereo analysis for extracting building heights, along with a general knowledge of building geometry and shadow information.

The paper explains the approach, presents experiments, and discusses the results.

Introduction

Along with the technological revolution in computing power in the past few decades, photogrammetry has progressed from analog stereoplotters, through analytical and computer-assisted instrumentation, to its current status. Nowadays, digital images are processed on softcopy photogrammetric workstations. These systems are capable of storing a huge amount of image data, and performing a wide variety of image processing tasks, such as image enhancement, edge detection, and spectral classification. Furthermore, these digital systems offer an integrated stereoplotting environment that can be easily connected to spatial databases and map production facilities.

The main advantage of digital photogrammetric systems is their potential to automate photogrammetric tasks, such as DEM generation, orthophoto production, aerotriangulation, feature extraction, and object recognition. Significant progress has been accomplished in automating the DEM generation task (see, e.g., Krzystek and Ackermann (1995)), which has led also to an automatic orthophoto production. Recent research studies show encouraging concepts and results with automatic aerotriangulation (Ackermann and Krzystek, 1997; Schenk, 1997; Toth and Krupnik, 1996). However, automatic extraction of man-made objects, which has been studied for more than a decade by the photogrammetric and computer vision communities, is still in its early research stage. The main reason is that recognizing semantic information, which is a relatively easy task for a human operator, is an ill-posed problem that is a major obstacle for a computer.

This paper discusses the main reasons for the lack of working systems for cartographic object extraction, and describes our efforts to define an efficient, reliable approach for extracting man-made objects in general, and buildings in particular, from large-scale images. The described approach comprises two major concepts: (1) extraction is performed from the object view point, i.e., in three-dimensional (3D) object space; and (2) objects are extracted semiautomatically. Integrating both concepts in a conventional, well-established photogrammetric procedure complies with the primary requirement of the mapping community, which is to have an efficient extraction system, while maintaining the required accuracy and reliability.

The general approach is demonstrated by basic tools for extracting building outlines from a stereo pair of large-scale images. The use of the two basic concepts is justified for this case. Buildings are 3D objects, and the use of the third dimension, i.e., the elevation of the building edges, may greatly improve the extraction procedure. Concerning the second concept, compromising on a semiautomatic procedure (compared to a fully automatic approach) does not reduce the efficiency of the object extraction task significantly. Buildings are noticeable objects that can be easily identified interactively and marked for automatic extraction by a window or even a single point.

Previous Work, Motivations, and Concepts

Extracting building outlines from large-scale aerial images has been performed manually by human operators, with high accuracy and reliability, for many years. Nevertheless, it is a time and labor consuming task.

At first glance, buildings may seem to be simple objects, which can be easily identified and extracted. However, automatic extraction of buildings from large-scale images must deal with several difficulties:

- Buildings appear differently from different viewpoints;
- Parts of the building may be occluded by interfering objects such as other buildings or trees, or darkened by shadows;
- Buildings are man-made objects and therefore designed in varied shapes and sizes; and
- Using one image, usually only 2D information is extracted while, for many applications, height information is essential. Even the 2D information is not necessarily correct if perspective images are used, and no information is available about elevations.

The examples presented at the results section demonstrate some of these difficulties.

In the past two decades, many studies, presenting different approaches for building detection and extraction, have been published (e.g., Haala and Anders (1997); Huertas and Nevatia (1988); Irvin and McKeown (1989); Shi and Shibasaki (1996); and Weidner and Förstner (1995)). These studies

Department of Civil Engineering, Technion - Israel Institute of Technology, Haifa, 32000 Israel (krupnik@tx.technion.ac.il.).

Photogrammetric Engineering & Remote Sensing,
Vol. 65, No. 4, April 1999, pp. 459–465.

0099-1112/99/6504–459$3.00/0

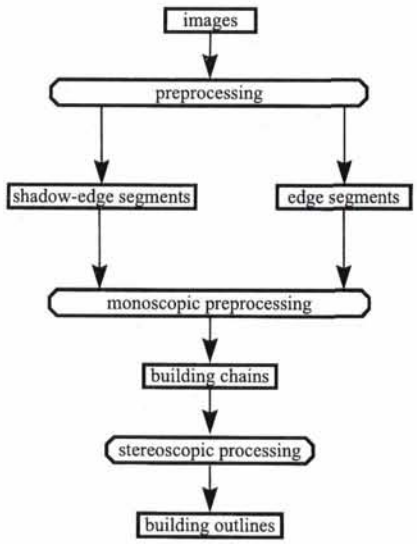

Figure 1. General outline of the proposed approach.

are aimed mainly at identifying the buildings and not necessarily at accurately measuring their outlines. Lang and Förstner (1996) describe a method for accurately modeling buildings. In their study, however, building types (models) are selected and approximately fitted to their appearance in a single image. The difference between the latter study and the idea described in this paper is that no model fitting is used, and the building is searched in a window according to *a priori* knowledge.

One of the major questions that may be raised after more than a decade of extensive research on the subject of buildings extraction is "why is cartographic object extraction still performed manually on softcopy photogrammetric systems?" It is obvious that digital systems that are capable of automatic extraction of objects may save many human working hours. Therefore, with all the sophisticated algorithms studied by the computer vision and photogrammetric communities, we would expect that such systems would be available by now.

The answer to this question may be implicit in the following observations:

- Accuracy, reliability, and efficiency are primary requirements for an acceptable system for cartographic object extraction.
- Photogrammetric systems (either analytical or digital) constitute a well-established, widely used, accurate, and reliable environment for interactive cartographic object extraction. Therefore, replacing them with automatic procedures must guarantee higher efficiency.
- Automatic systems for object extraction, as reflected in the literature, do not necessarily generate fully reliable results. Despite the fact that from an automated system viewpoint 80 to 90 percent correct results are considered to be a success, this is not the case from the cartographic viewpoint.
- Any case where results are not fully correct entails long hours of post processing, which includes interactively locating incorrect results and correcting them. It is clear that the photogrammetric/cartographic production communities will find it hard to accept any "black-box" type system that will allow for "some" incorrect results.

The study presented in this paper proposes a semiautomatic technique for building extraction. The technique is based on the following three general key points:

- Buildings are 3D objects, and the information implicit in this property is valuable. Therefore, extraction of buildings is performed in the object space. Three-dimensional locations of building outlines are obtained by using two images simultaneously. As explained in the following sections, building heights are not only calculated at the end of the extraction process. They are also used as a significant source of information during the matching stage.
- Buildings are 3D objects that cast shadows on the ground surface. These shadows can be used to improve the outline identification procedure, help with distinguishing between buildings and other 2D objects, and for verifying building hypotheses. In addition, shadows are detected in order not to confuse them with buildings.
- There is no attempt to detect buildings within an entire scene. A pair of image patches that contain the same building (and possibly other objects and features) is used for extracting the building outline. Such a pair is available either by a manual procedure or a different utility that only identifies potential buildings.

Reconstruction of Building Outlines

The proposed approach for reconstructing building outlines consist of three stages (Figure 1): Preprocessing, monoscopic processing, and stereoscopic processing. This general approach is suitable for extracting any type of building outlines, depending on the *a priori* knowledge implemented. Nevertheless, within the scope of this paper, the following assumptions are made:

- Buildings have planar rooftops. Objects may appear on the roof; however, corners of the building contours have the same elevation.
- Buildings may be described by four characterizing points, each of which is a corner of a circumscribing rectangle.
- A building outline may be constructed of more than four line segments (see Results Section). However, additional corners should not interfere with the previous assumption, i.e., the general shape is still a rectangle.

Preprocessing

The preprocessing stage takes a raw image as an input and produces two sets of edge segments: segments that belong to shadow areas, and segments that belong to other features. This operation is performed for each image separately.

The preprocessing stage consists of the following steps (Figure 2):

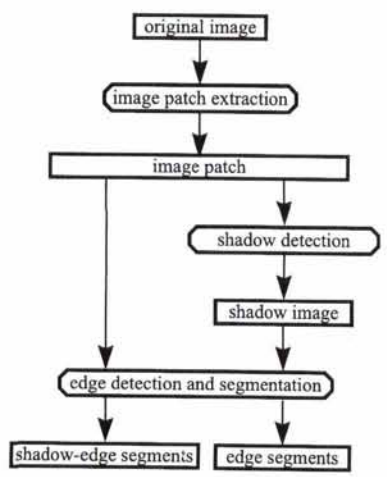

Figure 2. The preprocessing stage.

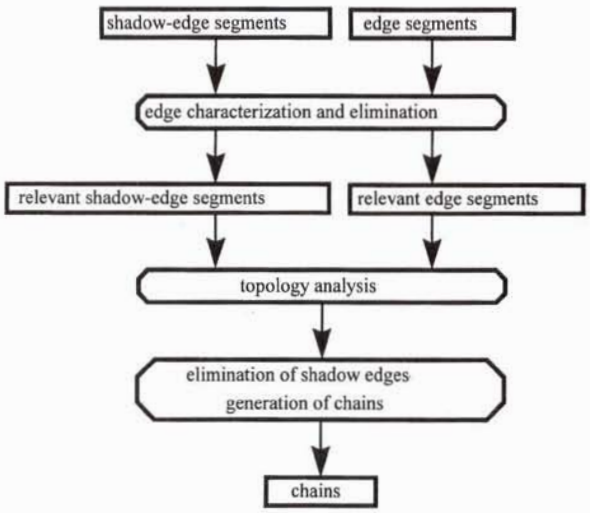

Figure 3. The monoscopic processing stage.

Image Patch Extraction
As described earlier in this paper, the proposed approach is semiautomatic. An image patch (referred to as a window) that contains only one building is extracted from the large aerial image. It should be noted that, although it is assumed that only one building is included in the window, other interfering features, such as parts of other buildings, shadows, and objects on the ground, may be presented.

Shadow Detection
Shadows projected on the ground around the building are features that can help with defining the height of the building, as well as with the decision on what side of an edge is inside a building and what side is not. On the other hand, shadows may appear as nice closed features, which may be wrongly detected as buildings. Fortunately, shadows projected on the ground do not depend on the color or texture of the building but only on the shape, the direction of illumination, and the terrain. They appear as the darkest regions on the image. Therefore, it is relatively easy to detect them by, for example, elementary supervised classification. The result of this classification is a binary shadow image. One should note, however, that analyzing the relations between the shadows and the buildings is a non-trivial task, which is only briefly touched on in this research.

Edge Detection and Segmentation
Building contours appear as edges in images. Edge detection is therefore performed on the window, in order to find potential lines that describe the contours. A similar operation is performed on the shadow image, to detect the outline of shadows. Ideally, the shadow edges constitute a subset of the edges from the gray value image. Edges are than grouped and segmented to straight line segments, characterized by their length, magnitude, end points, and orientation. In the current implementation, the Optimal Zero Crossing Operator (OZCO) (Sarkar and Boyer, 1991) is used for edge detection, grouping and segmentation.

Monoscopic Processing
The monoscopic processing stage is aimed at selecting topologically organized sets of edge segments, referred to as *chains*, out of the set of edge segments. A chain consists of up to four nearly right-angled corners, which describe a potential outline of a building. In each of the two image patches, a building may have a few potential chains, to be used later on in the stereoscopic processing. One should realize that, as this procedure is aimed at detecting a certain type of building, there is no attempt to use any sophisticated mechanism of grouping that produces all possible surfaces out of the edges (such as presented in, e.g., Fradkin and Ethrog (1997)). This approach allows a relatively simple computation model, which is suitable for a semiautomatic building detection environment.

The monoscopic processing stage consists of the following steps (Figure 3):

Elimination of Irrelevant Edges
The edge segmentation procedure produces an excessive number of edge segments. Many of these segments describe features that do not belong to the building contour. A reasonable assumption concerning edges that belong to building outlines is that they are strong and sufficiently long. Edges that are shorter or weaker than predefined values are therefore not considered for further processing. Nevertheless, they are stored for a potential final verification step.

Definition of Topological Relations between Edge Segments
In order to create chains, a definition of topological relations between the line segments is essential. Three such relations are defined here between each pair of lines: parallelism, perpendicularity, and adjacency. Parallel and perpendicular pairs of lines are found by their orientation. Two line segments are considered parallel if the absolute value of the difference between their orientations is not greater than a predefined threshold $d\alpha$. In a similar way, two line segments are defined as perpendicular if the absolute value of the difference between their orientation is in the range $90° \pm d\alpha$. Adjacency is defined by the distance between line ends, or between a line end and the other line. Line segments are considered adjacent (or neighbors) if the distance between them is shorter than a predefined value dl. Having these relations, a pair of perpendicular and adjacent line segments constitutes a corner.

Generation of Chains
Chain generation is performed by grouping corners. The ideal case is a set of four corners that is ordered as a rectangular shape. In practice, not all corners are detected. Therefore, other configurations, pairs and triplets of corners, are grouped. A pair is formed either from two adjacent corners (Figure 4a) or from two corners on a diagonal of a rectangle (Figure 4b). Selecting a pair of corners is controlled by logical geometric constraints. With these constraints, only pairs that may form a rectangle (like those presented in Figures 4a and 4b) are selected, while others (like those presented in Figures 4c and 4d) are not.

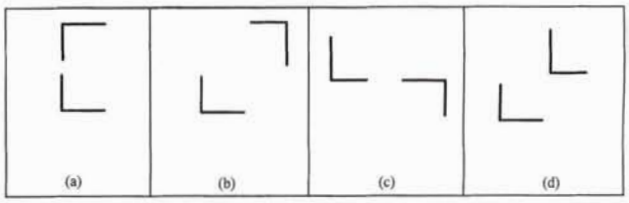

Figure 4. Valid (a,b) and non-valid (c,d) corner configurations.

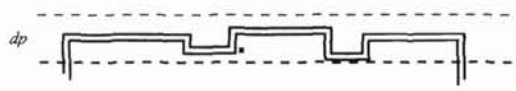

Figure 5. A band that defines a building edge.

In the ideal case, having a pair of adjacent corners, a leg that belongs to one corner lies on the same straight line as a leg from the other. Because the edge detection, segmentation, and grouping steps do not always produce the entire line between the corners, only fragments of this line pertain in the data. Moreover, in many cases not all these fragments lie exactly on the same line. There are two possible reasons for this situation. The first case is when a building edge is, in fact, not exactly a straight line. The second reason is that there may be more than one parallel line near the building edge, like the case where there is a concrete belt around the roof top. The definition whether two corners are "adjacent" on a chain therefore relies on an additional threshold, i.e., the distance *dp* between parallel lines. In practice, this distance defines a band along the building edge, as shown in Figure 5.

Elimination of Shadow Edges

Some of the generated chains belong to shadow outlines, which also consist of relatively long, strong edge segments. These are eliminated by matching them with edge segments that appear in the set of shadow edges generated in the preprocessing stage. Because a shadow edge resulted from the original window and one that resulted from the shadow window appear to be virtually in the same location, this matching procedure is as simple as finding a subset of edges according to their coordinates and geometric description.

Adding Missing Corners

As mentioned earlier, ideal chains should contain four corners that represent a rectangular shape of a building. Unfortunately, on many occasions this is not the case. If two corners on the diagonal are found, the other two corners are generated by a simple line intersection procedure. In other cases, i.e., when only two adjacent corners of a rectangle are found, the other corners cannot be added at this stage. In most cases, it is possible to add them during the stereoscopic processing.

At the end of this stage, a building outline has several candidate chains in each image. Apparently, parts of them may belong to other rectangular features in the image. The latter are eliminated in the following stage.

Stereoscopic Processing

In this stage, potential building chains from the two images are used for determining a single 3D outline of a building. For that purpose, *a priori* knowledge about the 3D geometry of a building is utilized. In order to generate 3D coordinates of a corner, a determination of which corner from the left image contour is compatible with a corner from the right image (and vice versa) is required. Detected corners are labeled according to their position within a chain (determined by the orientation of the legs in each corner). Allowing a small rotation between the corners on the two images, the corner labeled as 1 in one image is matched to a corner with the same label on the other image and so on. Three-dimensional coordinates are then calculated. In cases where there are several corners with the same label on each image, all possible combinations are checked. In practice, the number of possibilities is relatively small at this stage.

The stereoscopic processing stage consists of the following steps (Figure 6):

Generation of Missing Corners

Missing corners, in the case in which only two adjacent corners of the rectangle exist, cannot be added during the monoscopic processing stage. In order to add these missing corners, stereo information is used. Calculating the coordinates of a point in an image based on its location on the other image of a stereo pair, requires geometric constraints. One example is the case of one matched corner between the left and right chains. Having at least one matched corner in a combination (a corner for which 3D coordinates are calculated) enables adding corners detected in one image and missing on the other. This is done by using the assumption that the height of the missing corner is the same as the height of the matched one. The location of the point on the other image is determined by calculating the intersection of the building edge and the epipolar line. A different constraint is used to solve the problem where no corners are matched, and there are adjacent corners in a chain in each image. In such cases, the 3D location of corner 1 is constrained to be on line 1–2 (see Sahar (1997) for further details). Other cases are treated in a similar way.

Selection of the Correct Building Outline

At this stage, there exist several candidates for a building outline. Obviously, only one of them describes the building. The rest represent other objects or interfering connected contours. Initial elimination is performed using the assumption that corners on the roof should have the same height or minimal differences in heights. Selecting the correct outline from the remaining candidates is done by a scoring process, which is based on the 2D and 3D information obtained in earlier steps. In particular, the scoring process involves the following parameters:

- Type of corners (an "L" corner gets a higher score than a "T" corner);
- Distance between the line ends of the corners;
- Total edge coverage between the corners;
- Corner configuration in 2D: a pair of adjacent corners gets a higher score than a pair that lie on the diagonal;
- Rectangular shape of the 3D building outline; and
- Elevation differences between corners.

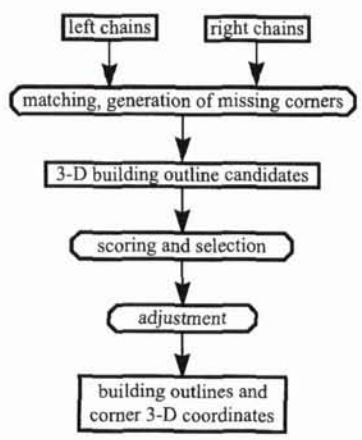

Figure 6. The stereoscopic processing stage.

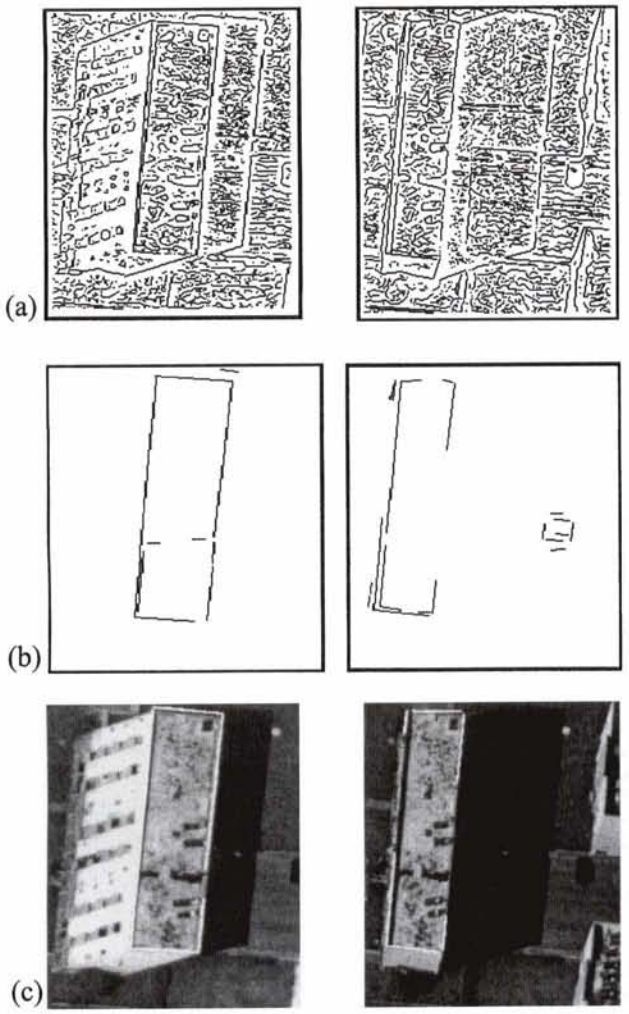

Figure 7. Results for building no. 1: (a) extracted edges; (b) chains after monoscopic processing; (c) resulting corners, overlaid on the image patches.

Adjustment of the Selected Outline

Due to displacement of edges, and due to interfering objects, the shape of selected building outline may not be a perfect rectangle that lies on a horizontal plane. An adjustment procedure has been used to adjust the corners to fit a horizontal rectangle, similar to what is done in commercial mapping systems.

Experimental Results

The approach described in the previous section has been implemented and tested on two data sets. Both scenes contain high buildings, were taken with a wide angle lens, were scanned with a photogrammetric PS1 scanner, and are of medium quality. The first data set (CAMPUS) contains three photographs that were taken with a Zeiss RMK camera, have a photo scale of 1:4000, and were scanned at a 15-μm resolution. The second data set (URBAN) contains two photographs that were taken with a Wild RC20 camera, have a photo scale of 1:6000, and were scanned at a 14-μm resolution.

A total of seven buildings were tested: three in the CAMPUS data set and four in the URBAN set. Because the CAMPUS area is covered by three photographs, there was a possibility to use the same building on a different stereo model. Therefore, two of the buildings were tested in different models.

Figures 7, 8, and 9 show sample results. These are presented in three stages:

- Edges detected in the preprocessing stage;
- Chains extracted in the monoscopic processing stage; and
- Selected outlines, obtained in the stereoscopic analysis, projected to and overlaid on the original image patches.

In the building presented in Figure 7, it can be clearly seen that the concrete belt surrounding the roof resulted in a shadow band on parts of the roof. This line was detected during the edge detection procedure and created additional corners. During stereo processing, the chain that was selected is the one that is actually placed on the outline of the building.

Figure 8 presents a more complex building. The outline of the building, in this case, is formed from four corners with gaps between them. The building was detected properly except for the left lower corner on the left image, where the concrete belt generated a strong shadow edge. The corner was displaced and adjusted in the final processing stage.

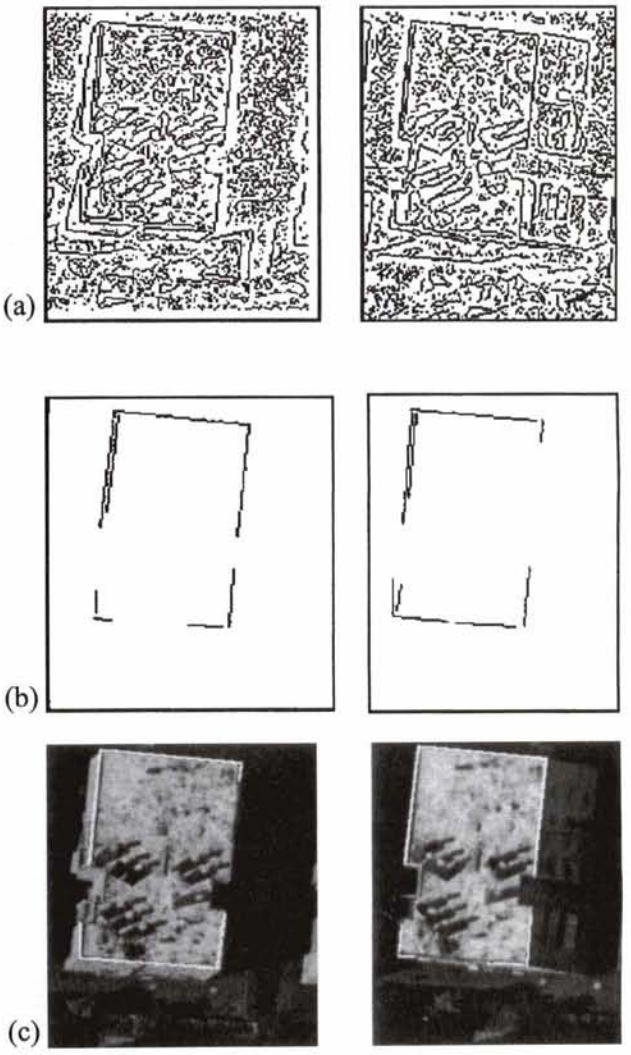

Figure 8. Results for building no. 2: (a) extracted edges; (b) chains after monoscopic processing, (c) resulting corners, overlaid on the image patches.

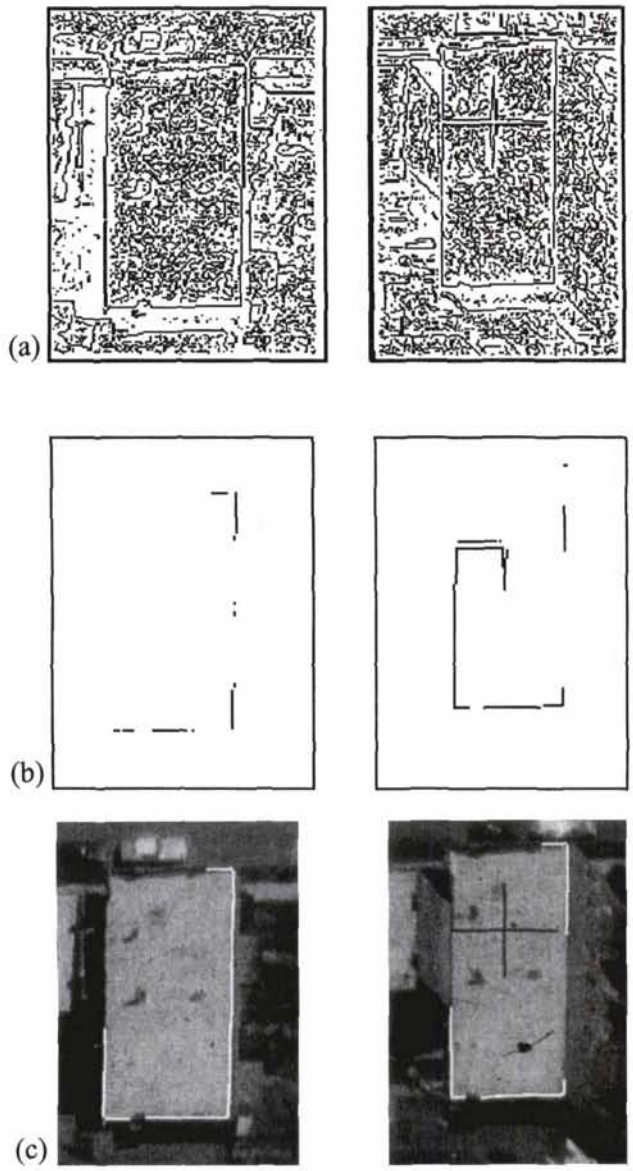

Figure 9. Results for building no. 3: (a) extracted edges; (b) chains after monoscopic processing, (c) resulting corners, overlaid on the image patches.

In Figure 9 an interfering réseau mark appears on the building roof, resulting in edges that were joined to form additional chains. The chains that were found on the right image are triplets (and not full contour as it may seem), and within the stereo processing the outline of the building was selected correctly. Nevertheless, the right lower and right upper edges in the right image are displaced.

The resulting 3D coordinates of the corners were compared with manually digitized coordinates of the building corners. These were digitized interactively on a photogrammetric softcopy station. The differences are summarized in the histograms depicted in Figures 10 and 11.

From the histograms, it is seen that approximately 50 percent of the corners in both data sets and in both the *xy* and *xyz* coordinates are displaced by less than 10 cm. More than 80 percent are displaced by less than 20 cm. The larger differences were found in locations where there are interfering objects such as the concrete belt, as described in the previous section.

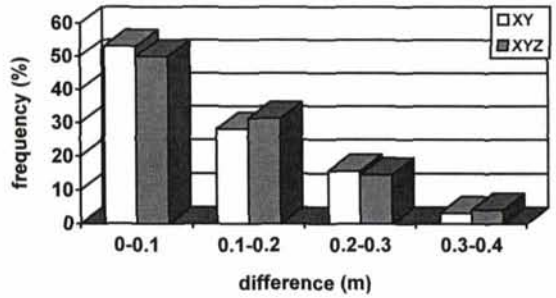

Figure 10. Differences (in metres) in corner locations for the URBAN dataset: *xy* differences on the left column, *xyz* differences on the right column.

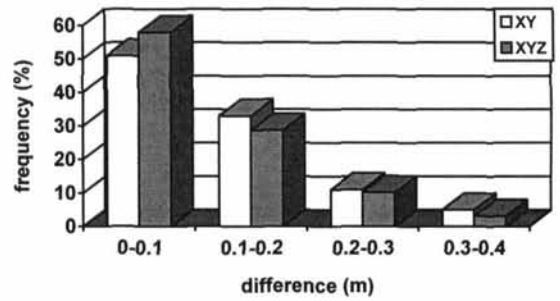

Figure 11. Differences (in metres) in corner locations for the CAMPUS dataset: *xy* differences on the left column, *xyz* differences on the right column.

Conclusions and Future Work

A new concept for building extraction is presented and discussed. The concept is based on a semiautomatic approach, where buildings are detected interactively and 3D building outlines are extracted automatically. The process comprises edge, shadow, and stereo information.

Outlines of seven buildings from two data sets of large-scale images were automatically extracted. Coordinates of extracted corners were compared with coordinates of manually measured points. The results show that approximately 80 percent of the extracted coordinates (in all three dimensions) are within 20 cm of the manually measured coordinates, which corresponds to approximately two pixels in image space.

It should be noted that, in the current implementation, there was no attempt to trace the detected corners back to the images. Compared coordinates are based on the original detected edges that are most likely to be shifted due to interfering objects and other artifacts. A careful corner detection procedure at the detected locations is expected to produce even better results.

The described results are very encouraging. They show that it is possible to detect outlines of buildings accurately and reliably, using an automatic procedure that exploits stereo, shadows, and *a priori* geometric information.

References

Ackermann, F., and P. Krzystek, 1997. Complete automation of digital aerial triangulation, *The Photogrammetric Record*, 15(89): 645–656.

Fradkin, M., and U. Ethrog, 1997. Feature matching for automatic generation of distortionless digital orthophoto, *Proceedings of SPIE, Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision III* (D.M. McKeown, Jr., C.J. McGlone, and O. Jamet, editors), 3072:153–164.

Haala, N., and K.H. Anders, 1997. Acquisition of 3-D urban models by analysis of aerial images, digital surface models and existing 2-D building information, *Proceedings of SPIE, Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision III* (D.M. McKeown, Jr., C.J. McGlone, and O. Jamet, editors), 3072:212–222.

Huertas, A., and R. Nevatia, 1988. Detecting buildings in aerial images, *Computer Vision, Graphics and Image Processing*, 41:131–152.

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

Krzystek, P., and F. Ackermann, 1995. New investigations into the practical performance of automatic DEM generation, *Proceedings, ACSM/ASPRS Annual Convention*, Charlotte, N.C., 2:488–500.

Lang, F., and W. Förstner, 1996. 3D City modeling with a digital one-eye stereo system, *International Archives of Photogrammetry and Remote Sensing*, 31(B4):780–785.

Sahar, L., 1997. *Semiautomatic Extraction of Buildings from Digital Aerial Images*, M.Sc. Thesis, Department of Civil Engineering, Technion — Israel Institute of Technology (in Hebrew).

Sarkar, S., and K.L. Boyer, 1991. Optimal infinite impulse response zero crossing based edge operator, *Computer Vision, Graphics and Image Processing*, 54(2):224–243.

Schenk, T., 1997. Toward automatic aerial triangulation, *ISPRS Journal of Photogrammetry & Remote Sensing*, 52(3):110–121.

Shi, Z., and R. Shibasaki, 1996. Towards automated house detection from digital stereo imagery for GIS database revision, *International Archives of Photogrammetry and Remote Sensing*, 31(B4): 780–785.

Toth, C.K., and A. Krupnik, 1996. Concept, implementation and results of an automatic aerotriangulation system, *Photogrammetric Engineering & Remote Sensing*, 62(6):711–717.

Weidner, U., and W. Förstner, 1995. Towards automatic building extraction from high resolution digital elevation models, *ISPRS Journal of Photogrammetry and Remote Sensing*, 50(4):38–44.

(Received 01 December 1997; accepted 05 June 1998)