Computer Vision and Image Understanding
Vol. 74, No. 2, May, pp. 138–149, 1999
Article ID cviu.1999.0750, available online at http://www.idealibrary.com on IDEAL®

# Automatic Object Extraction from Aerial Imagery—A Survey Focusing on Buildings

Helmut Mayer

*Chair for Photogrammetry and Remote Sensing, Technische Universität München, D-80290 Munich, Germany*
E-mail: helmut@photo.verm.tu-muenchen.de

Received March 31, 1998; accepted January 28, 1999

**This paper surveys the state-of-the-art automatic object extraction techniques from aerial imagery. It focuses on building extraction approaches, which present the majority of the work in this area. After proposing well-defined criteria for their assessment, characteristic approaches are selected and assessed, based on their models and strategies. The assessment gives rise to a combined model and strategy covering the current knowledge in the field. The model comprises: the derivation of characteristic properties from the function of objects; three-dimensional geometry and material properties; scales and levels of abstraction/aggregation; local and global context. The strategy consists of grouping, focusing on different scales, context-based control and generation of evidence from structures of parts, and fusion of data and algorithms. Many ideas which have not been explored in depth lead to promising directions for further research.** © 1999 Academic Press

## 1. INTRODUCTION

Automatic extraction of objects such as buildings or roads from digital aerial imagery is not only scientifically challenging but also of major practical importance for data acquisition and update of geographic information system (GIS) databases or site models.

In this paper, the term "extraction" is used for the detection, as well as for the reconstruction, of objects. "Detection" means that objects are found, based on simpler features and camera models, resulting in simple two- (2D) or three-dimensional (3D) models. On the other hand, for a highly accurate "reconstruction," knowledge about the object's geometry and especially its topology is assumed to be given and more complex camera models, as well as high quality data, are used. The basic reason why the two terms are combined in this paper is their interdependence; the semantics of an object (detection) depends directly on its geometric extent (reconstruction). More practically speaking, some recent approaches give an indication that, only by a precise reconstruction of an object, enough evidence can be achieved to exclude wrong object types.

The paper surveys the state-of-the-art automatic object extraction techniques from (digital) aerial imagery. The survey is extensive, but it does not claim to be complete. Some surveys give an overview of the whole area [1, 2], whereas others review only building extraction techniques [3, 4], or the role of artificial intelligence (AI) [5]. This survey includes approaches for object extraction from satellite images, which influence the extraction from aerial imagery. It only covers models and strategies. Specific algorithms or techniques derived from them are not reviewed, to limit the extent of the survey. Though the combination and integration of the models and strategies with human interaction to build semi-automatic systems [6–11] is of major practical importance, the survey only deals with the automatic parts of the extraction. This is due to the fact that the problems linked to human–computer interaction constitute a challenge in their own right.

The basic idea of the survey is to present a combined model and strategy covering the knowledge in the field. To give insight into how it arose, some approaches for building extraction were selected which are described in more detail and assessed according to some well-defined criteria. The detailed analysis is given only for the building extraction approaches, because they present the majority of the work.

The survey consists of four parts: In Section 2 criteria are introduced which allow for an assessment of the approaches according to their complexity. Then, characteristic approaches, selected according to their relevance at the time of their development, exemplifying various ways to extract buildings from aerial imagery are assessed, starting with the complexity of data an approach can handle (cf. Section 3). The actual assessment is split into a characterization of the models and strategies, as well as a classification of the approaches.

The assessment of the approaches gives rise to the combined model and strategy representing the current knowledge in the field (cf. Section 4). Because of the diversity of the approaches, the combined model and strategy is based on the state-of-the-art derived from buildings and other object types. "Combined" expresses that, at least potentially, all objects which are depicted in the given data are modeled. The section is complemented with outstanding issues whose importance has become clear only recently. Finally, after a short summary several highly promising directions of further research are explored in Section 5.



1077-3142/99 $30.00

## 2. CRITERIA FOR THE ASSESSMENT OF THE APPROACHES

This section proposes criteria, which allow for an assessment of different approaches according to their complexity. After showing the link between assessment and complexity, the data and their complexity are treated in more detail. The criteria for the assessment of the complexity of images, models, and strategies conclude the section.

### 2.1. Assessment and Complexity

The assessment of the approaches is based on the idea of [12] to distinguish strategies according to their suitability for data, i.e., images, and models of different complexity. For aerial imagery this has the consequence that the complexity of their content is relatively high (cf. Subsection 2.2). Additionally, all models, for example, the model for a building, are quite complex. They possess high variability and exhibit many details. Because of this double complexity, the so-called "combined strategy" [12], i.e., a combination of other strategies, is suited best. Nevertheless, not all approaches use the same strategy and therefore the complexity of the strategy is assessed in this paper, too.

The suitability of the combined strategy does not exclude the use of simpler strategies for parts of the problem. In the most simple case a "feature vector classification" of a pixel or a region is enough. If one wants to find, e.g., cars as evidence for a road, possibly fitting simple models to radiometry might be promising.

Another way of assessment which distinguishes a signal- or feature-based representation of the images and a geometry/physics-based or biologic/semantics-based representation of the models, is presented in [13]. It is similar to the one chosen here, as the signal- or feature-based representation of images can be seen as expressing low or high complexity.

### 2.2. Data and Its Complexity

The **complexity** of aerial imagery is mostly due to the large number of different objects depicted. Problems which arise from the dynamics of the scene, e.g., by moving objects, the seasons, etc., are not considered here. The number of objects depends mainly on two factors. The first is *content* (cf. Fig. 1). An old European city center is considerably more complex than a rural part of the midwest of the United States. A further classification of the content according to density (rural, suburban, urban), object complexity (residential, industrial, military), architecture (elaborate, plain, none), terrain (flat, hilly, mountainous), or vegetation (none, moderate, heavy) is not considered in this paper, though it might be useful for future attempts to evaluate the performance of object extraction approaches. The second factor is the *observability* of the different objects. It depends on several prerequisites. The most important is resolution. It can be low, i.e., $>1$ m, medium, i.e., $\geq 0.2$ m and $\leq 1$ m, or high, i.e., $<0.2$ m. Other prerequisites are image quality, in terms of contrast and noise, and the season. Both are not taken into account for the remainder of the paper, as they are assumed to be optimized for the given task.

Since digital aerial imagery is generated in most cases by scanning an analog film, the resolution is also dependent on the image scale which can be small, medium, or large. Typical image scales vary between approximately 1 : 70,000 and about 1 : 4000. A scanning of the latter ones with a resolution of 15 $\mu$m results in a ground resolution of 6 cm. This high resolution is especially needed for the extraction of buildings which may comprise small details such as gutters not visible in larger ground pixels. For other objects such as roads, a ground pixel size of 20 cm or even more is enough which can be gained by scanning images with scales like 1 : 12,000.

Another important prerequisite for observability is the number of images a scene can be found in. Mono, stereo-, and multiple images can be distinguished. The more images an object can be seen in, the better its 3D-geometry can be reconstructed. This is especially important for buildings. What is more, multiple images also reduce problems with occlusion. Color is also very helpful to extract buildings and vegetation. For vegetation in many cases color infrared images are used.

There is a multitude of other data besides aerial imagery which can be used for object extraction. According to its resolution, satellite imagery, for instance, is approaching medium scale aerial imagery. Besides three-line scanners such as MOMS-02/D2 [14] with fore and aft along track stereo, a ground resolution of about 6 m in the nadir view and about 18 m in the forward and backward looking views, as well as multispectral


**FIG. 1.** Resolution and content (a) low resolution ($>1$ m)—simple content; (b) medium resolution ($\geq 0.2$ m and $\leq 1$ m)—medium to complex content; (c) high resolution ($<0.2$ m)—medium to complex content [83].

**TABLE 1**

**Visual Recognizability of Various Object Types Depending on the Minimum Ground Pixel Size and Mono/Stereo in Satellite Images Based on [14]**

| Object type | Ground pixel size and Mono/stereo |
|---|---|
| Building | 2 m, stereo |
| Path | 2 m, stereo for occlusion by vegetation |
| Minor road | 5 m, stereo for occlusion by vegetation |
| Hydrology | 5 m, stereo for occlusion by vegetation |
| Main road | 10 m |
| Building block | 10 m |

capabilities, sensors with less than 1-m ground resolution for the pan-chromatic and less than 4 m for the multispectral channels are planned for the near future [15]. Especially the latter can be more suitable for some applications than aerial imagery.

Extremely interesting for an automatic interpretation are approaches which directly measure range/height data by means of laser scanning [16]. The ground resolution can be down to 0.25 m with a precision for the height of about 0.1 m.

For an empirical examination of the complexity, content and resolution are linked. The question is: Which objects can be mapped at which resolution? Basically, analogously to the Nyquist theorem, an object has to be sampled with a spatial resolution which is half the size of the object to be distinguished from other objects. Though much smaller objects can be seen when their contrast to the surroundings is relatively high. More specifically, [14] presents the visual recognizability of different object types for satellite imagery, depending on the minimum ground pixel size (cf. Table 1) which should be mostly transferable to aerial imagery. Recognizability means that the location, as well as the object type, can be determined. For acquiring GIS data there are higher demands concerning resolution if attributes of the objects are to be acquired, too.

In summary, the complexity of images depends on the scene content and on the observability of objects and, therefore, on the resolution. Avoiding unnecessary complexity which could disturb the extraction is especially important for automatic extraction. Basically, there is a minimum complexity to solve the problem (cf. Fig. 2). If the problem is solvable, the complexity may not be optimum in two ways: either the optimum solution is not reached, or it is achieved inefficiently by using too complex data, model, or strategy.

Since the complexity of the data depends strongly on the resolution, the latter should be chosen sufficiently small so that the important details of the objects can be recognized. It is also possible that there is more than one optimum solution. For roads, a coarse resolution of about 10 m is, for instance, ideal to detect hypotheses for highways, a medium resolution of 3 to 5 m to detect the different traffic lanes, and a high resolution down to 10 cm to extract the borders of the pavements and the markings. In [17] it is shown how, based on different scale-spaces [18], values for the resolution/scales can be determined analytically. What is more, it is also shown analytically how disturbing objects like cars can be eliminated from the road by means of scale-spaces.

### 2.3. *Complexity Criteria for the Assessment of Images, Models, and Strategies*

Before presenting the complexity criteria, the criticism of the current state-of-the-art image analysis [19–21] is considered. Its combination with the more specific criticism of the approaches for object extraction from aerial imagery in [22] and the postulated “enlarging the peephole,” in terms of spatial, spectral, temporal, and contextual components of [23], results in the following points of special importance:

1. The performance of the extraction should be evaluated for general validity using as many images as possible. This should not only be done visually, but also based on performance measures.
2. Spatial resolution should be appropriate for the problem.
3. Knowledge is important as not all information is contained in the image.
4. The extraction should be done in object-space, as only there can much of the knowledge about the real world be used. It is, for instance, hard to compare the width of a road in pixels with the model. For this, knowledge about the sensor and its orientation is a must.
5. It is better to integrate different kinds of methods than to use only one technique for the solution of the problem: “multiple

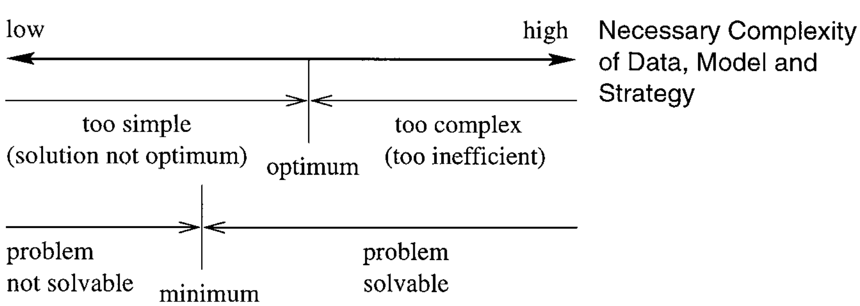


**FIG. 2.** Necessary complexity of data, model, and strategy.

**TABLE 2**
**Criteria for the Assessment of Models and Strategies**

| Criterion | Assessment |
|---|---|
| Complexity of data | Low / medium / high |
| Resolution | Low / medium / high |
| Content | Simple / medium / complex |
| Complexity of model | Low / medium / high |
| Representation formalisms | Implicit / type |
| Geometry and radiometry | Geometric + / radiometric |
| Kind of representation | 2D/3D; fixed values / parametric / generic |
| Sensor model | Simplified / detailed |
| Object model | Weak / medium / detailed |
| Scene model | None / weak / medium / detailed |
| Function | Not modeled / weak / strong; implicit / explicit |
| Complexity of strategy | Low / medium / high |
| Fusion | Type |
| Grouping | 2D + / 3D; simple / complex |
| Control | Type |
| Search | Type |
| Evaluation | None / simple / complex; type |

*Note.* ";" separates different assessments; "+" means "and"; "/" means "or."

methods are found along the path to enlightenment; there are no silver bullets" [22]. The use of spectral or temporal information is of special importance.

6. Object extraction should be done in a highly integrated fashion. More specifically, the context of the extraction should be adapted; as soon as new information is available it is used to simplify the extraction.

The first and the second points correspond to the key point *complexity of data* (an approach can handle) in Table 2. It is subdivided into resolution and content, according to Subsection 2.2. The first point is more implicitly contained, as the performance of object extraction is admittedly of first interest for the application, but it is not evaluated for most approaches up to now.

The third point agrees with the key point *complexity of model*. It is split first into the representation formalisms ("implicit" in Table 2 means that the knowledge is hidden in the program code) and the focusing on geometry and/or radiometry. The last two are, according to [23], important ways to use spatial knowledge. Further points are the type of representation formalism and the sensor model. Only by means of a detailed sensor model is the transition from image to object space addressed in the fourth point possible. By clarifying the degree of detail of the scene and of the object model, deficits of the model are shown which are due to the restriction to only one or a few object types and their relations, or due to a too unspecific modeling of the individual objects. A modeling of the function of the objects which is done as explicitly as possibly can help for a further improvement of the extraction.

The fifth and the sixth points both belong to the key point *complexity of strategy*. The integration of methods corresponds to the fusion of different techniques or data. Grouping is quite common for many approaches. Object extraction done in a highly integrated manner corresponds to the different levels of complexity and types of control and search. The criteria for the assessment of the approaches are concluded by the complexity of the (internal) evaluation.

## 3. SELECTED APPROACHES FOR BUILDING EXTRACTION AND THEIR ASSESSMENT

The goal of this section is to introduce and assess several characteristic examples of approaches for building extraction from aerial imagery. Together with the presentation of the approaches the complexity of data they can handle is assessed (cf. Subsection 3.1). Only then, the "relative" assessment of the approaches is presented in Subsections 3.2 and 3.3, according to the *complexity of model* and the *complexity of strategy* as motivated in Subsection 2.3. "Relative" means that only the comparison between the approaches is important. The assessment according to an "absolute" goal is carried out by a classification of the models and strategies (cf. Subsection 3.4), only so far as a very coarse estimate of the distance to the most important goal, the extraction of objects for GIS databases or site models, is given.

### *3.1. Complexity of Data*

Herman and Kanade's approach [24] was selected as it is an often cited early approach for the extraction of buildings which uses AI-focused 3D-reasoning, in combination with heuristics about the vertical and horizontal directions of lines to extract buildings as rectangular prisms. Some years later buildings still were modeled as rectangular prisms, but the 3D-structure was generated by matching higher level structures such as rectangles in stereo images found by grouping rectangular or parallel edges [25]. What can be done in single, possibly oblique, images to extract flat and peaked roof buildings using shadows and visible vertical edges based on vanishing points is exemplified by [26]. Wang *et al.* [27] demonstrate that the semantics of buildings is not restricted to the geometry. From several images optimally rectified images of walls are calculated and used to extract part structures of windows and doors in the walls which can help to raise the probability of the extraction. That prismatic as well as parameterized buildings, like peaked ones, can be detected and extracted from a digital surface model (DSM) is shown by [28]. Some of the most recent and sophisticated approaches match primitives such as edges or corners in several images, based on a detailed image model, and a complex strategy as presented by [29, 30]. Whereas [30] only tries to find parts of the roof, in [29] a generic model for buildings was elaborated which consists of a combination of parameterized building parts. Table 3 gives the complexity of data an approach can handle that is split, according to Subsection 2.3, into resolution and content. It can be seen that the most recent approaches are the ones which can handle the most complicated scenes.

**TABLE 3**
**Complexity of Data for Building Extraction**

| Approach | Complexity of data | Resolution | Content |
|---|---|---|---|
| 3D reasoning, Herman and Kanade, 1984 [24] | Low–medium | Medium | Simple |
| Matching of grouping-based structures in stereo images, Mohan and Nevatia, 1989 [25] | Medium | Medium | Medium |
| 3D-interpretation of mono images using shadows and, vertical edges, Shufelt, 1996 [26] | Medium | Medium | Medium |
| Evidence from building parts, Wang *et al.*, 1997 [27] | Medium | Medium | Medium |
| Building extraction from digital surface models, Weidner, 1997 [28] | Medium (DSM) | Medium | Simple–medium |
| Matching of primitives in multiimages, Fischer *et al.*, 1998 [29] | High | High | Medium–complex |
| Matching of primitives in multiimages, Henricsson, 1998 [30] | High | High | Medium–complex |

### 3.2. *Characterization of Models*

The models for building extraction show only a weak tendency to use knowledge-based representation formalisms (cf. Table 4). Concerning radiometry and geometry there is a significant trend to use edges, i.e., a geometric representation, for the extraction. Since buildings are 3D objects, mostly 3D representations, partly supported by 2D representations, are exploited. Generic, as well as parametric models, are utilized. While the former have the advantage of generality, for the latter it can be more easily checked whether the extracted object is a building. For the first approaches the sensor model was often simplified, and it is only a recent tendency to use a detailed photogrammetric sensor model for 3D reconstruction. The object models are becoming more and more detailed, based on complex, generic, and parametric 3D structures (cf. Table 5) and they are partially extended by part structures. The height of the buildings is modeled in oblique views by an object model comprising vertical walls and in nadir-looking views with a scene model including shadows. Shadows are considered to be a part of the scene model, as they include information beyond the object. The function of objects is only used implicitly when gaining evidence from the structure of building parts, such as doors, windows, or vents.

In summary there is an evolution from general techniques to approaches customized for the object type, i.e., the building. It is considered to be very important to utilize the specific knowledge, i.e., the models and strategies, as completely as possible. Though, if it is possible without too much effort, an approach will still be designed for as many object types as possible. Overall there is also a trend to focus on 3D geometry.

### 3.3. *Characterization of Strategies*

Many approaches fuse two (stereo) or more images (cf. Table 6). Additionally in [24] multitemporal images are used. This leads to the problem of updating, for which the approach taken is relatively simple. In [30] colored images and a DSM

**TABLE 4**
**Characterization of the Models for Building Extraction I**

| Approach | Representation formalisms | Geometry and radiometry | Kind of representation | Sensor model |
|---|---|---|---|---|
| 3D reasoning, Herman and Kanade, 1984 [24] | Implicit | Geometric | 3D; generic (rectangular prism) | Simplified |
| Matching of grouping-based structures in stereo images, Mohan and Nevatia, 1989 [25] | Constraints | Geometric | 2D + 3D; generic (rectangular prism) | Simplified |
| 3D interpretation of mono images using shadows and vertical edges, Shufelt, 1996 [26] | Implicit | Geometric | 3D; parametric (flat + peaked roof) | Detailed |
| Evidence from building-parts, Wang *et al.*, 1997 [27] | Implicit | Geo- + radiometric | 2D + 3D; walls | Detailed |
| Building extraction from digital surface models, Weidner, 1997 [28] | Implicit | Geometric | 3D; parametric + generic | — |
| Matching of primitives in multi-images, Fischer *et al.*, 1998 [29] | Constraints | Geometric | 3D; generic, based on basic types | Detailed |
| Matching of primitives in multi-images, Henricsson, 1998 [30] | Implicit | Geometric | 3D; generic (roof) | Detailed |

**TABLE 5**
**Characterization of the Models for Building Extraction II**

| Approach | Object model | Scene model | Function |
|---|---|---|---|
| 3D reasoning, Herman and Kanade, 1984 [24] | Weak (prism) | None | Not modeled |
| Matching of grouping-based structures in stereo images, Mohan and Nevatia, 1989 [25] | Weak (prism) | None | Not modeled |
| 3D interpretation of mono images using shadows and vertical edges, Shufelt, 1996 [26] | Medium (flat + peaked roof, wall) | Weak (shadow) | Not modeled |
| Evidence from building parts, Wang *et al.*, 1997 [27] | Medium (3D blocks + structures of parts) | Weak (shadow) | Weak; implicit (doors, windows) |
| Building extraction from digital surface models, Weidner, 1997 [28] | Medium (standard building types + complex flat roofed) | None | Not modeled |
| Matching of primitives in multi-images, Fischer *et al.*, 1998 [29] | Detailed (complex 3D structure, aspects) | None | Not modeled |
| Matching of primitives in multi-images, Henricsson, 1998 [30] | Detailed (complex 3D structure) | None | Not modeled |

[31] are exploited to generate hypotheses for buildings. Grouping is of significant importance and relatively complex schemes are used. It is done in the 2D image as well as in the 3D scene and [30] even combines both. The control of the approaches is mostly data-driven. Only [29] is based on "hypothesize and verify," where grouping is used for the hypotheses. Wang *et al.*'s approach [27] is model-driven as the walls can only be extracted from the image using already existing information. Constraint satisfaction is used in [25, 29]. If it is there at all, internal evaluation is mostly simple. Only [29, 28] use a more complex evaluation, based on probability theory or minimum description length (MDL).

In summary there is a change from simpler to more complex strategies. There is a tendency to use more than two images, also color images or DSM. Grouping is focused on. The trend is toward complex grouping in 2D as well as in 3D. The tendency

**TABLE 6**
**Characterization of the Strategies for Building Extraction**

| Approach | Fusion | Grouping | Control/search | Evaluation |
|---|---|---|---|---|
| 3D reasoning, Herman and Kanade, 1984 [24] | Stereo at different times | 3D, simple | Data-driven; — | No |
| Matching of grouping-based structures in stereo images, Mohan and Nevatia, 1989 [25] | Stereo | 2D; complex | Data-driven; constraint satisfaction | Simple |
| 3D interpretation of mono images using shadows and vertical edges, Shufelt, 1996 [26] | — | 3D; complex | Data-driven; — | Simple |
| Evidence from building parts, Wang *et al.*, 1997 [27] | Several images | — | Model-driven; — | No |
| Building extraction from digital surface models, Weidner, 1997 [28] | — | — | Data-driven; — | Complex; MDL |
| Matching of primitives in multi-images, Fischer *et al.*, 1998 [29] | Several images | 3D; complex | Hypothesize and verify; constraint satisfaction | Complex |
| Matching of primitives in multi-images, Henricsson, 1998 [30] | Several images, color, DSM | 2D + 3D; complex | Data-driven; — | Simple |

goes toward mixed strategies. Some recent approaches use the more general strategy of hypothesize and verify. For internal evaluation more and more complex modeling is used.

### 3.4. Classification of Models and Strategies

The classification of the different approaches regarding their models and strategies is done in two ways: first, by comparing it with the ultimate goal, the extraction of objects for GIS databases or for site models, and second, by showing the most sophisticated approaches.

Compared to the ultimate goal, the results based on the models and strategies of all presented approaches are for images of different characteristics and complex contents far from being useful in practice. Nevertheless, for restricted domains the most recent approaches are getting closer to being useful in practice. This might be the reason why quantitative performance evaluation has received much interest recently [26, 32–34]. As is to be expected, the most sophisticated approaches for building extraction are the newest ones [29, 30].

## 4. A COMBINED MODEL AND STRATEGY FOR OBJECT EXTRACTION IN AERIAL IMAGERY

The combined model and strategy condenses, on one hand, the results of the assessment in Section 3. On the other hand, knowledge about the extraction of object types other than buildings is included here, because there is a big overlap in the model, as well as in the strategy, and the overlap can be used to widen the scope without too much effort. The combined model and strategy is subdivided into model and strategy (cf. Subsections 4.1 and 4.2). References to additional approaches are given which utilize a part of the model or the strategy especially well. Their order shows how well they make use of it in relation to each other. Finally, in Subsection 4.3 outstanding issues for the combined model and strategy are presented whose importance has become clear only recently.

### 4.1. Model

The model is organized into general parts and specific parts for buildings. The fact that the general parts are much larger than the specific parts illustrates that the model is for the most part generic.

The **general parts of the model** are:

- *Characteristic properties are often the consequence of the function of objects* [35, 36]. Very importantly, they integrate knowledge about the 3D real world into the model. Typical examples for knowledge sources are, apart from constraints concerning the usefulness for humans [37], construction instructions for different types of buildings or roads. For large parts of the knowledge about function it seems to be enough to take them into consideration for modeling. I.e., it is not necessary to integrate them into the system.
- *Modeling of material properties* [38] causes the interpretation to be not so much affected by sensor characteristics. This is especially important when different kinds of sensors, like optical and radar, are utilized.
- *2D geometric/topologic regularities* [39, 26, 33, 25, 40] are used for buildings as well as for road pavements (parallel edges). For buildings they are especially suited for simple types, e.g., with perpendicular outlines and flat roofs, and the construction of hypotheses for more complex buildings. Though, for modeling of the latter ones they are not sophisticated enough.
- *A detailed image model* [4, 41, 30, 42–44] with rich attributes and a feature adjacency graph exploits much better the information contained in the image. Especially for matching in more than one image, the probability can be raised significantly, compared to approaches based only on one feature type (mostly edges).
- *Levels of abstraction and scale* [35, 45–47] are especially useful for roads and vegetation but might be also useful to detect building blocks. For roads in coarse scale, many disturbances are eliminated (cf. Subsection 2.2). The elimination of disturbances helps to bridge gaps and to get a more complete road network. Coarse scales can be generated artificially using scale-spaces. The local modeling of single objects such as trees by using appearance-based approaches [48] is exploited in fine scale. This avoids the transition from image space to object space.
- *The geometric/topologic neighborhood* [49, 50, 38, 51], i.e., the spatial context, describes the spatial arrangement of objects. For example, shadows can be used to detect buildings. Intersections or cars have a direct relation to a road, whereas trees or buildings are needed because they cast shadows or occlude roads on one hand and because they form rows parallel to the road, on the other hand.
- *Global and local context* [35, 52] subdivides the geometric/topologic neighborhood by a spatial partitioning. Very local structures like a tree casting a shadow or a building occluding a road are distinguished from global structures such as *suburb_urban*, *forest*, or *open_rural*. The latter ones restrict the frequency and the characteristics of the former ones: For instance, in *open_rural* areas there are only few buildings, located often well separated.
- *Structures of parts* [27, 53, 35, 49, 54], also called substructures, can be used as local evidence for objects. Typical examples comprise cars on the road, doors or windows in the wall, as well as dormer windows or vents on the roof. These objects show a characteristic arrangement with each other and regarding the object they are part of. Often it is useful to rectify the images before the extraction of the part structures to get a standardized situation.
- *Statistic modeling* [55, 56] extends the widely used, more or less functional and deterministic modeling. With probabilistic methods the uncertainty of the data as well as of the model can be propagated and used for controling the analysis.

**Specific parts of the model for building extraction** comprise:

• *Shadows* [34, 57, 26, 58] and *walls (vertical edges)* [26, 34, 57, 59] are very good evidence for the 3D interpretation of mono images. Nevertheless, in some cases there are problems: for shadows they emerge from nonplanar terrain close to the building, from shadows cast on buildings close-by, or from the fact that shadows could be occluded by the object itself in oblique views. On the other hand, vertical walls are mainly visible in oblique views but can be occluded by other objects. Nevertheless, it is not clear, why shadows and, especially, vertical edges are used so seldom for building extraction from two or more images.

• *The 3D geometry in ≥2 images* [29, 30, 42, 33, 60–63], based on a camera model and given orientations, gives a valuable indication for the existence of the 3D structure characteristic for buildings. The more images from different directions are used, the higher is the chance to exclude wrong matches.

• *A generic 3D model* [29] which consists of surfaces and a constructive solid geometry (CSG) modeling may be the best starting point for a more generic building extraction. All other representations either cannot describe a complex building properly, or it seems difficult to decide that a structure cannot be a building [30]. As a restrictive comment it has to be added that it is not yet clear if there are structures which "cannot be a building."

• *Aspects* [4, 29] derived from a generic description consisting of building terminals and connectors allow for a parallel modeling in a 2D image and a 3D object model by enabling an explicit transition between these two.

### 4.2. Strategy

The parts of the strategy are organized in the same way as in Subsection 4.1.

The **general parts of the strategy** are:

• *Appearance-based approaches* [53, 45] avoid the explicit transition from an image model to an object model [48]. With them, objects like trees can be extracted, for which the modeling is quite difficult by other means because they have a relatively varying appearance. They also can be used for the extraction of details on the rooftop or cars on roads when the resolution is close to the point where these objects cannot be extracted any more.

• *Grouping* [39, 26, 33, 25, 64, 40], i.e., the search for geometric/topologic regularities, allows for focusing on parts of objects and therefore limits the search space. An often encountered problem is that the regularities specified are not strict enough to ensure a reliable extraction. Grouping should therefore be accompanied by verification.

• By means of the *focus on different scales* [35, 65] the extraction is at the same time accelerated, as well as improved. By using multiple scales one can start with reliable structures in coarse scale and use them to focus the extraction on the specific areas and object types in fine scale. In many cases, instead of changing the scale in the image by means of scale-spaces [18], image pyramids can be used. This significantly accelerates the processing of the small scale.

• *Hypotheses generation and search/resegmentation based on spatial context* [49, 1] is done by predicting an object given another object with a spatial relation to it. Many objects receive their semantics only in this way, which is especially true if they cannot be recognized, or are at least hard to extract by themselves.

• When *focusing on contexts* [35] the distinction between global and local context is used for a further improvement of hypotheses generation and search/resegmentation. In many cases it is useful to first segment the image into the global contexts and only then to start the extraction of the objects in the easiest or most promising global context. For roads, these are, for instance, the *open_rural* areas, in which the objects and the local contexts made up from them are analyzed. Objects in the local context such as trees or shadows can prevent the extraction of roads. Other objects such as cars can help to validate roads.

• The *generation of evidence from structures of parts/substructures* [53, 35, 27, 49, 54] improves the probability of hypotheses. Here, it is assumed that part structures cannot be extracted directly in many cases. However, if there is an hypothesis about the object to be extracted, its spatial constraint makes the extraction of the part structures possible. For buildings, single objects such as doors or windows cannot be interpreted by themselves (e.g., black blobs), but their arrangement makes their semantics and, at the same time, the semantics of the object itself clear.

• *Balancing image information versus the geometric model in an automatic process* [8, 66, 42, 67] enables a geometric improvement of objects with already clear semantics, but weakly defined outlines. Typical examples are snakes [68] or "model-based optimization" [8]. Recent results on the extraction of roads in shadowed regions [66] show that snakes are also useful to extract objects when only a stabilized geometry makes the extraction of useful image features possible.

• The *fusion of data and of algorithms* [49, 28, 30, 42] comprises not only color but also multispectral images and images from different sensors. Although, the color in images is not stable due to the indirect lighting of shadowed objects, color images are, for instance, useful for limiting the search space of building extraction by using the fact that many roofs are red. Additionally, there can be, more or less, unexpected colors of the roofs, for example, old green copper roofs. The fusion of algorithms is a very general technique which can be used for different areas. Examples are the treatment of scale-transitions or of different kinds of viewpoint of the image function, e.g., various types of definition of edges or regions.

The **specific parts of the strategy for buildings** consist of:

• By *matching of primitives in several images* [29, 30, 42, 33, 61, 62, 43, 69] valuable information about the 3D geometry of parts of buildings can be gained, especially when using many images from different directions. To get a good approximation, and therefore to improve the probability of the matching, a DSM should be used when available. Methods to produce dense and reliable DSM for urban areas by image matching are, e.g., [70, 71].

- The *use of aspects* [29], i.e., the sequence 3D points ⇒ building part ⇒ building ⇒ matching of building parts to image primitives, allows for a direct transition from image to object model and vice versa. A modeling based on aspects as described in Subsection 4.1 is a prerequisite for this.
- The results of the *extraction of hypotheses for buildings from DSM* [28, 72, 70, 16, 73–78] are a little imprecise and unreliable, especially as vegetation is often mistaken for buildings. Therefore, the results are used for applications with not too high requirements, such as telecommunication planning or as robust approximation for further extraction using image information.

### *4.3. Outstanding Issues for Model and Strategy*

The importance of three outstanding issues has become clear only recently:

First, **scale** or resolution of an image are not only important due to their link to the observability of objects or parts of objects. More importantly, by the abstraction in the coarse scales generated by scale-spaces, features in the image such as lines can be linked directly to objects like roads [65, 35, 17].

Second, the **context** and its spatial organization renders a highly effective means to impose structure on the knowledge [38, 35, 52]. This makes it possible to construct large consistent models and strategies tackling the complexity of the objects.

Third, the **3D structure** of objects such as vegetation and especially buildings is the key to their recognition. There are two ways which ideally should be combined:

- DSM are quite effective for detecting vegetation and buildings [28, 70, 73]. The reliability is improved considerably if a DSM from active laser scanning is used [72, 16, 79].
- The matching of features in more than two passive images using the information from the DSM as an approximation and a detailed image model results in highly reliable 3D structures [41, 30]. These can be combined with the knowledge from the model [29].

## 5. SUMMARY AND DIRECTIONS FOR FURTHER RESEARCH

This paper has surveyed the state-of-the-art automatic object extraction techniques from aerial imagery. Characteristic approaches were assessed, based on their models and strategies using well-defined criteria. From the assessment a model and a strategy representing the knowledge in the field were condensed.

Nevertheless, there are many ideas which have not been explored in depth in the combined model and strategy, pointing to several promising directions for further research:

- Different kinds of objects should be investigated concerning their scale-space behavior in various scale-spaces. Characteristic textures in different scales are a closely related phenomenon.
- Local contexts can be modeled more in depth. It could be analyzed which objects should be treated, which relations are of highest importance, and how objects and relations should be formalized best.
- For the structures of parts it could be investigated how relevant details such as cars, dormer windows, doors, and windows are for their extraction, as well as how important they are for the verification of hypotheses.
- The function of objects should be used more explicitly. This could be done, for instance, by autonomous agents, simulating pedestrians or cars, which are checking the functional plausibility of the result of the extraction.
- Concerning fusion it could be investigated which additional sensors are useful. The investigation should be accompanied by a further exploration of sophisticated algorithms, especially evaluating their performance for different applications.
- Other sensors could aid to make more use of the material properties. In particular, data from future imaging, possibly multispectral laser scanners is of interest [79], as they make available reflection information, more or less, independent of lighting conditions.
- To improve the versatility of object extraction, machine learning techniques [80], such as evolutionary algorithms [81] or storing sets of parameters, e.g., for textures, or widths of roads come to mind. An example for learning parts of the model for buildings is presented in [82]. The reference data needed for learning could come from a GIS. Storing the sets of parameters could, for instance, be made dependent on local or global context and on even more general types of surroundings, like rural areas close to the mountains and industrial landscape. For the time being it is assumed that the basic structures have to be formulated before learning can help with the details.
- The variability of the objects is treated insufficiently. That is, most approaches assume that an object has only one kind of appearance. For example, buildings are mostly assumed to be polyhedra. Roof gardens or highly nonplanar shapes of buildings designed by modern architects are not considered. Here, a "multimodel" is proposed for the appearance which takes this variability into account. A strategy might be to tackle the appearance which is most prominent in an area first and only to use other ones if this fails.
- Of utmost importance is a more detailed modeling. To extract more general types of buildings, not only dormer windows or gutters, but also knowledge about architecture should be incorporated. For a more detailed modeling, again, machine learning could be used (see above).
- Internal evaluation is closely related to the previous item. In many cases it is not enough to construct a deterministic model, but it has to be complemented by a statistical component, i.e., probabilistic methods, to decide about the semantics of the objects. For every object and its relations there must be a separate evaluation. For determining the a priori and conditional probabilities, again, machine learning could be helpful.

- Apart from its importance for the decision about the semantics the internal evaluation of objects is also needed for control. It depends not only on the knowledge about the local or global context (model and strategy) but also on the data, i.e., the objects to be extracted and their surroundings. In an addition to the missing evaluation of the individual local or global contexts, it is not clear how—given these quite complex conditions—a consistent evaluation can be achieved at all.

Altogether, there are several promising directions to go, maybe some which have not been thought of yet. Data from imaging laser scanners have opened up ways which few people could imagine some years ago. Since object extraction in aerial imagery has received much attraction recently, there is a chance to make good progress and to reach the point where automatic object extraction becomes feasible for practical applications.

## ACKNOWLEDGMENTS

I express my gratitude to Professor Heinrich Ebner for always encouraging my work. To Professor Wolfgang Förstner I am grateful for innumerable helpful comments and great ideas. I also thank the anonymous reviewers for their helpful remarks.

## REFERENCES

1. T. Matsuyama and V. S.-S. Hwang, *SIGMA—A Knowledge-Based Aerial Image Understanding System*, Plenum, New York, 1990.
2. E. Gülch, Cartographic features from digital images, in *Proceedings, Contributions to 2nd Course in Digital Photogrammetry, Bonn*, Royal Institute of Technology, Department of Geodesy and Photogrammetry, Stockholm, Sweden, 1995.
3. V. Venkateswar and R. Chellappa, *Intelligent Interpretation of Aerial Images*, Technical Report USC-SIPI-137, Signal and Image Processing Institute, University of Southern California, Los Angeles, CA, 1989.
4. C. Braun, T. H. Kolbe, F. Lang, W. Schickler, A. B. Cremers, W. Förstner, and L. Plümer, Models for photogrammetric building reconstruction, *Comput. & Graphics* **19**(1), 1995, 109–118.
5. F. Ade, The role of artificial intelligence in the reconstruction of man-made objects from aerial images, in *Proceedings, Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*, pp. 23–32, Birkhäuser, Basel, Switzerland, 1997.
6. S. Heuel and R. Nevatia, Including interaction in an automated modelling system, in *Proceedings, IEEE International Symposium on Computer Vision, 1995*, pp. 383–388.
7. S. Airault, O. Jamet, and F. Leymarie, From manual to automatic stereoplotting: Evaluation of different road network capture processes, in *Proceedings, International Archives of Photogrammetry and Remote Sensing, 1996*, Vol. (31) B3/III, pp. 14–17.
8. P. Fua and C. Brechbühler, Imposing hard constraints on soft snakes, in *Proceedings, Fourth European Conference on Computer Vision, 1996*, Vol. II, pp. 495–506.
9. A. Grün and H. Dan, TOBAGO—A topology builder for the automated generation of building models, in *Proceedings, Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*, pp. 149–160, Birkhäuser, Basel, Switzerland, 1997.
10. E. Gülch, Application of semi-automatic building acquisition, in *Proceedings, Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*, pp. 129–138, Birkhäuser, Basel, Switzerland, 1997.
11. J. J. Pearson and L. A. Oddo, A testbed for the evalution of feature extraction techniques in a time constrained environment, in *Proceedings, Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*, pp. 13–22, Birkhäuser, Basel, Switzerland, 1997.
12. P. Suetens, P. Fua, and A. J. Hanson, Computational strategies for object recognition, *ACM Comput. Surv.* **24**(1), 1992, 5–60.
13. W. Förstner, A future of photogrammetric research, *NGT Geodesia* **93-8**, 1993, 372–383.
14. G. Konecny and J. Schiewe, Mapping from digital satellite image data with special reference to MOMS-02, *ISPRS J. Photogrammetry & Remote Sensing* **51**, 1996, 173–181.
15. L. W. Fritz, The era of commercial earth observation satellites, *Photogrammetric Engineering & Remote Sensing* **62**(1), 1996, 39–45.
16. P. Axelsson, Integrated sensors for improved 3D interpretation, in *Proceedings, International Archives of Photogrammetry and Remote Sensing, 1998*, Vol. (32) 4/1, pp. 27–34.
17. H. Mayer and C. Steger, Scale-space events and their link to abstraction for road extraction, *ISPRS J. Photogrammetry & Remote Sensing* **53**, 1998, 62–75.
18. T. Lindeberg, *Scale-Space Theory in Computer Vision*, Kluwer Academic, Boston, MA, 1994.
19. R. C. Jain and T. Binford, Ignorance, myopia and naiveté in computer vision systems, *Computer Vision, Graphics, and Image Processing: Image Understanding* **53**(1), 1991, 112–117.
20. T. Pavlidis, Why progress in machine vision is so slow, *Pattern Recognition* **13**, 1992, 221–225.
21. J.-M. Jolion, Computer vision methodologies, *Computer Vision, Graphics, and Image Processing: Image Understanding* **59**(1), 1994, 53–71.
22. D. M. McKeown, *Top Ten Lessons Learned in Automated Cartography*, Technical Report CMU-CS-96-110, Digital Mapping Laboratory, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, 1996.
23. T. Strat, Advancing computer vision through advances in photogrammetry, *Z. Photogramm. Fernerkundung* **5/94**, 1994, 151–160.
24. M. Herman and T. Kanade, The 3D MOSAIC scene understanding system: Incremental reconstruction of 3D scenes from complex images, in *Proceedings, Image Understanding Workshop, 1984*, pp. 137–148.
25. R. Mohan and R. Nevatia, Using perceptual organization to extract 3-D structures, *IEEE Trans. Pattern Anal. Mach. Intell.* **11**(11), 1989, 1121–1139.
26. J. A. Shufelt, exploiting photogrammetric methods for building extraction in aerial images, in *Proceedings, International Archives of Photogrammetry and Remote Sensing, 1996*, Vol. (31) B6/VI, pp. 74–79.
27. X. Wang, J. Lim, R. T. Collins, and A. R. Hanson, Extracting surface textures and microstructures from multiple aerial images, in *Proceedings, Computer Vision and Pattern Recognition, 1997*, pp. 301–306.
28. U. Weidner, digital surface models for building extraction, in *Proceedings, Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*, pp. 193–202, Birkhäuser, Basel, Switzerland, 1997.
29. A. Fischer, T. H. Kolbe, F. Lang, A. B. Cremers, W. Förstner, L. Plümer, and V. Steinhage, Extracting buildings from aerial images using hierarchical aggregation in 2D and 3D, *Computer Vision and Image Understanding* **72**(2), 1998, 185–203.
30. O. Henricsson, The role of color attributes and similarity grouping in 3-D building reconstruction, *Computer Vision and Image Understanding* **72**(2), 1998, 163–184.
31. O. Henricsson, *Analysis of Image Structures Using Color Attributes and Similarity Relations*, Mitteilung 59, Institut für Geodäsie und Photogrammetrie an der Eidgenössischen Technischen Hochschule Zürich, 1996.
32. O. Henricsson and E. Baltsavias, 3D building reconstruction with ARUBA: A qualitative and quantitative evaluation, in *Proceedings,*

*Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*, pp. 65–76, Birkhäuser, Basel, Switzerland, 1997.

33. R. T. Collins, C. O. Jaynes, Y.-Q. Cheng, X. Wang, F. Stolle, E. M. Riseman, and A. R. Hanson, The ascender system: automated site modeling from multiple aerial images, *Computer Vision and Image Understanding* **72**(2), 1998, 143–162.
34. C. Lin and R. Nevatia, Building detection and description from a single intensity image, *Computer Vision and Image Understanding* **72**(2), 1998, 101–121.
35. A. Baumgartner, C. Steger, H. Mayer, and W. Eckstein, Multi-resolution, semantic objects, and context for road extraction, in *Proceedings, Semantic Modeling for the Acquisition of Topographic Information from Images and Maps*, pp. 140–156, Birkhäuser, Basel, Switzerland, 1997.
36. M. de Gunst and G. Vosselman, A semantic road model for aerial image interpretation, in *Proceedings, Semantic Modeling for the Acquisition of Topographic Information from Images and Maps*, pp. 107–122, Birkhäuser, Basel, Switzerland, 1997.
37. L. Stark and K. Bowyer, Achieving generalized object recognition through reasoning about association of function to structure, *IEEE Trans. Pattern Anal. Mach. Intell.* **13**(10), 1991, 1097–1104.
38. R. Tönjes, Control of scene reconstruction using explicit knowledge, in *Proceedings, 3rd IEEE Workshop on Applications of Computer Vision, 1996*, pp. 15–20.
39. U. Stilla and E. Michaelsen, Semantic modelling of man-made objects by production nets, in *Proceedings, Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*, pp. 43–52, Birkhäuser, Basel, Switzerland, 1997.
40. P. Fua and J. Hanson, Resegmentation using generic shape: locating general cultural objects, *Pattern Recognition Letters* **5**, 1987, 243–252.
41. F. Lang and W. Förstner, Surface reconstruction of man-made objects using polymorphic mid-level features and generic scene knowledge, *Z. Photogramm. Fernerkundung* **6/96**, 1996, 193–201.
42. T. Moons, D. Frère, J. Vandekerckhove, and L. Van Gool, Automatic modelling and 3D reconstruction of urban house roofs from high resolution aerial imagery, in *Proceedings, Fifth European Conference on Computer Vision, 1998*, Vol. I, pp. 410–425.
43. H. Wiman, Least squares matching for three dimensional building reconstruction, in *Proceedings, Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*, pp. 223–232, Birkhäuser, Basel, Switzerland, 1997.
44. C. Steger, C. Glock, W. Eckstein, H. Mayer, and B. Radig, Model-based road extraction from images, in *Proceedings, Automatic Extraction of Man-Made Objects from Aerial and Space Images*, pp. 275–284. Birkhäuser, Basel, Switzerland, 1995.
45. R. J. Pollock, A Model-based approach to automatically locating individual tree crowns in high-resolution images of forest canopies, in *Proceedings, First International Airborne Remote Sensing Conference and Exhibition, 1994*, pp. 357–369.
46. H. P. Pan, Production, inversion and learning of spatial structure: A general paradigm to generic model-based image understanding, in *Proceedings, International Archives of Photogrammetry and Remote Sensing, 1992*, Vol. (29) B3/III, pp. 930–937.
47. H. Bischof, W. Schneider, and A. Pinz, Multispectral classification of landsat-images using neural networks, *IEEE Trans. on Geoscience and Remote Sensing* **30**(3), 1992, 482–490.
48. J. Ponce, M. Hebert, and A. Zisserman, Report on the 1996 international workshop on object representation in computer vision, in *Proceedings, Object Representation in Computer Vision II*, pp. 1–8, Springer-Verlag, Berlin, 1996.
49. P. Garnesson, G. Giraudon, and P. Montesinos, An image analysis system, application for aerial imagery interpretation, in *Proceedings, 10th International Conference on Pattern Recognition, 1990*, pp. 210–212.
50. Y. Xu, A prototype object-oriented environment for image understanding, *GIS* **9**(2), 1996, 26–34.
51. D. M. McKeown, W. A. Harvey, and J. McDermott, Rule-based interpretation of aerial imagery, *IEEE Trans. Pattern Anal. Mach. Intell.* **7**, 1985, 570–585.
52. G. Bordes, G. Giraudon, and O. Jamet, Road modeling based on a cartographic database for aerial image interpretation, in *Proceedings, Semantic Modeling for the Acquisition of Topographic Information from Images and Maps*, pp. 123–139, Birkhäuser, Basel, Switzerland, 1997.
53. R. Ruskoné, L. Guigues, S. Airault, and O. Jamet, Vehicle detection on aerial images: A structural approach, in *Proceedings, 13th International Conference on Pattern Recognition, 1996*, Vol. III, pp. 900–903.
54. F. Quint, MOSES: A structural approach to aerial image understanding, in *Proceedings, Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*, pp. 323–332, Birkhäuser, Basel, Switzerland, 1997.
55. K. Kulschewski, Building recognition with Bayesian networks, in *Proceedings, Semantic Modeling for the Acquisition of Topographic Information from Images and Maps*, pp. 196–210, Birkhäuser, Basel, Switzerland, 1997.
56. J. Klonowski and K. R. Koch, Two level image interpretation based on Markov random fields, in *Proceedings, Semantic Modeling for the Acquisition of Topographic Information from Images and Maps*, pp. 37–55, Birkhäuser, Basel, Switzerland, 1997.
57. C. O. Jaynes, F. Stolle, and R. T. Collins, Task driven perceptual organization for extraction of rooftop polygons, in *Proceedings, 2nd IEEE Workshop on Applications of Computer Vision, 1994*, pp. 152–159.
58. Y.-T. Liow and L. Pavlidis, Use of shadows for extracting buildings in aerial images, *Computer Vision, Graphics, and Image Processing* **49**, 1990, 242–277.
59. E. Gülch, A knowledge based approach to reconstruct buildings in digital aerial imagery, in *Proceedings, International Archives of Photogrammetry and Remote Sensing, 1992*, Vol. (29) B2/II, pp. 410–417.
60. R. Nevatia, C. Lin, and A. Huertas, A system for building detection from aerial images, in *Proceedings, Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*, pp. 77–86, Birkhäuser, Basel, Switzerland, 1997.
61. M. Roux, Y. C. Hsieh, and D. M. McKeown, Performance analysis of object space matching for building extraction using several images, in *Proceedings, Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision II, 1995*, Vol. 2486. SPIE, pp. 277–297.
62. O. Grau, A scene analysis system for the generation of 3-D models, in *Proceedings, International Conference on Recent Advances in 3-D Digital Imaging and Modeling, 1997*, pp. 221–228.
63. H. Wiman and P. Axelsson, Finding 3D-structures in multiple aerial images using lines and regions, in *Proceedings, International Archives of Photogrammetry and Remote Sensing, 1996*, Vol. (31) B3/III, pp. 953–959.
64. V. Venkateswar and R. Chellappa, Hierarchical stereo and motion correspondence using feature grouping, *International Journal of Computer Vision* **15**(3), 1995, 245–269.
65. M. M. Trivedi, Analysis of high-resolution aerial images, in *Proceedings, Image Analysis Applications*, pp. 281–305, Marcel Dekker Publishers, New York, NY, 1990.
66. H. Mayer, I. Laptev, and A. Baumgartner, Multi-scale and snakes for automatic road extraction, in *Proceedings, Fifth European Conference on Computer Vision, 1998*, Vol. II, pp. 720–733.
67. A. Grün and H. Li, Linear feature extraction with 3-D LSB-snakes, in *Proceedings, Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*, pp. 287–298, Birkhäuser, Basel, Switzerland, 1997.
68. M. Kass, A. Witkin, and D. Terzopoulos, Snakes: active contour models, *International Journal of Computer Vision* **1**(4), 1987, 321–331.
69. L. Spreeuwers, K. Schutte, and Z. Houkes, A model driven approach to extract buildings from multiple-view aerial imagery, in *Proceedings,*

*Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*, pp. 109–118, Birkhäuser, Basel, Switzerland, 1997.

70. N. Paparoditis, M. Cord, M. Jordan, and J.-P. Cocquerez, Building detection and reconstruction from mid- and high-resolution aerial imagery, *Computer Vision and Image Understanding* **72**(2), 1998, 122–142.
71. M. Berthod, L. Gabet, G. Giraudon, and J. L. Lotti, High-resolution stereo for the detection of buildings, in *Proceedings, Automatic Extraction of Man-Made Objects from Aerial and Space Images*, pp. 135–144, Birkhäuser, Basel, Switzerland, 1995.
72. N. Haala and C. Brenner, Interpretation of urban surface models using 2D building information, *Computer Vision and Image Understanding* **72**(2), 1998, 204–214.
73. C. O. Jaynes, A. Hanson, and E. Riseman, Model-based surface recovery of buildings in optical and ranges images, in *Proceedings, Semantic Modeling for the Acquisition of Topographic Information from Images and Maps*, pp. 211–227, Birkhäuser, Basel, Switzerland, 1997.
74. O. Dissard, C. Baillard, H. Maître, and O. Jamet, Above ground objects in urban scenes from medium scale aerial imagery, in *Proceedings, Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*, pp. 183–192, Birkhäuser, Basel, Switzerland, 1997.
75. S. Mason and E. Baltsavias, Image-based reconstruction of informal settlements, in *Proceedings, Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*, pp. 97–108, Birkhäuser, Basel, Switzerland, 1997.
76. M. Roux and H. Maître, Three-dimensional description of dense urban areas using maps and aerial images, in *Proceedings, Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*, pp. 311–322, Birkhäuser, Basel, Switzerland, 1997.
77. R. E. Fayek, Preserving topography in 3D data compression for shape recognition, in *Proceedings, International Archives of Photogrammetry and Remote Sensing, 1996*, Vol. (31) B3/III, pp. 186–191.
78. T. Kim and J.-P. Muller, Building extraction and verification from spaceborn and aerial imagery using image understanding fusion techniques, in *Proceedings, Automatic Extraction of Man-Made Objects from Aerial and Space Images*, pp. 221–229. Birkhäuser, Basel, Switzerland, 1995.
79. C. Hug, Extracting artificial surface objects from airborne laser scanner data, in *Proceedings, Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*, pp. 203–212, Birkhäuser, Basel, Switzerland, 1997.
80. R. Michalski, J. Carbonell, and T. Mitchell, *Machine Learning—An Artificial Intelligence Approach*, Springer-Verlag, Berlin, 1984.
81. T. Bäck, U. Hammel, and H.-P. Schwefel, Evolutionary computation: comments on the history and current state, *IEEE Trans. on Evolutionary Computation* **1**(1), 1997, 3–17.
82. R. Englert, Systematic acquisition of generic 3D building model knowledge, in *Proceedings, Semantic Modeling for the Acquisition of Topographic Information from Images and Maps*, pp. 181–195, Birkhäuser, Basel, Switzerland, 1997.
83. A. Grün, O. Kübler, and P. Agouris, editors, *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Birkhäuser, Basel, Switzerland, 1995.