

# PARSING SEGMENTED DIGITAL IMAGES

John Stokes, Dr

Department of Photogrammetry  
Royal Institute of Technology  
S-100 44 Stockholm, Sweden  
E-mail: john@fmi.kth.se  
ISPRS Commission III

## ABSTRACT

Computer interpretation of images requires a decision on what to look for and a strategy exploiting this knowledge. A generic model for which objects make up the image is designed, assuming low altitude aerial images. The generic model is exploited first to choose a representation of the image suitable for parsing, secondly for defining a parsing procedure. As representation, a segmentation is chosen consisting of segment boundaries plus descriptions of segment interiors. Segment boundaries are only of the kind recognized by the parsers, in this case straight line and smooth curve segments represented using strip trees. Secondly, this segmentation is used as an input to a set of parsers which use a list of properties of buildings in order to interpret the input. Both line parsers and region parsers are used. Each parser is successful for a limited task. A set of parsers is scanned until the parse is accepted in a consistency test. This strategy is chosen as it is considered easier to test if a parse is acceptable than to design a well performing general parser.

Key words: Image interpretation, image segmentation, parsing, knowledge based interpretation.

## 1 INTRODUCTION

The production of large scale maps using aerial images is one of the more important applications of photogrammetry. Well established procedures for the measurement of points in images and the estimation of their location in object space have existed for a long time. Also, the possibility of using computers when carrying through the necessary photogrammetric procedures has introduced algorithms like e.g. bundle adjustment, which have brought the bulk of photogrammetric know-how to a high degree of completion. There is, however, a very important exception: Working with digital images, all photogrammetric work today rests on a continuous interaction between the computer and an operator, who is responsible for everything having to do with interpretation. The possibility of using computers for automating digital image interpretation has not yet given any algorithms used in a computer production line. (Förstner, 1989) calls the problem "the stepchild of photogrammetric research", indicating an extraordinary low interest among photogrammetrists for the problem. After the paper by (Fua and Hanson, 1988) and the advent of the decision procedures designed by (Wallace, 1978) and (Rissanen, 1984), the problem has, however, been approached by several research-workers, e.g. (McKeown, 1991), (Herman et al, 1984) and (Förstner 1988).

The success of an automated procedure for locating and describing objects in digital images rests on a rational procedure for the interpretation of image contents. Several subtasks can be identified which have to be performed in such a procedure: the image must be preprocessed in order to obtain an input suitable for parsing, the object types must be defined using models in such a way that these models can be explicitly used for parsing, the parsing should be carried through qualitatively without side views on statistical bias, errors and contradictions must be taken care of, etc. Several methods to parse man made objects have been developed, this kind of objects having so much internal structure that a generic model for them easily can be conceived. (Walz, 1972) devised a parser for line drawings, a method which, while intuitively appealing, strongly rests on the correct identification of boundary lines. It is therefore very sensitive to errors due to missing lines and is probably not an appropriate method for parsing segmentations of grey level images. (Dickinson et al, 1990) use the idea of aspects and perform the parse at a high description level, where errors are comparatively simple to trace. Also this method is sensitive to missing lines. The region segment parser suggested below is however directly inspired by this approach.

The work presented here comprises a design of a procedure for automated interpretation of digital low-altitude aerial images. Section 2 gives an overall presentation of a procedure for interpretation. Section 3 discusses the problem of obtaining image representations suitable for parsing. In section 4, the introduction of image features into 3-D object space using fully oriented 2-D images is discussed. Section 5 presents line and region parsers followed by an example in section 6 and a closing discussion in section 7.

## 2 OVERALL STRATEGY

When interpreting a grey level image, the representation of the image usually has no connection to the expected information contained in the image. A computer-based interpretation will in these cases start in quite an arbitrary way. If, instead, an image representation is based on the relation to object types expected in the image, interpretation, i.e. parsing the representation, can be made simpler. The image is then represented in terms of the geometrical contents, i.e. primitives, relevant for objects assumed to be present. A suitable representation of aerial images used for mapping purposes is a segmentation describing grey level discontinuities in the form of straight line and curve segments as well as segment interiors.

Having an image representation in terms of a limited set of primitives, the parsing amounts to a listing of those properties in the image that have a unique

interpretation using the generic model. In a practical situation, this is performed testing primitives in the image by scanning a set of rules involving properties and relations of object parts.

Several subtopics have been identified when automatizing the process of image interpretation:

- Specification of generic models which define the objects and thus limit what can be and is to be described.
- Segmentation of the image according to the generic models in order to obtain a well defined set of primitives for future parsing.
- Parsing the segmentation output in terms of object types given by the generic models and reconstruction of objects in 3-D space using this parse.
- Least squares adjustment of the obtained parse to the original grey level image or the segmentation when quantitative object description and localization is desired. Description of the global, regional and local quality of the image as well as the performance of image operators such as segmentation, when run on the images used.

The description of complex objects present in digital images is thus approached using a procedure with several steps. In the first step, a segmentation procedure is run on the raw image data producing a set of closed polygons with interior descriptions. This procedure presents the segment borders in a language that is intelligible to a generic model for the objects to be described. In the case investigated here, the objects are buildings with the borders between roofs and walls assumed to be straight line segments, this making the output set of closed polygons and other geometrically simple figures in the segmentation procedure a natural choice. In a second step, the segmentation output from the first step is interpreted as an element in object space using a parser. Quantitative information on the images can be obtained from the parse. If necessary, high quality information as well as quality aspects can be assessed in a third step. A more detailed elaboration of the second step is given in (Gülch, 1992) and of the third step in (Zielinski, 1992).

### 3 IMAGE REPRESENTATION USING SEGMENTATION

Due to the methods existing today for acquisition of digital images, these are almost invariably represented by giving grey levels in a regular grid. Here, a strategy is proposed where the image representation is goal dependent, i.e. the image is represented in terms of primitives which are suitable for describing the objects possibly contained in the image.

Limiting the scope to objects in aerial images, the objects can be described in terms of specified parts in order to be able to keep to a limited set of primitives. The description of these primitives as well as their possible internal relations are given by a generic model. The representation of the image, chosen to suit to the parser, is given in terms of the same primitives. In digital aerial images the expected objects

are buildings, forests, fields, lakes, etc. These are all describable in terms of internally homogeneous regions with boundaries made up of straight lines, smooth curves and a limited set of corners. A suitable image representation is then given in terms of such regions together with a stochastic model for description of texture. This representation is in principle complete, i.e. the grey level image can be reconstructed from it, disregarding white noise. This kind of representation will here be called a segmentation.

There is an abundance of different methods to segment digital images including edge closure, region growing and other methods. Many methods are of the ad hoc type leaving the user with no information about the quality of the result. No method is known to us where possible region boundaries and other properties are restricted by object related model requirements. In a rigorous approach, the segmentation should be performed under such restrictions. Segmentation procedures using the principle of minimum description length (Leclerc, 1989), (Dengler, 1991) have recently been developed and appear quite suitable for this task. Region boundaries could be described using strip trees, which also give the description length. Such a procedure is under development. Until results are available, an existing procedure using region growing is used. The boundaries of the resulting segmentation are described using strip trees as described in section 3.1.

Having passed the image through such a segmentation procedure, the data set to be parsed consists of a set of closed polygons made up of straight line or curve segments, a description of segment interiors in terms of trends and texture parameters and finally a window coarsely defining what part of the segmentation is to be considered.

#### 3.1 Boundary descriptions using strip trees

The boundaries sought for in aerial images are assumed to be smooth with a limited number of corners. Linear segments are used to approximate this case. Strip trees are used to describe these boundaries following a method developed by (Ballard, 1981).

Letting a strip tree represent the boundary to a neighbour, each region segment refers to  $K$  strip trees. The following notation is chosen to represent the region boundary, see also figure 1:

- $a_0$  number of strip trees in boundary ( $=K$ ).
- $a_k$   $k=1, \dots, K$ . Pointers to the root strips  $S_k$  of the  $K$  strip trees generating region boundary.
- $S_k = (x_b, x_e, w_l, w_r, p_l, p_r)_k$ . Strip descriptions.
- $x_b, x_e$  coordinates of start and end points for strip.
- $w_l, w_r$  left and right width of strip.
- $w = w_l + w_r$ .
- $p_l, p_r$  pointers to left and right substrip.

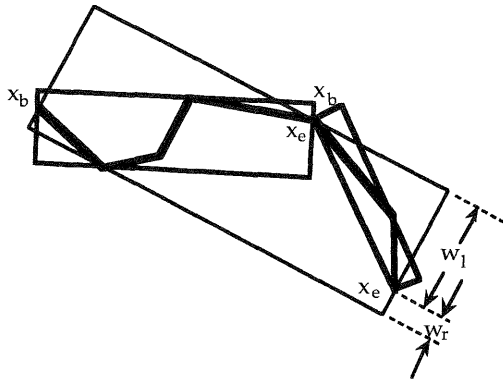


Figure 1. Two levels of strips in a strip tree.

Following Ballard, the strip is defined as a rectangle, two sides being parallel to the vector  $(x_e - x_b)$ , the other two sides having length  $w$  and containing the points  $x_b$  and  $x_e$ . Except at the bottom level, each strip is recursively divided into a left and a right substrip by choosing a new point as  $x_b$  or  $x_e$  on the strip border. From a principal point of view, new substrips can be defined until no more points occur between  $x_b$  and  $x_e$ , i.e. when  $w = w_1 = w_r = 0$ . In a discrete case, as with digital images, the process is halted when  $w \leq 2e$  where  $e$  is the pixel size.

In case of a strongly curved boundary, the description using a strip tree gives a tree with many levels and many leaves at the bottom level. If the boundary instead is a straight line, only one single strip is necessary for the description.

### 3.2 Structure of segment output

The line segments given by the segmentation procedure can have both ends connected to other line segments or have one or both end points free. All connections are at end points, i.e. a T-junction is described using three line segments with a 180 degree angle between two of them. All line segments are given in an ordered list.

Given the segmentation in the form of line segments, it can also be represented in the form of region segments. These are also given in an ordered list. The region is assumed to lie on the left hand when moving from the first to the second point of a boundary. An interest index is introduced giving the number of leaves in the strip tree describing the total segment boundary divided by the boundary length. This index will be small for segments mainly bordered by straight line segments. The group index is the sum of interest indices of the immediate neighbours plus the index of the region segment itself. The group index will be low when several neighbouring segments have low interest indices, a situation which usually will occur for man made objects such as buildings. Accordingly, possible buildings can be identified looking for minima of the group index.

## 4 INTRODUCING IMAGE FEATURES INTO OBJECT SPACE

The parsers described here assume fully oriented images. An advantage with exploiting the orientation information for an image is that object space is

described using a global frame of reference. When several different images are used to reconstruct the object in object space, a detail in the object which is present in several images will be localized approximately in the same place. Solving the matching problem then amounts to introducing a suitable neighbourhood relation.

### 4.1 Relation between image space and object space

When each image has been fully oriented, inner as well as outer orientation has been performed. The camera constant  $c$ , the camera location  $(X_0, Y_0, Z_0)$  and orientation  $(\varphi, \omega, \kappa)$  in object space are known.

Points  $(x, y)$  in the images can, when these parameters are known, be localized in object space as soon as their altitude  $Z$  is known:

$$\begin{aligned} X &= X_0 + (Z - Z_0) \cdot \frac{x r_{11} + y r_{12} - c r_{13}}{x r_{31} + y r_{32} - c r_{33}} \\ Y &= Y_0 + (Z - Z_0) \cdot \frac{x r_{21} + y r_{22} - c r_{23}}{x r_{31} + y r_{32} - c r_{33}} \end{aligned} \quad (1)$$

Here  $r_{ij}$  are the coefficients of the rotation matrix defined by  $(\varphi, \omega, \kappa)$ .

When line segments in the segmentation are said to be parallel, they are assumed to have a common point of convergency. When the image is the result of a projective transformation (which is the common situation), the convergency points can lie anywhere in the plane (i.e. they can be finite). The images assumed to be fully oriented, all convergency points are known.

In the procedures to be described below, a first point of the building will be chosen in the image and assigned an arbitrary height. Using various criteria or judgements for line segments being horizontal or vertical, the building will successively be reconstructed. Height differences between end points of vertical line segments are then estimated using the relation

$$h = [(H/c)/(A/c)]a \quad (2)$$

where  $H$  is the flying height in object space dimensions (usually meters),  $A$  is the distance of the closest point on the line segment to the nadir point and  $a$  is the length of the line segment, both lengths given in the same units, e.g. camera system units or pixels.

### 4.2 Drawing line segments in object space

When using the line segment parser, line segments are introduced into object space according to the following procedure: After deciding for a point lying on the ground, this point is introduced into object space at an arbitrarily chosen altitude. A line segment having a start point connected to a point already present in object space is introduced using this point as a departure. Vertical and horizontal line segments have end points with coordinates that can be determined using relations (1) and (2) above. If this end point is located within a threshold distance to the same point introduced earlier, this point is readjusted

according to the number of line segments used to determine it, else it is introduced as a new point. No further action due to the discontinuity is taken at this point. In the consistency check following the parsing it will be investigated if the discontinuity can be removed. Some rules of the parsers require introduction of line segments neither vertical, nor horizontal. In these cases arbitrary heights are introduced for later correction.

When the parsing is initiated using the region segment parser, the region segments are matched to an aspect prototype and introduced into object space using the known labels of the line segments in the prototype, placing a ground line segment at an arbitrary datum and otherwise following the same procedure as for the line segment parser. The meaning with the region segment parser is therefore mainly a preassigning of certain labels to specific line segments in the input.

## 5 PARSING THE SEGMENTATION OUTPUT

Details of the segmentation output are to be interpreted as details in object space using a generic model for possible objects and a parser. This model describes the building in terms of roofs, walls, doors, windows and garden regions, all described by geometrically simple regions and relations between these.

There is no reason to insist in a parser that succeeds with an appropriate analysis of any possibly incomplete segmentation of any possible building. If at all possible to construct, such a parser will be very complicated and probably quite slow. What is needed is instead a procedure which can tell when simple and fast parsers have failed. If the description of the object is accepted, a successful parse has been performed in a short time, if not, another parser is chosen. Using a set of parsers of moderate success rates will generate a compound procedure with a high success rate. The procedure described below is an example of such a set of parsers, designed to take care of the fact that images of horizontal line segments are not easily identified.

The parse can be performed either using line segments or region segments. There are two advantages using line segments: First, the probability of a region segment being erroneous is high as one missing or misinterpreted line segment makes the whole region segment erroneous. Second, the fact that all vertical line segments point towards the nadir point gives a simple criterion for the interpretation of these line segments. There are, however, also disadvantages: Vertical line segments of objects close to the nadir point are short or missing in the segmentation output. Two different kinds of parsers are therefore used in the procedure below, which starts with the set of line segment parsers and continues with the region parser if the first procedure breaks down.

### 5.1 Object recognition

The input to the parsers consists of the segmentation output together with a window identifying what part of the segmentation should be interpreted and described as a building. The window is provided by

the operator or given by a recognition procedure as for instance the following one:

A simple scan of the segmentation represented as region segments gives locations where the group index (see section 3.2) is minimum. These locations are used as probable locations of the objects looked for. In the case of buildings, the interest index was chosen so that building roofs should tend to give minimum group indices. The reference point of the segment, together with a window size determined from neighbouring object indices, determines a window to adjoin the segmentation presented as input to the parser.

### 5.2 A generic model for buildings

Complex objects like buildings are difficult to model using specific models. Although it is possible in principle to cover the possible logical structures using parameters, this is impractical. An alternative approach using generic models has been used by several investigators, e.g. by (Fua and Hanson, 1988). Here, a generic model is used, defining buildings as objects present in object space and consisting of sub-objects with certain properties and relations. A sketch of this model is given below, the details of which are given implicitly in the parsers in sections 5.3 and 5.4.

buildings:	connected set of roofs, walls, floors, windows, doors, chimneys
	properties:
ground:	set of nonbuildings
	properties:
roofs:	geometry: polygon(closed, plane, nonvertical)
	neighbours: roofs, walls
	properties: a neighbour at a sloping border is a roof
walls:	geometry: polygon(closed, plane, vertical)
	neighbours: roofs, walls, ground
	properties: a neighbour at a horizontal border is not a wall
floors:	geometry: polygon(closed, plane, horizontal)
	neighbours: walls
	properties: at least one neighbour is beneath the floor
windows:	geometry: polygon(closed, plane, vertical)
	neighbours: walls
	properties: one surrounding neighbour
doors:	geometry: polygon(closed, plane, vertical)
	neighbours: walls, roofs
	properties: one semi-surrounding neighbour
chimneys:	vertical structure, part of roof
polygons:	connected set of line segments
line segments:	2 points
points:	(x,y,z)

Table 1. Sketch displaying principle of generic model for buildings

### 5.3 The parser for line segments

The most conspicuous properties of images of buildings are that they can be described using line drawings, where some of the line segments represent vertical or horizontal lines. A natural approach when parsing a segmentation including buildings is then to identify vertical line segments as those pointing at the

nadir. Also, when the nadir point is close to the image centre (a usual situation in aerial photography), those horizontal line segments which generate buildings are either parallel or orthogonal within the building. These ideas are used in the line segment parser described below.

The general procedure for reconstruction in object space is to start from a vertical line connected to several horizontal lines and introduce these into object space at an arbitrary datum. Vertical and horizontal lines connected to already introduced lines are then successively introduced. Logically weaker rules are introduced when no more lines can be introduced. As soon as a line has been found and introduced into object space, the procedure returns to the top level looking for more horizontal and vertical lines. A descending line of successively weaker rules used in this way are activated only when no stronger rule is applicable.

The search for new lines is always made in order of numbering. As soon as a line has been introduced into object space, it can be used as a connection for new lines. The various steps are given below, most steps being illustrated in section 6 and figure 3.

- 1) **Generation of line drawing.** Strip the segmentation of all line segments that cut the window boundary. The remaining structure is a bounded segmentation.
- 2) **Grouping.** Sort the line segments of the segmentation into groups  $G_i$ ,  $i=1,..I$ , each group containing parallel line segments of a direction  $\phi_i$ ,  $i=1,..I-1$ ,  $\phi_i \neq \phi_j$ ,  $i \neq j$ . The  $I$ :th group contains all line segments not parallel to any others.
- 3) **Vertical lines.** Find the group of line segments pointing at the nadir point and call it group 1. This group can be empty.
- 4) **Search strategy.** Sort the groups 2 to  $I-1$  according to a predefined criterion. This criterion defines the search strategy, i.e. it gives the order in which groups are searched in point 5). Several different search strategies will be used, the procedure 4) to 15) running through each one in turn until the segmentation is accepted. Examples of search strategies are for instance sorting the groups according to decreasing size, first considering line segments that are orthogonal or parallel or sorting line segments according to length.
- 5) **Initial vertical line for parsing.** Beginning with group 2 and proceeding to group  $I-1$ , look for two line segments belonging to the same group, both of which are connected to a line segment in group 1. Avoid T-junctions. If no such pair is found, try the region segment parser. Introduce the three line segments found into object space.
  - a) Assume that the altitude  $Z$  at the point closest to nadir of the line segment from group 1 is zero. Given the image coordinates  $(x,y)$ , this point is given the point  $(X,Y,0)$  of object space using the transformation (1). The other point of this line segment is placed in  $(X,Y,h)$  of object space, where  $h$  is given by (2).

- b) The other two line segments, both assumed to be horizontal, are connected to the end points of the line segment introduced in a). Their unconnected end points are ascribed altitudes  $Z$  according to the already known altitude of the connected end points. Both line segments are introduced into object space using the transformation (1).

6) **Connected horizontal and vertical line segments.** Check all groups 1 to  $I$  for line segments connected to those already introduced in object space.

- a) If a line segment so found does not belong to group 1, is parallel or orthogonal in the line drawing to line segments already classified as horizontal, it is itself classified as horizontal and introduced into object space using the altitude  $Z$  of the connection point in the transformation (1).
- b) If a line segment so found does belong to group 1, it is assumed to be vertical and introduced into object space using (2) to determine the altitude of the free end point.
- c) Line segments so found, but not qualifying under a) or b), are neglected.

7) **Do 6)** until no more line segments qualifying under a) or b) are found.

8) **Sloping roof line segments.** Look for two already classified line segments having the same height in object space and belonging to the same group. The two end points  $P_1, P_2$  should connect to two line segments  $L_1$  and  $L_2$ , where the continued lines have a common point  $P_3$ . At least one of these line segments should be unclassified. The common point  $P_3$  should be on the continuation of a line segment  $L_3$  being either parallel or orthogonal to the horizontal line segments.

The line segments  $L_1$  and  $L_2$  are interpreted as sloping roof line segments.  $L_3$  is interpreted as a roof line.  $L_3$  divides the roof into two roof parts to which  $L_1$  and  $L_2$  belong. If there exist line segments  $L_1'$  parallel to and belonging to the same segment as  $L_1$ , and  $L_2'$  parallel to and belonging to the same segment as  $L_2$ , both of which have a common point on the continuation of  $L_3$  at the other end point, the gable part defined by  $P_1, P_2, P_3$  is assumed to be vertical. In this case, do a) else assume the roof to be symmetrical and do b).

- a) A line through  $P_3$  and nadir cuts the line between  $P_1$  and  $P_2$  at  $P_4$ . The line through  $P_3, P_4$  is considered vertical. The altitude of  $P_3$  and so of  $L_3$  is determined using the relation (2).
- b) Choose a point  $P_4$  in the image, centred between  $P_1$  and  $P_2$ . A line through  $P_4$  and nadir cuts a line colinear to  $L_3$  at  $P_5$ . The line segment  $P_4, P_5$  is considered vertical. The altitude of  $P_5$  and so of  $L_3$  is determined using the relation (2).

$L_3$  can now be introduced into object space using (1).  $L_1$  and  $L_2$  are introduced between the points  $P_1, P_2$  and the end point of  $L_3$ .

In the case a),  $L_1'$  and  $L_2'$  are introduced with their end points determined using their lengths and their parallelity to  $L_1$  and  $L_2$ .

In the case b), if there exist line segments  $L_2'$  parallel to and belonging to the opposite segment as  $L_1$ , and  $L_1'$  parallel to and belonging to the opposite segment as  $L_2$  (i.e. crosswise parallelity), both of which have a common point on the continuation of  $L_3$  at the other end point, the line segments  $L_2'$  and  $L_1'$  are introduced into object space using their lengths and parallelity to  $L_1$  and  $L_2$ .

9) Do 6)-8) until no more line segments are introduced.

10) **Colinear lines.** Look for a line segment  $L_1$  which is colinear to an already classified horizontal line segment  $L_2$ . Assuming that  $L_1$  and  $L_2$  are colinear also in object space, the altitude of the end points of  $L_1$  is known. If they adhere to at least one of the following properties,  $L_1$  is introduced into object space. The properties to be tested are

a) Another set of colinear line segments  $L_1'$  and  $L_2'$  can be found, where  $L_2'$  is already classified, and where  $L_1'$  and  $L_2'$  are orthogonal to  $L_1$  and  $L_2$ . If  $L_1$  connects to  $L_1'$  while  $L_2$  and  $L_2'$  have the same altitude,  $L_1$  and  $L_1'$  are both introduced into object space.

b) Another set of colinear line segments  $L_1'$  and  $L_2'$ , where at least  $L_2'$  is already classified (not necessary the same altitude as  $L_2$ ), and parallel to  $L_1$  and  $L_2$  can be found.  $L_1$  and  $L_1'$  should at least partly project on each other. If  $L_1$  is unclassified, it can be introduced into object space using the same altitude as  $L_2$ .

c) A line segment  $L_3$  can be found having the same altitude as  $L_2$ . Also, a line cutting  $L_1$  and  $L_3$  through their end points cuts  $L_1$  at right angles and has  $L_1$  and  $L_3$  on the same side.

11) Do 6)-10) until no more line segments are introduced.

12) **Miscellaneous.** Check the remaining line segments.

a) **Windows and doors.** If they among themselves generate closed polygons completely within anyone of the building elements already found, and if some of the line segments of such a closed polygon belong to group 1, while some belong to groups containing horizontal line segments, they are classified as windows or doors. If they can not be classified this way, they are rejected.

b) **Doubly connected line segments.** Introduce into object space all line segments connected at both ends to line segments already introduced.

13) Do 12) until no more line segments are introduced.

14) Do 6)-13) until no more line segments are introduced.

15) **Consistency check.** Check the building reconstructed in object space for consistency. If it is consistent (complete or incomplete), the parsing is terminated, if it is inconsistent, restart from 4).

#### 5.4 The parser for region segments

When buildings lie on or close to the nadir point of an image, their walls are generally invisible. In this case it is natural to build the parsing procedure on sets of region segments, aspects, describing composite parts of the roof. The possible combinations of a low number of geometrical objects describing a part of a roof are not too many, while they at the same time define specific properties of the building to be described. It is in no way necessary to describe the complete roof in an aspect, rather only so much that the identified part is sure to be a roof part. It is, of course, possible also to include wall parts in these aspects.

The segments contained in the input window of the segmentation are tested for matching against the set of aspect prototypes shown in figure 2. Note that only the structure of the aspects is important, i.e. lines that are parallel or orthogonal in the prototypes should be parallel or orthogonal in the segments to be tested. Lengths of line segments are not important, nor are symmetry properties in the prototypes. In order to use as much information as possible in an aspect prototype, the test for matching is performed starting with the most complex prototypes (those containing most line segments) and continued with matching on simpler ones only as long as no match has been found.

When a matching prototype has been encountered, the corresponding line segments in the segmentation are introduced into object space. The interpretation of all line segments in the prototype is then assumed to be known (this is chosen in an obvious way once and for all). Line segments placed on the ground are assumed to be horizontal, given the altitude zero and introduced using relation (1). Vertical line segments are tested for passing through the nadir point and introduced into object space using relation (2). Horizontal line segments connected to points already

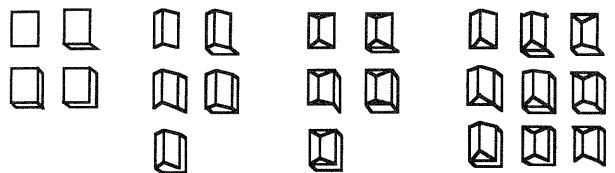


Figure 2. Prototypes for building parts to be identified in segmentations.

introduced into object space are introduced at the appropriate altitude using relation (1). Finally, obvious missing lines are introduced into object space. If the prototype contains no vertical line at all, the horizontal lines are interpreted as roof lines and introduced into object space at a preassigned altitude.

When the parsing has been initialized according to these rules, the rest of the parse is continued using the line segment parser and starting at rule 6).

### 5.5 Alternation between parsers

In the procedure described above, the parsing begins with trying a line segment parser. If the parse produced is unsuccessful, other versions of the same parser are used trying the various strategies in rule 4. If no parse passed the consistency test, the region segment parser is used. If also this attempt is a failure, the user must be asked for help.

As images of buildings never display the complete structure, the consistency test will be performed in such a way that also partially parsed buildings are accepted. A typical result of parsing the segmentation of a single image is therefore a partial building structure in object space plus a set of line segments in the segmentation, where the parsing has failed. These remaining line segments might belong to the building or they might describe the ground surrounding it.

In order to take care of the line segments which belong to the building, the complete parsing procedure is started again, now skipping the line segments already included in object space. A new vertical line segment connected to horizontal line segments is selected according to rule 5), the sequence of line segment parsers and region segment parser being carried through until no more line segments are parsed. Again, this complete procedure is repeated until no more line segments are parsed. In this way weakly connected parts of the same building, and also different buildings contained in the window to be parsed, are introduced into object space at approximately correct locations.

Remaining line segments or polygons, which can be connected to a ground point of the reconstructed building are assumed to be ground line segments in a horizontal environment. They can be introduced using the transformation (1). If there still are line segments left, they are unconnected to anything parsed.

### 5.6 Consistency check for buildings in object space

The general procedure given above for interpretation of the input segmentation is to use a series of moderately successful parsers and terminate the procedure when a consistency check indicates an acceptable object description. This approach rests on the assumption that it probably is easier to design a high quality consistency check than a well performing parser. The consistency check is designed so as to check all rules contained in a generic model. As a consequence, this procedure is an interpretation process. Substructures, in the form of 2-D regions in object space, which have passed the consistency test, are ascribed labels (wall, roof, etc).

As the parse to be checked usually will contain inconsistencies which can be removed, the consistency check should include procedures for such removals. Examples are discontinuities due to erroneous parses of individual line segments and also missing lines which sometimes must be introduced in order that the the 2-D regions in object space be plane.

### 5.7 Introduction of missing lines

Check that the mappings in object space of all parsed region segments are planar. Denoting the points  $r_k = (x_k, y_k, z_k)$ ,  $k=1, \dots, n$  in the associated closed polygon, check that the following determinants are zero:

$$\begin{vmatrix} x_k & y_k & z_k & 1 \\ x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \end{vmatrix} = 0; \quad k = 4, \dots, n$$

Segments not satisfying this requirement are assumed to be composed of several segments divided by line segments missing in the segmentation. Subsegments are produced and tested for planarity by introducing line segments  $L_{kj} = (r_k, r_j)$ ,  $k=1, j=3, \dots, n-1; k=2, \dots, n-2, j=k+2, \dots, n$ . When a planar subsegment has been found the line segment is introduced into object space. The remaining subsegment is again tested for planarity, continuing the procedure until no segment is left.

### 5.8 Interpretation

After checking that all polygons are closed and planar, they are labelled walls, roofs, floors, windows, etc. according to the properties of these subobjects as specified by the generic model. This amounts to checking surface orientation and neighbours.

- Starting with identifying all vertical region segments, these are checked for generating horizontally connected structures. If unconnected vertical boundaries exist, this is due to effects of perspective giving rise to hidden regions. All region segments participating in such structures are labelled walls.
- Sloping regions above walls are labelled roofs.
- Horizontal regions above walls are labelled floors, below walls are labelled ground.
- Unclassified vertical regions located in the interior of walls or roofs are interpreted as doors or windows, depending on them sharing the bottom line segment with the parent region or not.

### 5.9 Miscellaneous tests

Besides the rules given in the generic model, certain criteria must be fulfilled due to the projection under which the image was made. Also other tests can be performed in order to ensure a correct parse. Examples of such rules are that all line segments from the given segmentation are visible, i.e. as seen from the sensor; they cannot lie behind surfaces generated by other line segments. Also no line segments can cut through building surfaces of any kind.

### 5.10 The decision of acceptance

After running the consistency check on a given parse, the final decision of accepting the parse or not has to be made. Local failures can be accepted as well as local errors, which are just removed. The kind of parsing errors which should lead to rejection are global: e.g. when the line segment parser chooses to start with a line pointing at nadir and this line in fact is non-vertical. Criteria for this and other fatal mistakes are yet to come.



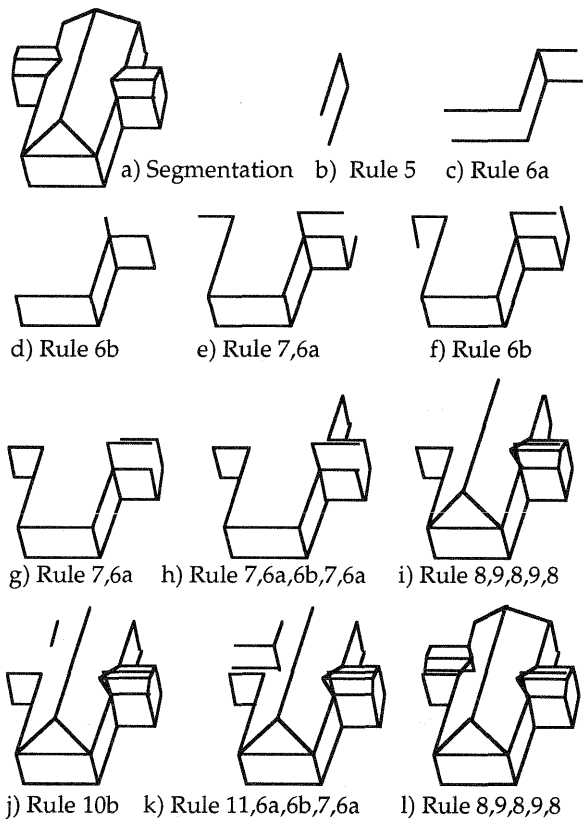


Figure 3. Line parser run on synthetic segmentation.

## 6 EXAMPLE

The line segment parser was tested manually on a synthetic segmentation of a building shown in figure 3a. The different rules used for the reconstruction are shown step by step in figures 3b-l. As is easily seen, rule 6b erroneously introduces vertical lines twice (figures 3d and 3k), this due to sloping roof lines directed towards the nadir point. These errors are corrected in the consistency check, requiring that segments be closed and plane. Further discussion of examples are given in (Gülch, 1992).

## 7 DISCUSSION

The procedure described here splits the problem of interpretation into two steps: segmentation and parsing. For segmentation, long being a bottle-neck of image analysis, procedures now exist giving results of acceptable quality. However, the requirement in this paper that segmentation boundaries should be given by the same models as those used by the parsers, is crucial: without this requirement parsing instead of segmentation becomes an approximation process. Parsing, being a task only involving logical decisions, has nothing to do with approximation. Segmentation procedures with this property are now being developed. Our implementation has not yet proceeded far enough to give inputs for the parsers described.

When this goal is achieved, more realistic tests will be performed with the parsers presented here, which have given very promising interpretations when using synthetic inputs.

## 8 ACKNOWLEDGEMENTS

The support of the Swedish National Board for Technical Development under project nr 89-1537 and the Swedish Council for Building Research under project nr 890743-2 is gratefully acknowledged.

## 9 REFERENCES

- Ballard, D. H., 1981. Strip Trees: A hierarchical representation for curves, *Comm. ACM* 24, 310-321.
- Dengler, J., 1991. Estimation of discontinuous displacement vector fields with the minimum description length criterion. *Proc. CVPR'91*, Lahaina, Maui, Hawaii, IEEE Press.
- Dickinson S. J., Pentland A. P., Rosenfeld A., 1990. Qualitative 3-D Shape Recovery using Distributed Aspect Matching. CAR-TR-505, CS-TR-2484, AFOSR-86-0092, University of Maryland, College Park, Maryland.
- Fua P., Hanson A. J., 1988. Generic feature extraction using probability-based objective functions. SRI International.
- Förstner, W., 1988. Model based detection and location of houses as topographical control points in digital images. In: *Int. Arch. Photogramm. Remote Sensing.*, Kyoto, Vol 27, B11, pp. 505-17.
- Förstner, W., 1989. Image analysis techniques for digital photogrammetry. *Proc. 42:nd Photogrammetric Week*, Stuttgart, FRG.
- Georgeff, M. P., Wallace, C. S., 1984. A general selection criterion for inductive inference. In *ECAI 84: Advances in Artificial Intelligence*. Elsevier, pp. 473-482.
- Gülch, E., 1992. A knowledge based approach to reconstruct buildings in digital aerial imagery. In: *Proceedings XVII ISPRS Congress*, Washington D.C. USA.
- Herman, M., Kanade, T., Kuroe, S., 1984. Incremental acquisition of a three-dimensional scene model from images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. PAMI-6, No. 3.
- Leclerc, Y. G., 1989. Constructing Simple Stable Descriptions for Image Partitioning. *IJCV*, 3, pp.73-102.
- McKeown, D. M., 1991. Information fusion in cartographic feature extraction from aerial imagery. In *Ebener, H., (ed.) Dig. Photogramm. Systems*. Karlsruhe: Wichmann.
- Rissanen, Y., 1989. *Stochastic Complexity in statistical inquiry*. World Scientific, Teaneck, NJ.
- Walz D. L., 1972. *Generating Semantic Descriptions from Drawings of Scenes with Shadows*. MIT, Internal report. Reprinted as "Understanding Line Drawings of Scenes with Shadows" in *Winston P.H. (Ed.) (1975): The Psychology of Computer Vision*, MacGraw-Hill, pp 19-92.
- Zielinski, H., 1992. Line photogrammetry with multiple images. In: *Proceedings XVII ISPRS Congress*, Washington D.C., USA.