Using Generic Geometric Models for Intelligent Shape Extraction

Pascal Fua and Andrew J. Hanson *

Artificial Intelligence Center
Computer and Information Sciences Division
SRI International

Abstract

Object delineation that is based only on low-level segmentation or edge-finding algorithms is difficult because typical edge maps have either too few object edges or too many irrelevant edges, while object-containing regions are generally oversegmented or undersegmented. We correct these shortcomings by using model-based geometric constraints to produce delineations belonging to generic shape classes. Our work thus supplies an essential link between low-level and high-level image-understanding techniques. We show representative results achieved when our models for buildings, roads, and trees are applied to aerial images.

I. Introduction

Our goal is to find and delineate probable instances of generic object classes in real images. The shape delineation task described here is critical for the extraction of objects from images that are too complex to be handled by syntactic approaches alone.

We choose as our application domain aerial images of intermediate resolution, that is, images with resolution adequate for humans to perceive shapes clearly, but not so fine that small details and textures would dominate the description given by a human observer. In Figure 1, we present a typical aerial image of this class that contains a combination of suburban features, along with a corresponding edge image [Canny, 1986] and a segmentation [Laws, 1984].

A standard low-level approach to the task of extracting objects such as buildings from Figure 1a would attempt to match region boundaries or edge groups with the edges of a building template. However, when we examine the data, we see that neither regions nor edges correspond reliably to building objects. The segmentation boundaries tend either to break a building roof into pieces, or to merge extraneous areas with those identifiable as roofs. The Canny edges, on the other hand, do not include several critical edges in the center building or the road, even though these are extracted as region boundaries by the segmentation.

*This research was supported in part by the Defense Advanced Research Projects Agency under Contract No. MDA903-86-C-0084.

Clearly, no single parameter setting for conventional segmentation or edge-finding techniques can be expected to handle all the desired objects in one image, much less in multiple images.

Therefore, an intermediate step is required for complex scenes: model-based shape parsing procedures must be provided in order to generate object delineations that are sufficiently reliable to be useful for applications such as context-specific labeling systems [see, e.g., Brooks, 1981; McKeown et al., 1985].

The key elements of our approach to solving this problem are the following:

- **Define Generic Shape Models.** We avoid the drawbacks of rigid template models and produce delineations that are not necessarily tied to any specific labeling scheme by defining shape models for generic classes of objects. When we supplement low-level data with the predictive power of such models, we are able to recover information that is more likely to be semantically meaningful.
- **Integrate Edge-Based and Area-Based Geometric Constraints.** Both the edges and areas of a feature contain geometric information relevant to the task of identifying it as an instance of a generic model. We use edges to generate overall geometry and to provide estimated area outlines. Areas that are associated with edges are tested for compatibility with the object model and with one another; we use the RANSAC random sample consensus technique [Fischler and Bolles, 1981] to compute optimal model fits that systematically discount gross anomalies. Furthermore, *multiple sources* of information are incorporated in the geometric search procedure by using a set of segmentations produced by a progression of parameter settings.
- **Predict and Verify Model Components** Missing components of models are predicted and checked using model-based adaptive search procedures; our implementation uses a gradient ascent method [Leclerc and Fua 1987] to search for predicted edges with the required geometry. Thus, for example, we can reconstruct and locate building boundaries and road edges that might be unrecoverable using conventional methods; if one chose an edge-detector parameter setting

weak enough to find such missing edges at the beginning, the edge map would be dominated by irrelevant noise.

II. Generic Modeling

People can classify instances of various object categories accurately even though a particular instance may have a unique shape they have never seen before. Generic shape classes provide an effective approach to automating this human ability. Generic models that we have found useful for analysis of *real images* possess the following characteristics:

- **Strong edge geometry.** The elementary edge or line data extractable from an image must be related in some direct and computable way to the object. In particular, the model must suggest explicit rules for dealing with anomalies and predicting likely locations of missing geometrical components. Typical models include edge geometry characterized by long, straight edge segments, by edges or lines with uniform local curvature, and by edges with good statistical signatures characterizing their jaggedness. In addition, there must be mechanisms for the production of coherent area-enclosing structures. Thus, for example, parallel edges, corners, equidistant curved lines and edges outlining a compact shape are reasonable geometric substructures that can be used to delineate areas that are portions of the larger structure.
- **Strong area signature.** Areas contained within a substructure of a generic object should be characterizable by a computable signature. If anomalies are expected, they should be clearly distinguishable by using the area signature and should ideally comprise no more than a small fraction of the area. Examples of such areas are parking lots with cars or roofs with chimneys. Typical area signatures would be the presence of uniform or uniformly changing intensity values or textures. Anomalies in such a background are easily located and discounted using a RANSAC procedure to fit planes to the intensity values within the delineated area.

The models we have implemented – buildings, roads, and trees – contain the following universal components: (1) edge definition, (2) composite structure definition, (3) linking geometry specification for composite structures, (4) area signature specification, and (5) a geometric completion model. The components of each of these models are are summarized in Table I.

The most general model-parsing procedure that we have needed to interpret each of these models in an image includes the following steps:

1. **Build the edges according to the edge definition.**
2. **Find edge relationships based on geometric constraints combined with tests on signatures of enclosed areas.**
3. **Build closed graphs of related edges that enclose consistent areas as well as matching the model geometry.**
4. **Predict, search for, and fill in missing elements of the model geometry.**
5. **Compare the resulting delineation with the characteristics of the original model.**

The overall approach can clearly be extended to any other object for which appropriate characteristics can be formulated, e.g, cylindrical oil tanks, drainage patterns, and buildings with perspective distortion.

In the following subsections, we outline the features of our models for buildings, roads, and trees, and illustrate how these models fit into the general framework. Where space allows, we mention some details of the individual requirements of the model parsing framework outlined above.

A. Buildings–Rectilinear Networks

Our most extensive work so far has been devoted to the task of delineating rectilinear, presumably cultural, structures [Fua and Hanson, 1985, 1986].

We characterize buildings and related cultural structures (e.g., parking lots, patios, gardens, and courtyards) as rectilinear networks of adjacent or joinable area-enclosing straight-edge groups that contain areas with planar intensity.

The basic parsing procedure for generic rectilinear structures follows the pattern given above. Since region boundaries of a histogram-based segmentation [Laws, 1984; Ohlander et al. 1978] tend to correspond to high image gradients, the straight edges are extracted as sequences of pixels with consistent gradient directions. While single segmentations often have inadequate characteristics, segmentations with increasingly permissive parameters produce regions that are first undersegmented, then well segmented and finally oversegmented as shown in Figure 2. The multiple data sources lead to a network of geometrically consistent straight edges, shown in Figure 2f, that serve as a basis for the geometric processes.

In practice, region boundaries may deviate by a few pixels from the actual edge location; we optimize their locations using the gradient ascent procedure. In each of the segmentation regions, edges that are parallel or perpendicular are singled out for special consideration. These associated edges, together with the region they come from, define areas in the image. Areas are tested for consistency with a RANSAC planar fit in intensity space, and edges that generate qualifying areas are retained for further parsing.

We note that the same edge can belong to several structures. If the structures are compatible with respect to the structure-linking specification and enclosed-area characteristics, new geometric relationships between edges,

such as collinearity, are instantiated. The result is that edges are grouped into networks defined by graphs of the geometric relations among them.

Rectilinear geometric relationships are used to predict how the edges should be linked and where missing edges might be. The predictions are fed to the adaptive straight-edge finder [Leclerc and Fua, 1987] or to the F^* edge finder [Fischler, et al., 1981] if a straight edge link is not found.

The networks of compatible composite structures are then connected to form closed contours and define new semantically motivated regions that are the final output of the current system. The candidate features can be scored using a measure of the closeness of the delineation characteristics to those expected in an ideal model instance.

In practice, the information required to assign meaningful labels to candidate cultural structures can be very primitive; we shall give some examples below in which even such simple techniques as clustering based on region-similarity measures are quite effective.

B. Roads–Curvilinear Parallel Networks

It is straightforward to modify the rectilinear cultural feature model to include smoothly curving road segments. The edges of such roads are almost straight in most places and can be detected locally using the techniques given above. They are then grouped into parallel structures and linked into elongated networks that may have large-scale curvature. To deal with winding roads, the straight edges can be replaced by smoothly curved edges while the rest of the approach is retained. See Table I for a summary.

The only major change in the road model is the rule employed to predict missing components of the geometric structure. First, the initial network of parallel edges is used to estimate the location of the road's center and width. Next, we fit a spline to the estimated center of the road and use it to define two parallel splines that correspond to the road's edges. Using the gradient ascent method, we optimize the location of the two splines under the constraint that they must remain parallel. This is a powerful technique because, wherever one side of the road is lost due to poor photometry or occlusions, the edge information present on the other side can still be utilized to guide the optimization procedure.

C. Trees–Irregular Clumps

Clumps of vegetation, typically small groups of trees, are characterizable as being complementary to the regular cultural-object models we have described so far. Since their edges are typically jagged and irregular, any compact object that has no components resembling roads or buildings could be a candidate for vegetation. Other irregular objects such as rock outcroppings, bodies of water, and drainage patterns would have similar signatures. The tree model is summarized in Table I.

The parsing procedure for vegetation clumps starts with the boundary of a given segmentation region and optimizes the location using gradient ascent. In contrast to the building model, jagged parts of the boundary, rather than linear parts, are selected as edge candidates. Within single regions, jagged edges that have consistent neighboring areas computed by local chamfering are incorporated into composite structures. The F^* general linear-feature utility is then used to connect the edges along the path with the strongest image gradient. This generates the required closed regions delineating the vegetation candidate.

III. Typical Results

In this section we present some representative results of applying our approach to aerial imagery.

To illustrate the behavior of the system on buildings, we have chosen two images that are especially challenging in terms of shape complexity and faint edge photometry. In Figure 3, we show the results of analyzing the image shown in Figure 1a. Figure 3a illustrates the initial set of networks, selected in this case on the basis of a size filter; if we add a selection criterion based upon clustering areas with similar intensity characteristics, one of the clusters is the set of house candidates in Figure 3b.

Figure 4a shows another example of an image containing difficult-to-parse cultural structures; in particular, note the extreme weakness of many relevant roof edges. Figure 4b contains a cluster of bright enclosures that can be identified as sunlit roofs, Figure 4c shows a corresponding cluster of shaded roof sections, and Figure 4d gives the complete roof structures.

Turning our attention now to linear features, we take the image shown in Figure 1a and apply the model for generic road segments. The system finds the initial set of straight edges shown in Figure 5a, groups them into equidistant parallels, connects those that seem to be collinear or smoothly curving, and then uses them to predict the approximate delineation of the road as shown in Figure 5b. Finally, the predicted shape is optimized with respect to variations in the global width and local curve skeleton, thus yielding Figure 5c.

Finally, we apply the parsing procedure to vegetation clumps. In Figure 6a, we show an image containing typical vegetation clumps, along with one of a set of segmentations in Figure 6b. The initial candidates for vegetation clumps are shown in Figure 6c, and a final selection filtered with respect to image intensity in Figure 6d.

IV. Conclusions

In this work, we have proposed an approach based on generic models and a combination of edge-driven and photometry-based geometric reasoning to delineate several classes of objects in aerial images. Such delineations may be utilized in a variety of ways, but are especially appropriate as input to high-level knowledge-based systems. Since these structures are generic, there is no a priori commitment to any particular labeling or modeling system.

We have devised methods for the following:

- **Integration of Multiple Geometric Data Sources.**

  Data-driven edge extraction and image segmentation processes do not perform well on multiple target objects. We combine multiple information sources and use both edge geometry and enclosed-area characteristics to generate and verify shape hypotheses; we thus make efficient use of the available geometric information in the image.

- **Generic Shape Extraction.**

  For many important tasks, the exact shapes of objects of interest are not known. We define and use generic models to deal with whole classes of objects. Within the context of such models, we recover expected but missing model components using adaptive search techniques, and compensate for photometric anomalies. In particular, we have proposed models for cultural structures, roads, and vegetation clumps, all of which fit into a universal format for model definition and parsing.

The system's effectiveness derives from the definition and use of generic shape models to refine and interpret low-level image information. The clear delineations that we can produce are an essential step toward application-oriented parsing schemes, and provide an adequate basis for rule-based labeling systems that could not function with traditional low-level data alone.

References

R.A. Brooks, "Symbolic Reasoning Among 3-D Models and 2-D Images," Artificial Intelligence Journal **16**, (1981).

J. Canny, "A Computational Approach to Edge Detection," IEEE Trans. PAMI **8**, pp. 679–698 (1986).

M.A. Fischler and R.C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," CACM, Vol. 24, No. 6, pp. 381–395 (June 1981).

M.A. Fischler, J.M. Tenenbaum, and H.C. Wolf, "Detection of Roads and Linear Structures in Low-Resolution Aerial Imagery Using a Multisource Knowledge Integration Technique," Computer Graphics and Image Processing **15**, pp. 201–223 (1981).

P. Fua and A.J. Hanson, "Locating Cultural Regions in Aerial Imagery Using Geometric Cues," Proceedings of the Image Understanding Workshop, pp. 271–278 (December 1985).

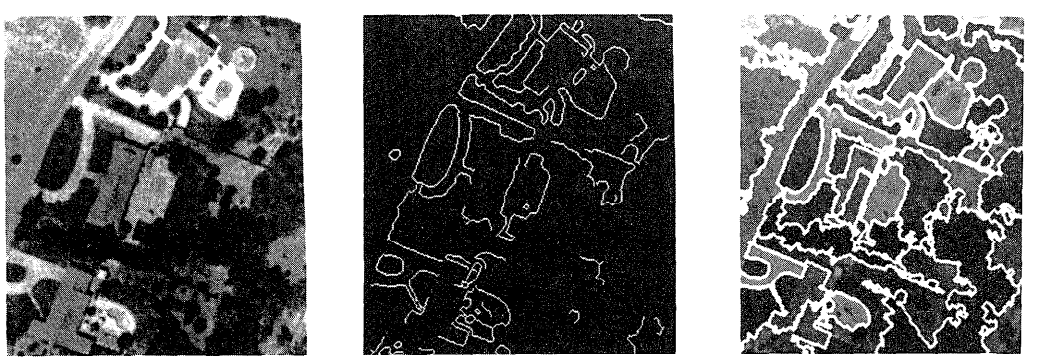
P. Fua and A.J. Hanson, "Resegmentation Using Generic Shape: Locating General Cultural Objects," Pattern Recognition Letters (1986) *in press.*

K.I. Laws, "Goal-Directed Texture Segmentation," Technical Note 334, Artificial Intelligence Center, SRI International, Menlo Park, California (September 1984).

Y. Leclerc and P. Fua "Finding Object Boundaries Using Guided Gradient Ascent," Proceedings of the 1987 Topical Meeting on Machine Vision, Lake Tahoe, CA, pp.168–171 (March 1987).

D. McKeown, W.A. Harvey, and J. McDermott, "Rule-Based Interpretation of Aerial Imagery," IEEE Trans. PAMI **7**, pp. 570–585 (1985).

R. Ohlander, K. Price, and D.R. Reddy, "Picture Segmentation Using a Recursive Region Splitting Method," Computer Graphics and Image Processing **8**, pp. 313–333 (1978).



(a) (b) (c)

Figure 1: (a) A typical aerial image with suburban features. (b) A Canny edge map. (c) A Laws histogram-based segmentation.

Model Component	Buildings	Roads	Trees
Edge definition	Straight	Curved	Jagged
Composite structure definition	Parallel and perpendicular	Parallel	Cluster
Linking geometry specification for composite structures	Rectilinear	Curvilinear	Free form
Area signature specification	Planar intensity	Planar intensity	Planar intensity
Geometric completion model	Straight edge search	Curved edge search	Connecting path search

Table I. Summary of the characterstics of each of three models described in the text.

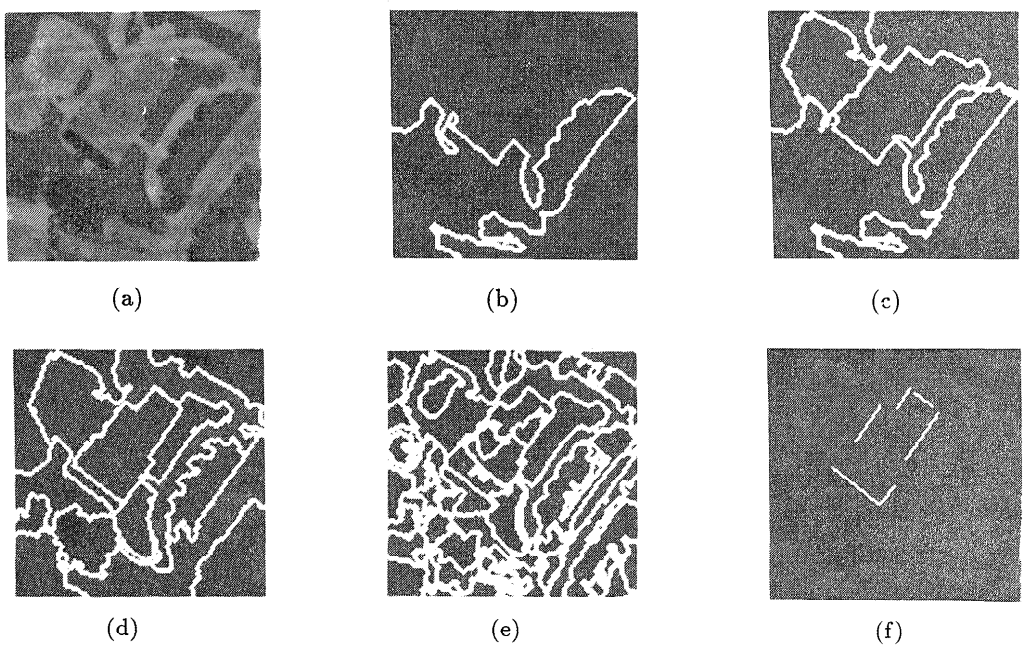


Figure 2: (a) A small image portion containing a cultural structure. (b) An extreme undersegmented partition. (c) An undersegmented partition. (d) An optimum partition for detecting the structure. (e) A highly oversegmented partition. (f) The set of long straight edges extracted from the partition boundaries using the criterion that the edges enclose as large a uniform rectilinear area as possible. These edges form a network.

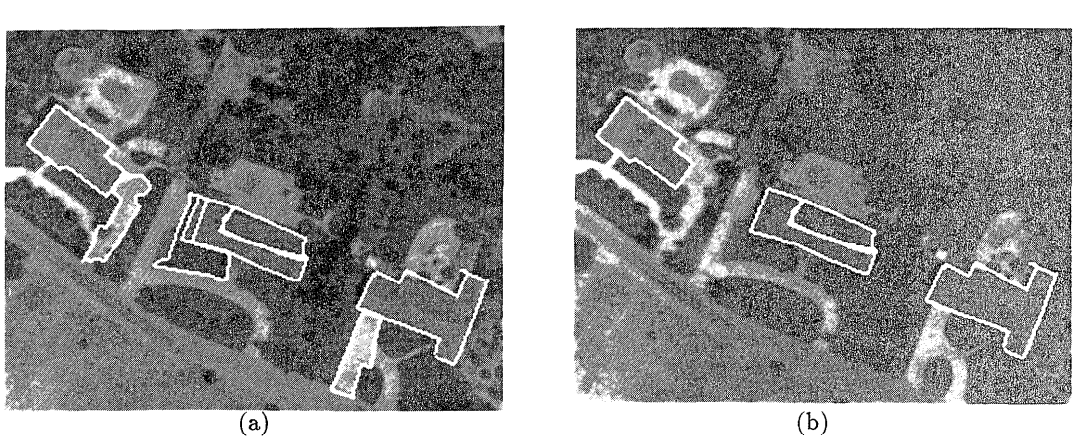
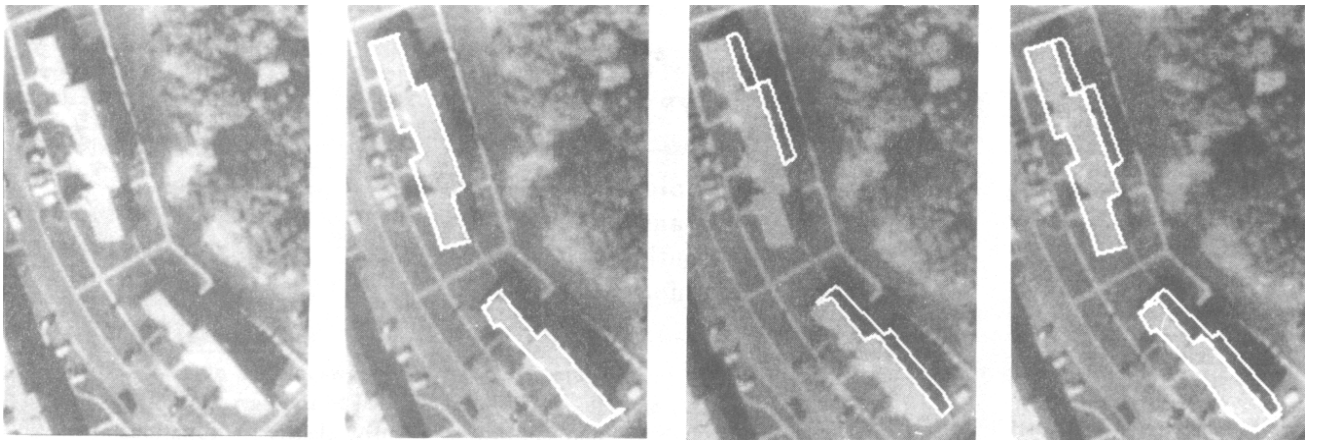
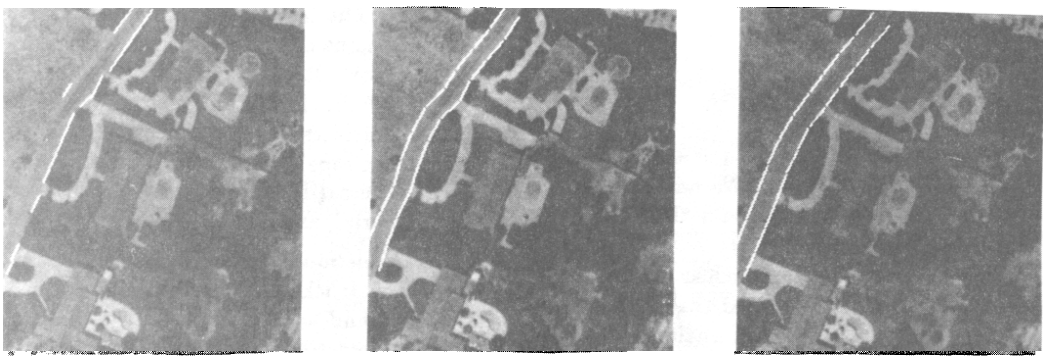


Figure 3: (a) Rectilinear networks meeting a size criterion. (b) House-like networks found by imposing in an additional region uniformity filter.



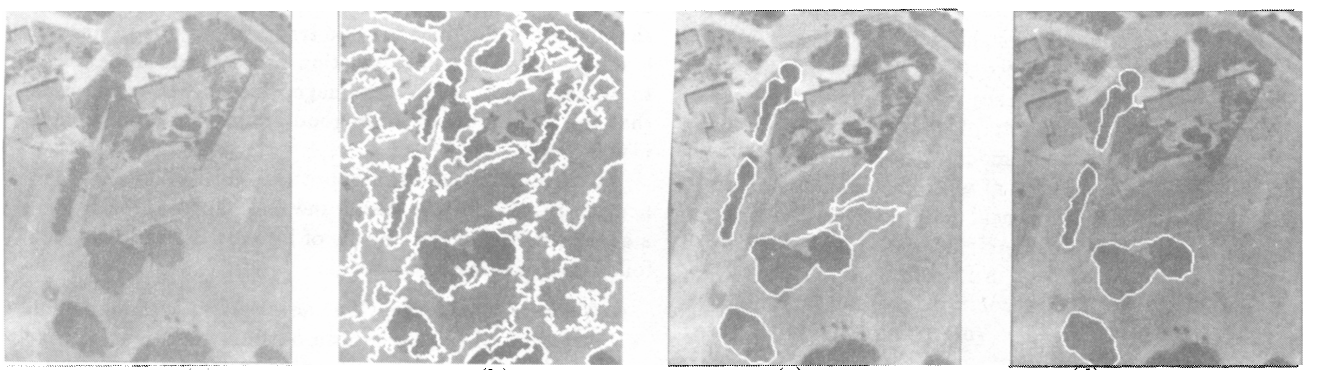
(a) (b) (c) (d)

Figure 4: (a) An image containing complex buildings with some faint edges. (b) A sunlit roof cluster. (c) A shaded roof cluster. (d) House candidates constructed by merging the sunlit and shaded roof candidates.



(a) (b) (c)

Figure 5: An example of a road segment. (a) The edges that are originally grouped together as a possible road structure. (b) Intermediate prediction of the road path given only the initial edges. (c) Final road position optimized to choose best path with the same (variable) width for the entire length.



(a) (b) (c) (d)

Figure 6: (a) An image containing vegetation clumps. (b) One of a family of segmentations used to derive edge candidates. (c) The initial set of vegetation clump candidates. (d) Vegetation candidates selected on the basis of the intensity of the enclosed area.