Task Driven Perceptual Organization for Extraction of Rooftop Polygons*

Christopher Jaynes, Frank Stolle and Robert Collins

Abstract

A new method for extracting planar polygonal rooftops in monocular aerial imagery is proposed. Through bottom-up and top-down construction of perceptual groups, polygons in a single aerial image can be robustly extracted.

Orthogonal corners and lines are extracted and hierarchically related using perceptual grouping techniques. Top-down feature verification is used so that features, and links between the features, are verified with local information in the image and weighed in a graph structure according to the underlying support for each feature.

Cycles in the graph correspond to possible building rooftop hypotheses. *Virtual features* are hypothesized for the perceptual completion of partial rooftops. Extraction of the "best" grouping of features into a building rooftop hypothesis is posed as a graph search problem. The maximally weighted, independent set of cycles in the graph is extracted as the final set of roof boundaries.

1 Introduction

Extraction of polygonal structures from an aerial image is an important step in building detection and model construction. We would like to determine the shape and location of buildings within an aerial image robustly and accurately by extracting the polygons that define rooftop boundaries.

Industrial and urban centers are typically complex and cluttered with structure. Occlusions, strong perspective effects, and variable lighting conditions are a few of the problems when dealing with aerial imagry of typical urban centers. Despite these difficulties, a successful system will discover rooftops that can be used for further image understanding tasks.

2 Task Driven Organization

The power of perceptual organization for the extraction of structure in natural scenes is well know. [?, 4, 6] In our approach, low level features are perceptually grouped to form *collated features* which are then used to hypothesize the final groupings. However, besides this bottom-up approach, each level of the hierarchy may search for features in a task driven, top-down manner. Grouping choices are driven by the goal of the system and the domain. We apply task driven perceptual organization to the process of polygonal rooftop extraction from aerial imagery.

2.1 Overview

The system proceeds in three steps; low level feature extraction, collated feature detection, and hypothesis arbitration. Each module generates features that are used at during the next phase and interacts with lower level modules through top-down feature extraction.

The low level features in this system are perspective image projections of orthogonal corners and straight line segments in the scene.[1] Mid-level collated features are sequences of corners and lines that are grouped together to form *chains*. High-level polygon hypotheses are formed from closed chains.

Because single collated features can be part of several closed polygons, the final set of closed polygons must be searched for the "best" independent set of closed chains. This is done using certainty measures that are

*This work was funded by the RADIUS project under ARPA/Army contract TEC DACA76-92-C-0041 and also by the National Science Foundation grant No. CDA-8922572

[1] That is, while the corners are orthogonal in the world they are not necessarily orthogonal in the image. The prespective projection is known and the shape of the image corner is computable.

maintained throughout the entire grouping process. As each feature is extracted it is assigned a certainty; the final grouping choice is then found as the independent set of closed chains that maximizes the overall certainty.

2.2 The Feature Relation Graph

Features and their groupings are stored in a graph structure called the *feature relation graph.* Low level features are nodes in the graph, and binary relations between features are represented with an edge between the corresponding nodes. Both nodes and edges are assigned a certainty that reflects the confidence of a feature or a feature grouping.

Cycles in the feature relation graph represent grouped polygon hypotheses. The maximally-weighted, set of independent cycles is then extracted from the feature relation graph to discover a set of independent high confidence rooftop polygons.

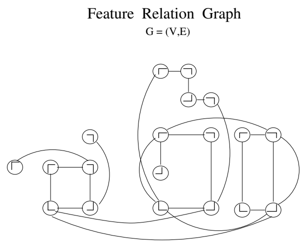

Figure 1: Features are stored in the feature relation graph. Low level features are represented as nodes, collated features as paths, and final polygons as closed cycles.

3 Low Level Feature Extraction

Because we are interested in the detection of human-made structure, orthogonal corners and straight lines were used as low level features. The low level features that are originally extracted are used to form collated features.

3.1 Straight Lines

Straight lines are extracted using two different methods. The primary, bottom-up method for extracting low level straight line features is the Boldt algorithm [1]. This algorithm hierarchically groups edgels into progressively longer line segments based on proximity and collinearity constraints. Figure 2 shows the Boldt lines extracted from a typical aerial urban scene.

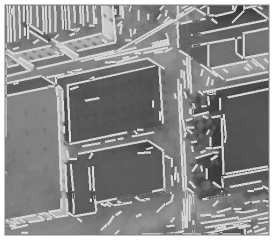

Figure 2: Boldt Lines

Boldt lines are assigned a certainty measure that is calculated during their extraction. The line certainty depends primarily on the contrast of the edge and on the least-squares residual error of the line fit to the grouped edges. For a detailed description of Boldt line certainty, see [1].

Another line detection scheme is used for top-down grouping verification. These *local lines* are extracted when possible groupings between features are being considered. This approach focuses the power of perceptual grouping to a predictive task and avoids reliance on single, globally extracted features.

For example, while attempting to construct a chain feature, it is necessary to discover ad verify if a local line lies between two corners. Each pixel in the image along a connecting line between the two corners is classified as a supporting edgel or nonedgel according to the image intensity gradient, as computed by an oriented Sobel mask, and the variance in the gradient magnitude. This is performed within a rectangular search window between the two corner features.

The final strength of the local line is determined by dividing the number of edgels, L, by the number of pixels in the search line, N. This value is thresholded in order to determine if there is enough edgeness to consider this to be a line. For the results shown here a line threshold of 70% was used.

These local lines are used to verify that a grouping hypothesis, between two corners for example, is justified by evidence in the image. Figure 3 shows a top-down line search between two corner features. The certainty of local lines is based on the contrast of the edge and the percentage of the search window that can be classified as a line.

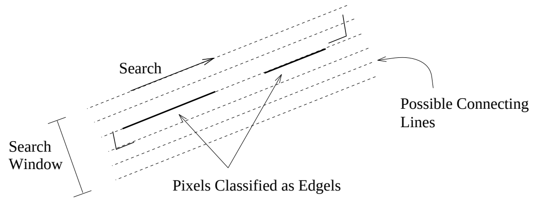

Figure 3: The local line finder is invoked by higher level processes to encourage possible groupings.

3.2 Orthogonal Corners

Our domain assumption is that rooftop polygons will be produced by flat horizontal surfaces with orthogonal corners. This describes a large majority of the building roofs in urban and industrial centers. Therefore, corner features are orthogonal and parallel to the ground plane in the 3D world. Of course, the apparent shape of an orthogonal corner in the image is not invariant under perspective projection, but varies predictably with respect to image position.

To further simplify processing, we assume that a majority of the buildings are aligned according to an approximate city grid. This assumption reduces the set of orthogonal roof corners to be considered to only four, which for the purposes of this paper are labeled NorthEast, NorthWest, SouthEast and SouthWest. The relative orientation of the city grid with respect to the camera completely determines how these four cardinal corner types will appear in the image. Currently, we compute this orientation from the given camera pose; however the city-grid orientation can also be computed more generally using vanishing point analysis [5]. Once this orientation information is known, the perspective transformation mapping 3D orthogonal corners into 2D image corners can be determined, and ideal corner masks can be generated to accurately extract these important low level corner features.

Four different corner masks are generated. Warping is performed by mapping the lines that define the orthogonal corner through the perspective transformation and into the new expected corner angle. This transformation is performed to sub-pixel accuracy.

The four masks are used to detect each of the possible orientations of a roof corner and to classify the corner's type. Corner types are important for later perceptual grouping of compatible corners.

The final masks, then, are typical $n \times n$ ideal corner detectors that are convolved with the image. In the results shown here 7×7 masks were used. Masks of this small size do well in localizing corners and in detecting non-obvious corners (see section 6), however they are sensitive to noise. The performance of template-based corner detection in grey level images with respect to detection, localization, and stability is discussed in [7]. In our system, we allow a large number of false positives when detecting corners and rely on the higher level grouping processes to discard incorrect low level features.

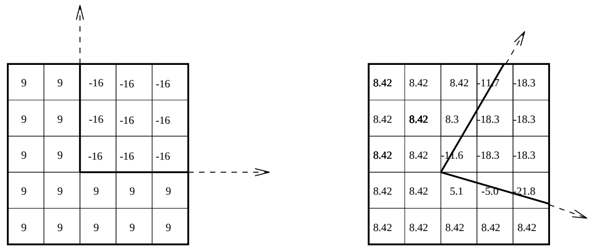

Figure 4: An original 5×5 mask and the corresponding perspectively transformed mask that is convolved with the image.

Once constructed, each mask is convolved with the image and the correlation value at each pixel in the image stored. The correlation measure is used as the measure of "cornerness" of each image pixel and is normalized by the maximum change in grey levels in the image over the size of the mask. Normalization is needed because the corners in typical aerial imagery range from high contrast to very dim. The correlation measure at each pixel is the certainty value for the corresponding orthogonal corner feature that is placed into the feature relation graph.

After convolution of each corner mask with the image, a large array of mask responses will be obtained. A large number of false positives are eliminated by thresholding the absolute value of the mask response. For the experiments an empirical threshold of 60% on corner uncertainty was used. Finally, non-maximal suppression over a 7×7 window of pixels is used to eliminate neighboring pixels that respond to the corner mask only partially. Figure 5 shows the results of orthogonal corner detection in an aerial scene.

Figure 5: Orthogonal Corners

4 Collated Features

Collated features are constructed from sets of lines and corners extracted from the image. A collated feature is a sequence of perceptually grouped corners and lines that form a chain. A valid chain group must contain an alternation of corners and lines, and can be of any length.

Low level features are grouped together according to the standard perceptual parameters of smoothness and symmetry. When such a group is formed, the corresponding nodes in the feature relation graph are connected with an edge. Paths in the feature relation graph become the chain features.

If a low level feature that is needed to complete a strong perceptual group is missing, a top-down feature detector is invoked and the missing feature is searched for in the image. Currently, the system is able to invoke the local line detector to complete a link of two corners if the lines extracted previously were insufficient.

4.1 Feature Groups

Standard perceptual grouping techniques are applied to the low level features in an attempt to group compatible corners and the line between them. These corner-line-corner triples are the "links" that, when followed as paths in the feature relation graph, form chains. Each link can be thought of as a polygon edge hypothesis, while chains are pieces of a polygon hypothesis. Closed chains are a special case and are treated as completed polygon hypotheses.

In order for a link to be formed, three conditions must be met (Figure ??). Given two orthogonal corners, they must first be of compatible types, where compatibility is defined according to corner type and axis information. For example, the east-pointing axis of a NorthWest corner cannot be grouped with a corner of type SouthWest. It is also not possible for a corner to be grouped with another corner of the same type. Secondly, grouped corners must be in proper spatial alignment with respect to each other. That is, corresponding axes of two corners to be linked must be roughly collinear. Finally, a perceptual link can be formed only

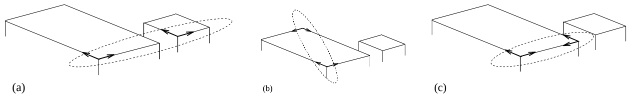

Figure 6: Perceptual grouping of low level features. (a) Incompatible corner types disallow group. (b) Improper alignment of corners. (c) Valid group, supported by line evidence.

if there is evidence for a supporting straight line between the corners. Figure 7 shows an example of perceptually grouped corners.

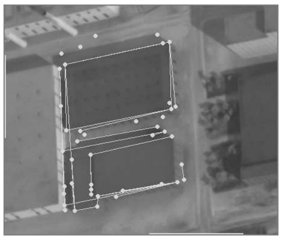

Figure 7: Perceptually Grouped Corner Pairs (Links)

To compute the certainty of a particular chain feature the weight of the corresponding path in the feature relation graph is computed. The certainty of a chain is the sum of the certainties of its parts. Thus, given chain C of length n, the certainty is computed as:

$$\kappa(C) = \sum_{i=0}^{n} \kappa(v_i) + \sum_{i=0}^{n-1} \kappa(e_i) \quad (1)$$

where v_i is node i in the path corresponding to C, e_i is edge i, and $\kappa(F)$ denotes the certainty of feature F.

5 Polygon Groupings

Extraction of final rooftop polygons proceeds in two steps. First, all possible polygons are computed from the collated features. Then, polygon hypotheses are arbitrated in order to arrive at a final set of non-conflicting, high confidence rooftop polygons.

Polygon hypotheses are simply closed chains, which can be found as cycles in the feature relation graph. All of the cycles in the feature relation graph are searched for in a depth first manner.

While searching for closed cycles, the collated feature detector may be invoked in order to attempt closure of chains that are missing a particular feature. The system then searches for evidence in the image that such a virtual feature can be hypothesized. Virtual features are hypothesized according to the parameters of perceptual completion. If a cycle is missing a single corner, for example, a virtual corner will be hypothesized at a position that is constrained both by symmetry and smoothness. Currently, the system is able to hypothesize virtual corners and then invoke lower level feature detectors to confirm the hypothesis.

After addition of a virtual corner, the image is searched by the local line finding algorithm for line data that supports this hypothesis. In the event that the evidence is sufficient, the new corner is generated as a low level feature and used to complete the cycle in the feature relation graph.

In this way, high level features do not rely on the original set of features that were extracted from the image. Rather, as evidence for a polygon accumulates, tailor-made searches for lower level features can be performed. This type of top-down inquiry increases the robustness of the system.

Once discovered, all cycles are stored in a dependency graph where nodes represent complete cycles. Nodes in the dependency graph contain the certainty of the cycle that the node represents. An edge between two nodes in the dependency graph is created when cycles have lower level features in common. The final set of polygons then, must be the set of nodes that are both independent (have no edges in common) and of maximum certainty.

A set of polygon hypotheses extracted from a typical image is shown in figure 5. Notice that, with the generation of virtual features such as corners, we are able to complete a polygon that is partly occluded by a neighboring building.

6 Results

In addition to the examples used above, two more examples are shown in order to demonstrate the system's robustness. Both of the images that were used had a variety of buildings, shadows, and many of the difficulties typically found in aerial imagery.

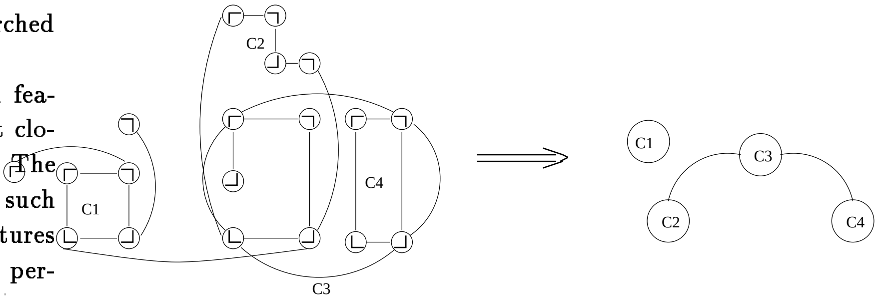

Figure 8: Cycles are extracted from the relation graph and placed, as nodes, into a dependency graph. The maximum independent set of nodes in the dependency graph is the final grouping choice.

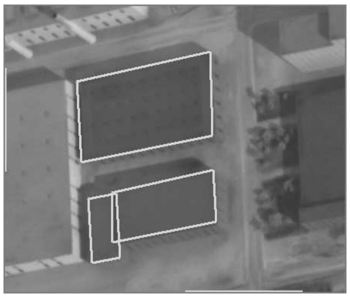

Figure 9: A set of polygon hypothesis extracted from the feature relation graph. When neighboring rooftops partially occlude a polygon, as in the above image, a virtual feature may be generated at the missing corner which can create final polygons that overlap in the 2D projection.

The images used are part of the RADIUS (Research and Development for Image Understanding) model board imagery. As before, the camera model and pose for each image is known. Both images were captured at an approximate height of 10,000 feet above the ground plane. The ground truth data supplied with the model board imagery is used to estimate the accuracy of the system. System accuracy was characterized in several ways:

- Polygons detected versus the true polygons comprising buildings in the image.
- Number of polygon vertices versus number of vertices in building rooftops.
- Average distance between polygon vertices and ground truth rooftop vertices.

Several points, at known rooftop corners in the world were selected for a more detailed performance analysis of the system. Each ground truth point was projected into the image using the known camera pose. The Euclidean image distance between the ground truth image corner and the extracted polygon corner was computed as the lower bound on the true 3D positional error[2]. This distance gives us an estimate of the polygon placement accuracy with respect to the actual building position and the estimated pose parameters.

6.1 First Test Image

The first example image[10] contains six distinct buildings of varying sizes. The strong shadows and different rooftop heights make this an interesting image for testing purposes. Nine ground truth points were used for an estimation of system accuracy. The low level features extracted are shown in figure 11. The value of virtual feature extraction is shown by building A. A shadow falls across a corner of the building and important low level information is missing. The corners are perceptually grouped and a final set of polygons is generated. The results of the test are shown in figure 12.

Results for Example 1	
Rooftops Detected	100.0 %
Vertex Coverage (No virtual features)	88.5 %
Vertex Coverage (With virtual features)	100.0 %
Avg. Vertex Displacement (Pixels)	3.6

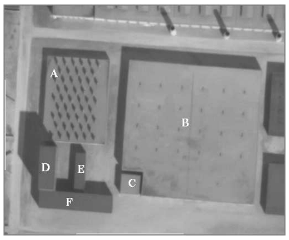

Figure 10: Image used in the first test sequence

6.2 Second Test Image

The second example image contains seven buildings of different sizes and complex shapes. Buildings E and F are very close but are known to be distinct structures. Eight ground truth points were selected for accuracy analysis. As before, the system was run on the image and the performance was analyzed.

[2] At this point, 3 dimensional position of the extracted polygons is not available.

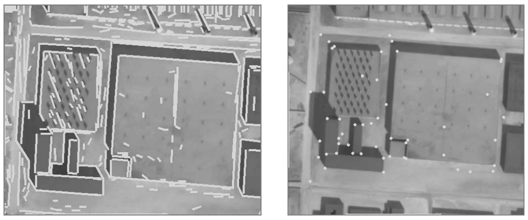

Figure 11: Low level features extracted from the first image: Boldt line data and orthogonal corners.

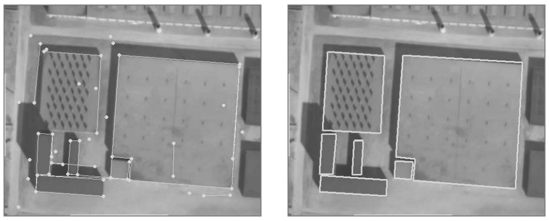

Figure 12: Perceptually grouped corners and the final set of polygon hypotheses

Results for Example 2	
Rooftops Detected	78 %
Vertex Coverage (No virtual features)	86.8 %
Vertex Coverage (With virtual features)	92.1 %
Avg. Vertex Displacement (Pixels)	4.74

Both buildings D and C were not entirely extracted. That is, the final polygons do not match the shape of the underlying structure. Building C is a two level structure and the system failed to discover the lower level rooftop boundary on the right. The corner detector failed to extract crucial low level features at the junction of the two roof heights.

Although the features were extracted on a similar structure to the right of building D, they were not well localized. With small structure, placement error becomes a problem and grouping is difficult. The orthogonal corner detector was designed to extract dihedral corners but can easily be extended to incorporate sun

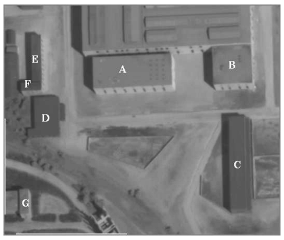

Figure 13: Second test image

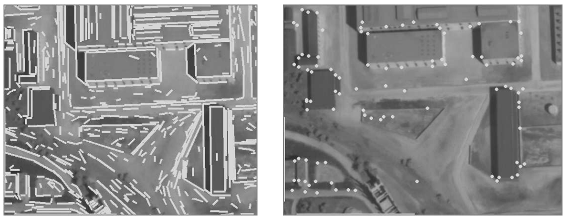

Figure 14: Low level features extracted from the second test image. Boldt line data and orthogonal corners.

angle information for trihedral corner detection. (See Section 7)

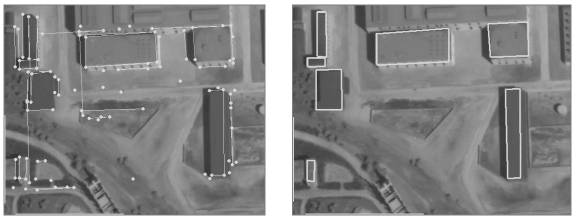

Figure 15: Perceptually grouped corners and the final set of polygon hypothesis

7 Conclusions and Future Work

The results from the proposed approach are encouraging. The system is expected to perform similarly on other aerial images and is currently being tested on a wide variety of aerial photos.

Currently, polygon detection is a piece of the larger aerial image understanding system being developed at U. Mass, Amherst under the RADIUS project. The hypothesized rooftop polygons are verified and refined through multi-image triangulation which computes a height for each polygon in the world. The final set of polygons are extruded to the ground plane for a final volumetric model of buildings.

An improved corner detection mask, that incorporates shadow angles from known sun position, will be constructed. Better methods for solving the maximum independent set problem will be explored, including approximation techniques such as simulated annealing.

The task driven approach to perceptual organization will be expanded to cover more general image understanding tasks. Relaxing restrictions such as flat roofs and orthogonal corners will be investigated so that a more general module can be used to group general structures in aerial imagery.

References

[1] R. Weiss and M. Boldt, "Geometric Grouping Applied to Straight Lines" *IEEE Computer Society on Computer Vision and Pattern Recognition*, 1986.

[2] R. Collins, A. Hanson, E. Riseman, and Y. Cheng. "Model Matching and Extension for Automated 3D Site Modeling." *Proc. DARPA Image Understanding Workshop*, 1992.

[3] M. Herman and T. Kanade. "The 3D MOSAIC Scene Understanding System: Incremental Reconstruction of 3D Scenes from Complex Images" *Proc. DARPA Image Understanding Workshop*, 1984.

[4] A. Huertas, C. Lin, and R. Nevatia. "Detection of Buildings from Monocular Views Using Perceptual Grouping and Shadows" *Proc. DARPA Image Understanding Workshop*, 1993.

[5] J.C. McGlone and J. Shufelt, "Incorperating Vanishing Point Geometry into a Building Extraction System," *Arpa Image Understanding Workshop*, Washington DC, pp.437-448, 1993.

[6] R. Mohan and R. Nevatia, "Using Perceptual Organization to Extract 3D Structures" *Trans. Pattern Analysis and Machine Intelligence*, 1989.

[7] A. Singh and M. Shneier. "Grey Level Corner Detection: A Generalization and a Robust Real Time Implementation" *Computer Vision, Graphics, and Image Processing*, 1990.

[8] V. Venkateswar and R. Chellapa. "Intelligent Interpretation of Aerial Images", *University of Southern California, Dept. of Electrical Engineering Technical Report 137* March 1989.