ARO 28258.3-PH

2



# Finding Regions of Interest for the Detection of Man-Made Objects in Non-Urban Scenes

**INTERIM TECHNICAL REPORT**

(TR-91-5-69)

H. Q. Lu and J. K. Aggarwal

March 1991

U. S. ARMY RESEARCH OFFICE

CONTRACT NO. DAAL03-91-G-0050



THE UNIVERSITY OF TEXAS AT AUSTIN
COMPUTER AND VISION RESEARCH CENTER
AUSTIN, TEXAS 78712

# Finding Regions of Interest for the Detection of Man-Made Objects in Non-Urban Scenes

**INTERIM TECHNICAL REPORT**

(TR-91-5-69)


H. Q. Lu and J. K. Aggarwal



March 1991

U. S. ARMY RESEARCH OFFICE

CONTRACT NO. DAAL03-91-G-0050





Computer and Vision Research Center

The University of Texas at Austin

Austin, Texas 78712

512/471-3259




**APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED**

MASTER COPY KEEP THIS COPY FOR REPRODUCTION PURPOSES

# REPORT DOCUMENTATION PAGE

*Form Approved OMB No. 0704-0188*



**1. AGENCY USE ONLY** *(Leave blank)*

**2. REPORT DATE**
March 1991

**3. REPORT TYPE AND DATES COVERED**
Interim Technical Report

**4. TITLE AND SUBTITLE**
FINDING REGIONS OF INTEREST FOR THE DETECTION OF MAN-MADE OBJECTS IN NON-URBAN SCENES

**5. FUNDING NUMBERS**
DAAL03-91-G-0050

**6. AUTHOR(S)**
H. Q. Lu and J. K. Aggarwal

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
Computer and Vision Research Center
The University of Texas at Austin
Austin, Texas 78712-1084

**8. PERFORMING ORGANIZATION REPORT NUMBER**
TR-91-5-69

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
U. S. Army Research Office
P. O. Box 12211
Research Triangle Park, NC 27709-2211

**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**
ARO 28258.3-PH

**11. SUPPLEMENTARY NOTES**
The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

**12a. DISTRIBUTION/AVAILABILITY STATEMENT**
Approved for public release; distribution unlimited.

**12b. DISTRIBUTION CODE**

**13. ABSTRACT** *(Maximum 200 words)*
This report presents a new approach for the detection of large man-made objects in a rural area using a single monochrome image. In this problem, man-made objects may be unspecified and the appearance of the objects is unpredictable. Prominent features discriminating man-made objects from natural objects are identified. A computational framework for applying perceptual organization and using the prominent features is presented. Techniques are developed to group low level image features hierarchically into a region of interest likely to contain man-made objects. These techniques include linear structure extraction, primitive structure formation, and region of interest location. Each of these methods presents its own unique property and advantage as compared with previous related work. Experimental results are presented using real images that have different kinds of man-made objects and a complex background. We show that the located regions of interest properly enclose the man-made objects in the scenes. The search space is therefore reduced from the whole image to the region of interest.

**14. SUBJECT TERMS**
computer vision, object detection, region of interest, perceptual organization, feature extraction, non-urban scene, geometric structure

**15. NUMBER OF PAGES**
39

**16. PRICE CODE**

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| UNCLASSIFIED | UNCLASSIFIED | UNCLASSIFIED | UL |

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std Z39-18
298-102

# Finding Regions of Interest for the Detection of Man-Made Objects in Non-Urban Scenes

INTERIM TECHNICAL REPORT

H. Q. Lu and J. K. Aggarwal

March 1991

U. S. ARMY RESEARCH OFFICE

CONTRACT NO. DAAL03-91-G-0050

Computer and Vision Research Center
The University of Texas at Austin
Austin, Texas 78712

512/471-3259

APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION UNLIMITED

# Finding Regions of Interest for the Detection of Man-Made Objects in Non-Urban Scenes*

H.Q. Lu and J.K. Aggarwal

Computer and Vision Research Center
Department of Electrical and Computer Engineering
University of Texas at Austin, Austin, Texas 78712

## Abstract

This report presents a new approach for the detection of large man-made objects in a rural area using a single monochrome image. In this problem, man-made objects may be unspecified and the appearance of the objects is unpredictable. Prominent features discriminating man-made objects from natural objects are identified. A computational framework for applying perceptual organization and using the prominent features is presented. Techniques are developed to group low level image features hierarchically into a *region of interest* (ROI) likely to contain man-made objects. These techniques include linear structure extraction, primitive structure formation, and region of interest location. Each of these methods presents its own unique property and advantage as compared with previous related work. Experimental results are presented using real images that have different kinds of man-made objects and a complex background. We show that the located ROIs properly enclose the man-made objects in the scenes. The search space is, therefore, reduced from the whole image to the ROI.

*This research was supported in part by the Army Research Office (contract DAAL03-91-G-0050), and in part by the General Dynamics (purchase order WPC001023).

# 1 Introduction

The computer perception of man-made objects in non-urban scenes is a challenging task in computer vision research. It also presents additional complexities and difficulties beyond the computer perception of objects in a well controlled laboratory or factory environment. This report presents a new approach for the automatic detection of large man-made objects in outdoor non-urban scenes. The methodology presented here is based on *perceptual organization*. The approach hierarchically organizes low level image features into a region likely to enclose man-made objects.

The environment we consider for the acquisition of images is a rural area in daylight hours. There may be large man-made objects, such as bridges, electric transmission towers, and tanks, among natural objects, such as trees, bushes, and vegetation, in the area. The man-made objects are unspecified and their appearances are unpredictable. Thus, we do not know what man-made objects may appear and whether there is a man-made object in the scene. Given a single monochrome image of such a scene, the goal is to automatically detect large man-made objects in the image. Because of the complexity and the variation of man-made objects and the uncertainty of the natural environment, the intermediate goal is to find in an image a *region of interest* (ROI) most likely to enclose man-made objects.

Most existing computer vision work concerning object recognition has focused on problems with pre-specified objects in a controlled environment. For example, most existent vision systems try to recognize objects in a given image with a uniform background and one or multiple objects whose exact models are known to the system [1]. Even with such seemingly well defined problems, numerous obstacles exist and considerable research efforts are being made. Finding man-made objects in a natural outdoor environment adds another dimension of complexity. One cannot arrange the natural environment. Natural objects, such as trees, vegetation, rivers, rocks, and clouds, co-exist in the scene with the possible man-made objects. As Fishler and Strat [2] point out, it is seldom possible to establish complete boundaries between objects of interest in natural scenes, and very few natural objects have compact shape descriptions. Therefore, our problem of detecting man-made objects in a natural environment is, in general, much more difficult than the tasks in a well controlled environment.

Although many researchers have investigated the automatic interpretation of outdoor natural scenes, few of them have investigated the detection of man-made objects [3]-[7] other than buildings and roads. Such research either uses more information, such as color or range, or has better knowledge (models) about the objects. Most of the other works focus on the following two areas: (1) the interpretation of aerial images, and (2) outdoor robot navigation. In the first area, techniques have been developed for detecting complex buildings and roads where buildings are modeled as the combinations of rectangles with uniform intensity and roads are parallel curves [8]-[12]. In the second area, most vision related work concerns road following [13, 14] and position estimation [15, 16]. Usually, a sequence of images is used for the navigation problems. Thus, when interpreting each image, a good initial estimation can be obtained from the interpretation of the previous images [13, 15].

The problem investigated in this research has the following features compared to the above mentioned research: (1) Man-made objects which may appear in the scene are unspecified. (2) In general, most man-made objects have more complicated structures than buildings appearing in aerial images and cannot be easily modeled using rectangles. (3) It is hard to predict the presence of man-made objects in the scene. (4) One monochrome image is used and there is no color or range information. In addition, the appearance change of the man-made objects caused by the view-direction change is a more severe problem than in the top view aerial images. The above aspects make the object description and detection task extremely difficult.

Generally speaking, humans can detect man-made objects easily. Psychologists have found that *perceptual organization* or *perceptual grouping* plays an important role in human perception. Perceptual organization refers to the human's visual ability to derive relevant groupings or structures from input images without prior knowledge of their contents [17]. For example, people can easily detect symmetry, collinearity, and parallelism. If we can derive similar groupings or structures computationally from an input image, that will be very helpful to the task of finding a region of interest for the detection of man-made objects, especially since we do not have any prior knowledge of the image's contents. Many researchers have worked on computational approaches to perceptual organization and applied the concept to various computer vision tasks [3],[8],[17]-[23]. Their work has important impact on our research. However, they concentrated on grouping features and recognizing objects using simple generic models or exact models for specified classes of objects. These approaches are inapplicable to our problem since the objects of interest are unspecified and, in general, have complicated structures.

This report presents a new approach for the detection of large man-made objects in a rural area. The system currently finds in the image a *region of interest* (ROI) likely to enclose man-made objects. Minimal knowledge and information about the objects and scenes are used in this work. Since it is desirable to have a general approach able to handle a variety of man-made objects, we identify prominent features discriminating man-made objects from natural objects. We then present a computational framework of applying perceptual organization and using the prominent features for finding an ROI in an image. Several techniques are derived to group low level image features hierarchically into the ROI. Different from other collinearization methods, our method of extracting linear structures performs line merge and line extension simultaneously. The technique of grouping line segments into parallel primitive structures considers more general situations than most of the previous relevant work. The method of finding a region of interest groups related primitive structures and eliminates those likely to be caused by accidental image events. Various examples, including different kinds of objects in non-urban scenes, are presented. These examples illustrate the ability of this approach to locate useful regions of interest in complex real images.

The ROI is useful for the initial screening of a large environment for man-made object detection. For automatic object recognition, ROI is useful in reducing the search space. When specific object classes are given, primitive structures composing the ROI can be used to match structures of the object models instead of matching individual features. The ROI can also be used in human-machine systems to find the focus-of-attention for human operators to further examine the image. This is applicable to real time operations, such as assisting an aircraft pilot by looking in alternate directions and providing ROIs, and to off-line processing involving a large number of images.

The rest of the report is organized as follows. Section 2 reviews previous research relevant to this work. Section 3 overviews our approach. Section 4 details various grouping techniques developed in this research. Section 5 gives implementation examples, and finally section 6 concludes the report.

# 2 Related Work

This section briefly reviews previous work relevant to our research. The review mainly includes the work of applying perceptual organization to computer vision tasks and the detection of man-made objects in outdoor natural scenes.

Perceptual organization has been studied since the early part of this century. Gestalt psychologists studied a large number of grouping phenomena and roughly categorized them into several *Gestalt Laws* (or *grouping rules*) [17, 24], which include proximity, similarity, continuation, closure, symmetry, and familiarity. In the past several years, perceptual organization has been introduced into computational vision research and the functional role of the former in the latter has been addressed [17, 25]. Lowe [17] argues that the most important functions of perceptual organization include segmentation, three-space inference, and the indexing of world knowledge. All of these lead to the reduction of search space for object recognition. McCafferty [18] formulates perceptual organization as an energy minimization problem. He quantifies the Gestalt Laws by defining individual energy terms. However, the selection of the weightings for these energy terms presents problems.

Recently, perceptual organization has been applied to solve practical computer vision problems [3, 8, 19, 20, 23]. Mohan and Nevatia [20] apply perceptual organization to segment images into visible object surfaces. They also investigate the detection and description of complex buildings in aerial images [8]. Assuming that roofs are the essential building structure seen in the image, they model the roof as a combination of rectangles. Baker *et al.* [3, 4] present an approach for the detection of concrete bridges. The straight line segments, once detected, are grouped into parallel lines. Intrinsic rectangles are extracted from the parallel lines. Color cues are then used to restrict the candidate artifacts and to produce confidence measures. However, the rectangle-type model may be unsuitable to other man-made objects in outdoor scenes with more complex structures. In this investigation, we deal with objects with complex structures as well as those with simpler structures.

Reynolds and Beveridge [23] examine the problem of searching for geometric structures in natural scene images. Straight lines are grouped using the geometric relations of collinearity, parallelism, orthogonality, and spatial proximity. The connected components of a graph representing lines and their relations are illustrated to correspond to significant geometric structures in the image. However, some components may contain many different image events. In addition, this method may not find some simple geometric structures, such as parallelograms except rectangles. Different from Reynolds-Beveridge's method, the work presented in this report groups low level features into primitive structures. The spatial relations among the primitive structures are used to find a region of interest most likely to enclose man-made objects. The advantages of finding primitive structures and identifying their relations are the abilities to extract a variety of geometric structures, to establish higher level relations among image features, and to use regional information.

Jacobs [19] presents a system called GROPER, which recognizes two dimensional objects using a library of many different objects. GROPER applies perceptual organization to reduce the search space for matching scene objects with models. GROPER is designed for a simplified world that contains only 2D polygonal objects, whereas the objects encountered in our research are 3D objects with complex structures.

Various techniques are presented in the literature for applying perceptual organization to group lower level image features, such as edge points, into higher level structures, such as straight lines and curves, and to detect junction, collinearity, parallelism, and symmetry [9, 21, 22, 26, 27]. Although similar to some of these techniques, the methods applied in our research for the first two levels of grouping have their unique properties and advantages. We discuss the differences and advantages in Section 4 after describing each of the methods.

There are other works concerning the interpretation of natural scenes other than those using perceptual organization. Brooks [5] presents the identification of aircraft in aerial images using ACRONYM, a model-based vision system. In ACRONYM, generic object classes and specific objects are represented by volumetric models using generalized cones along with sets of constraints relating to model parameters. Fua and Hanson [10, 11, 12, 28] present a sequence of papers concerning extracting features and locating *general cultural objects*, such as buildings, in aerial images. A new approach based on information theory to evaluate the correspondence between generic models and shape hypotheses in an image is presented in [10, 28]. Generic models for buildings are formulated and experimental results on several aerial images are presented. Beveridge *et al.* [29] present a method for identifying known 2D models in imperfect line data. The method is applied to complex outdoor scenes and good matches are demonstrated. However, the above approaches are unsuitable to detection problems where the potential objects are unspecified, as in the case of our research. Chu *et al.* [6, 7] present a system called AIMS to detect and recognize man-made objects in outdoor scenes. Multiple sensing modalities (range, intensity, velocity, and thermal) are integrated in AIMS to improve both low-level image segmentation and high-level image interpretation.

In summary, previous works concentrated on extracting groups of features; using simple generic models for specified classes of objects; recognizing objects with exact models; or using additional sensing information. Our task is to detect objects with minimal knowledge and information about the objects and scenes. The objects are unspecified and may have much more complicated structures than the objects considered in the previous research. Hence, none of the previous works, or any simple combination of these works, are applicable. A new approach must be developed, which we present in the next two sections.

# 3 Overview of the Approach

This section describes the basic concepts and an overview of our approach. The most important concepts are discriminative features and perceptual organization. The approach essentially organizes those features indicating man-made objects into structures, and finds the image region in which related structures reside.

In the initial stage of the research, we applied a line detector [30] to a number of images, including the one shown in Figure 1. Figure 2 illustrates the lines detected for the image in Figure 1. Obviously, in Figure 2, many of the tower's linear structures are fragmented and most junctions are broken or missing. Consequently, we usually cannot get a perfect line drawing of the tower. However, when this image (Figure 2) is presented to a person who did not see the original image (Figure 1) and has no knowledge of image processing, he can recognize the tower in the image without any difficulty. We believe that the perceptual grouping plays an important role here. Human vision has derived relevant groupings and structures from the line image. If a computational approach can be developed to derive a similar grouping, this will lead to, or at least be very helpful to, the detection of man-made objects. The questions are – what should be grouped in the image and how should the grouping be performed?

Since the goal is to detect man-made objects in natural scenes and since the objects are not particularly specified, features must be found that discriminate man-made objects from natural objects in an image. We believe that the most prominent features are the apparent regularity and relation. Most man-made objects have *linear structures* (LS) or linear boundaries. Such linear structures usually form certain regular patterns, such rectangles, parallels, and polygons, called *primitive structures* (PS). Primitive structures are usually related to each other and form the man-made objects. After line detection, many of such regularity and relations remain in the resulting image. In comparison, most natural objects do not have linear structures, and lines extractable from their images are usually randomly distributed.

The identification of the discriminative features indicates that to detect man-made objects, we should find linear structures in the image, regular patterns formed by the linear structures, and the regions occupied by such structures. Therefore, the computational

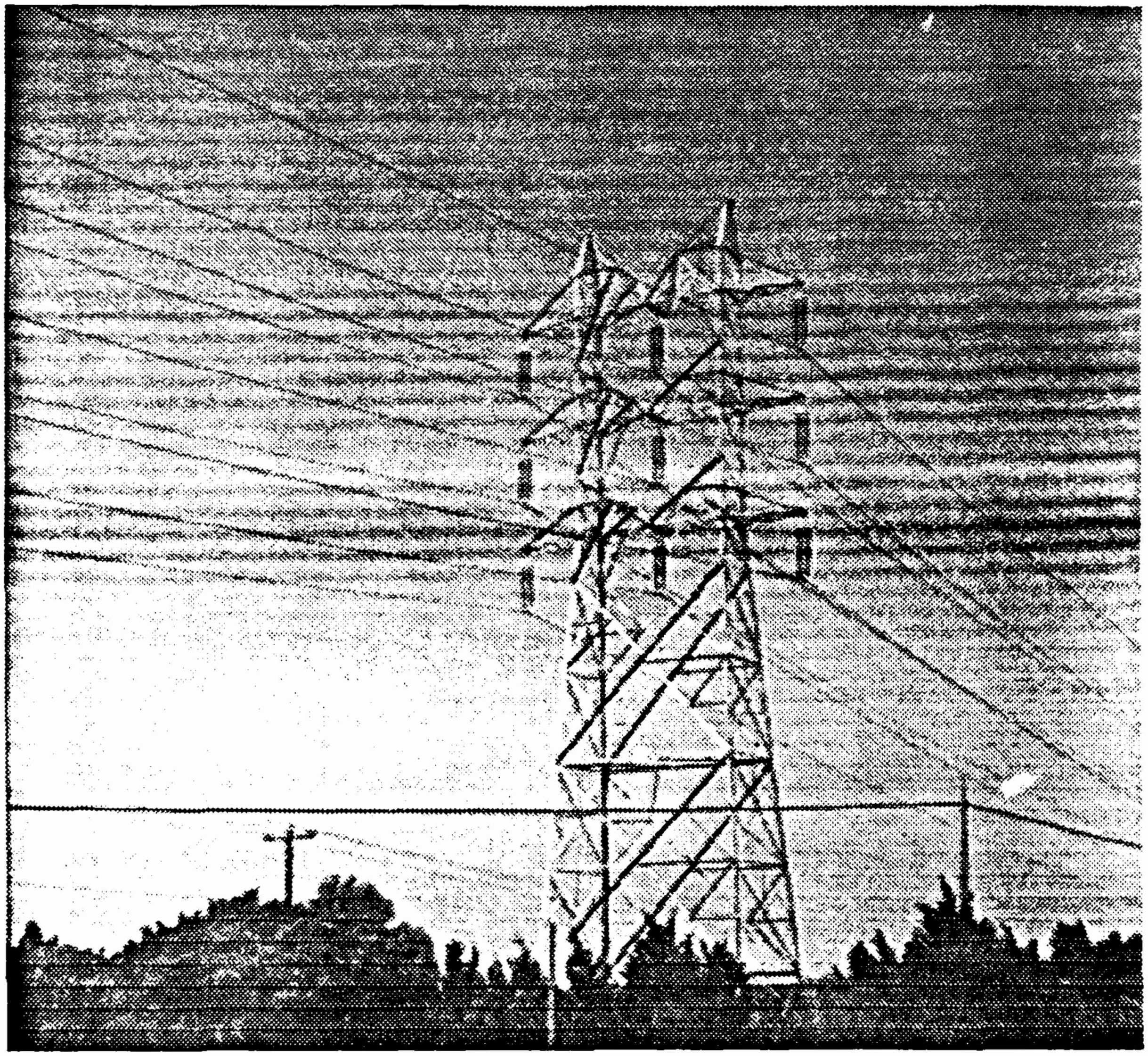


Figure 1: An image with an electric transmission tower.

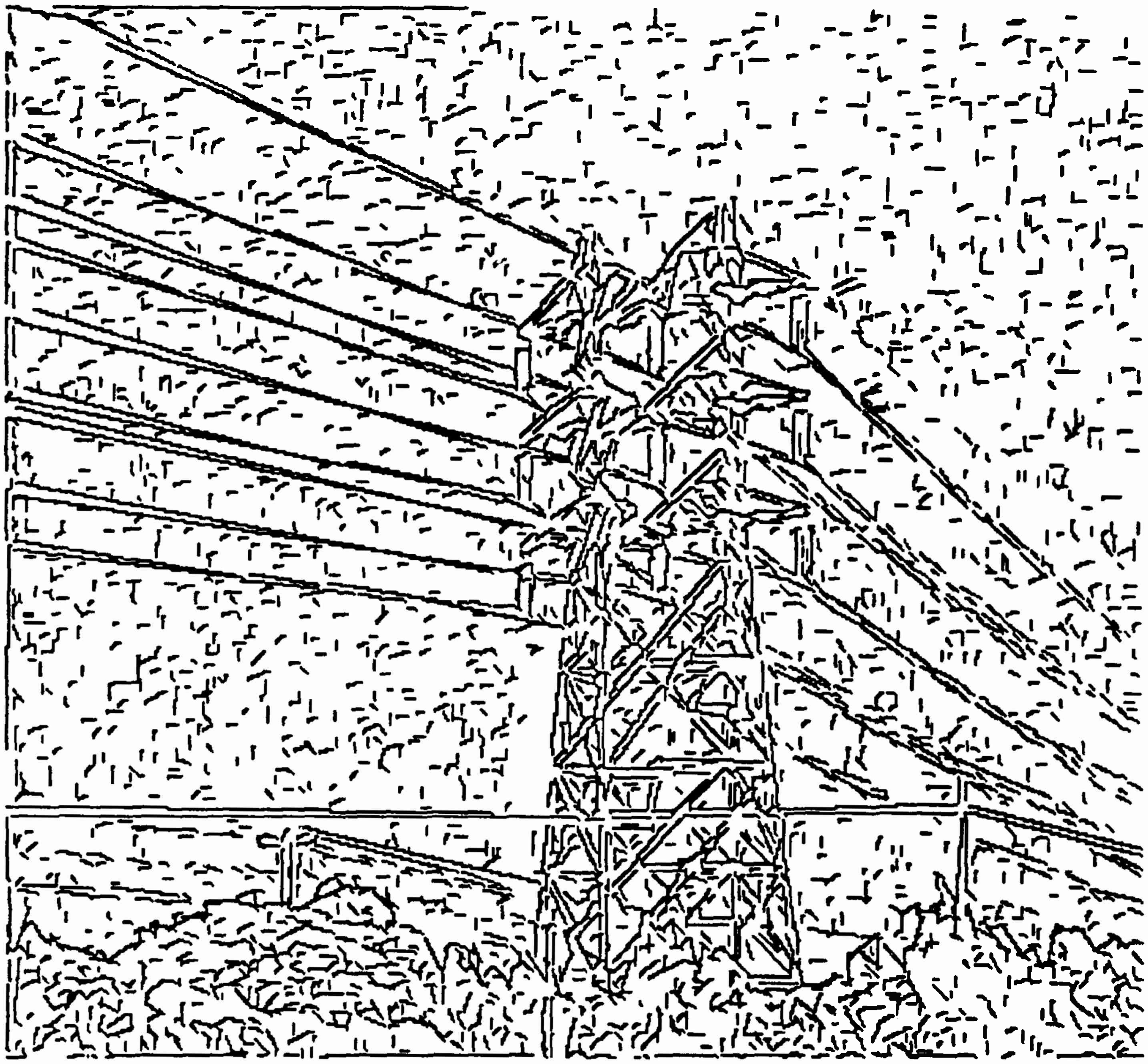


Figure 2: Linear edges detected from the tower image.

framework developed in this research includes three major steps: (1) grouping low level image features into linear structures, (2) organizing linear structures into primitive structures, and (3) finding related and *non-accidental* primitive structures. The non-accidentalness of an image event is an important concept. Lowe [17] argues that perceptual groupings are useful when they are unlikely to have arisen by an accidental viewpoint or position and, therefore, are likely to reflect meaningful scenic structures. Hence, the image region containing a collection of related non-accidental PSs is more likely to enclose man-made objects. We call such a region the *region of interest* (ROI). Corresponding to the above three steps, the approach has three modules: LS Extraction, PS Formation, and ROI Location. The following outlines the functions of each module:

- *LS Extraction*: This module first extracts linear edges (or line segments) from the input image. These edges are the basic information used for the perceptual organization. Linear structures are then extracted from the linear edges.

- *PS Formation*: This module finds primitive structures in the image by grouping the line segments satisfying certain criteria. Each PS may be an evidence indicating man-made objects. Only parallel PSs are implemented at this time.

- *ROI Location*: This module finds a region most likely to enclose man-made objects by grouping spatially related PSs, eliminating isolated PSs likely to be caused by accidental image events, and identifying the image regions occupied by these grouped PSs.

The region of interest hypothesizes the existence of man-made objects. Further analysis of the ROI may lead to the recognition of an object or to the rejection of the hypothesis. When specific object classes are given and models are established, primitive structures in the ROI can be matched to object models. The search space will be considerably reduced, since a primitive structure is a set of features with certain relations and, hence, implies more constraints. The ROI can also be used in human-machine systems to find the focus-of-attention so that human operators can further examine the image.

We should point out that the term *primitive structure* is also used in [25], where the primitive structure represents a larger class of entities, including edges, regions, parallelism, symmetry, repetition, and so forth. The primitive structure in this report represents the regular patterns formed by straight line segments. Basically, we consider the finer classification of the structural entities, since these entities are not grouped at the same level. For example, a line segment is a grouping of edge points whereas a parallel is a grouping of line segments.

# 4 Implementation

The previous section presents an overview of our approach and describes the functions of each system module. This section describes the techniques used in the various modules and compares these techniques to existent work.

## 4.1 Linear Structure Extraction

The module first detects straight line segments from the intensity image using Burns' line extraction algorithm [30]. Due to the imaging process, lighting conditions (e.g., sun direction), digitization, and the line detection process, many of the line segments along the linear structures of the tower are fragmented, skewed, or displaced, and some of them are missing. In many cases, two sets of near-parallel lines in the line image correspond to one linear structure in the intensity image. These linear structures should be recovered in order to properly interpret an image. Hence, the module post-processes the extracted line segments.

We wish to find a *representative line* to a set of closely bunched and similarly oriented linear edges, since they represent the linear structure of an object at a higher granularity level than themselves [8]. For example, we want to extract the linear structure implied by a set of line segments shown in Figure 3.

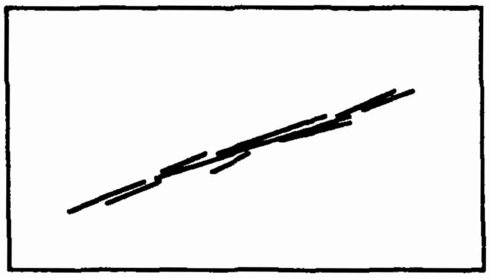


Figure 3: Extracting a linear structure from a set of line segments.

However, most of the collinearization techniques, such as [9, 21, 27], are unsuitable for extracting linear structures, since they only link near-collinear lines by examining the neighborhoods of the end points of each line; that is, they perform a line extension. Our objective of extracting linear structures is quite similar to that presented by Mohan and Nevatia [8]. In [8], the space around each line segment is *folded* onto the segment repeatedly to obtain a *single* line representing the grouped line segments. However, this folding technique does not consider the possible extension of the line segment. The techniques presented below perform both folding and line extension.

We developed two methods for line structure extraction. The rationales of both methods are the same. That is, close lines with similar orientations are likely to come from the same linear structure and, hence, should be merged into one line. This is also consistent with the Gestalt Laws of proximity, similarity, and continuation. However, the grouping procedures of the two methods are quite different. The first one, the ***neighborhood method***, groups lines in the neighborhood of each line, and the second one, the ***classification method***, groups lines by classifying them. Each of these methods can be used alone. Using both may result in better results. The advantages of this technique are that (1) the grouping of closely bunched near-parallel line segments and the extension of the line segments are performed simultaneously; and (2) both local and global information are used in the grouping.

### 4.1.1 The Neighborhood Method

This method iteratively groups and merges lines in a neighborhood of each line segment. The neighborhood grows as the length of the line increases. We first introduce definitions and then present the technique.

Two lines have *similar* orientations if the angle between the two lines is less than a threshold, *similarity-angle*. The *neighborhood* of a line segment $L$ is a symmetric elongate region with $L$ as the *medial axis* of the region [31]. Two line segments are *close* if at least one end point of one line segment is in the neighborhood of the other line segment.

The idea of the grouping process is as follows. The neighborhood of each line is searched to find all the lines with orientations similar to the current line, called the *base line*. The resulting set of lines, including the base line, are then replaced by a *representative line*. The process continues until no replacement occurs.

To reduce the search space, a line segment is represented by its two end points and is indexed by the image pixels corresponding to the end points. When searching for lines close to a base line, the neighborhood of the base line is searched. Hence, only those lines whose end points fall in this neighborhood are examined. After a set of lines $S$ is found with respect to a base line $L$, with $L \in S$, a representative line $L_r$ of $S$ is computed. $L_r$ passes through the point that is the geometric center of the line segments in $S$. The orientation of $L_r$ is the length weighted average of the orientations of the lines in $S$. To determine the end points of $L_r$, all the end points of the line segments in $S$ are orthogonally projected onto $L_r$. The two furthest apart projection points are the end points of $L_r$. $L_r$ replaces the lines in $S$. The process continues until no merge occurs.

#### 4.1.2 The Classification Method

The basic idea of this method is to classify line segments according to orientation, collinearity, and proximity. The method is implemented with the following three steps:

**Step 1: Orientation Classification.** The range of the orientation angles of all the line segments is divided into uniform overlapping intervals. The length of each interval equals the *similarity-angle* and an one degree overlap exists between adjacent intervals. Lines whose orientations fall into the same interval are classified into one cluster. All the clusters are processed separately using the same mechanism.

**Step 2: Collinearity Classification.** Let $G$ be a cluster. For each line $L \in G$, all the other lines in $G$ approximately collinear with $L$ are grouped into a set $S$ that includes $L$. Specifically, let $U$ be a strip along $L$ with $L$ in the middle, as Figure 4 shows. Then the

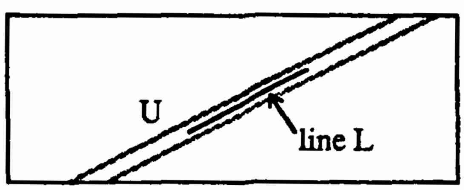



Figure 4: A strip formed around the line $L$.

set $S$ is

$$S = \{L, L_i | L_i \in G, Len(L_i \cap U) = \delta Len(L_i), i = 1, 2, ...\}$$

where $Len(\cdot)$ represents the length of a line and $\delta = 80\%$. The above equation says that a line in $G$ is in $S$ if 80% of the line is in $U$. A representative line $L_r$ of $S$ is found, as in the neighborhood method, except that the end points are not computed.

**Step 3: Proximity Classification.** This step merges spatially closed line segments in $S$. All the lines in $S$ are orthogonally projected onto $L_r$. The lines whose projections overlap or are close are merged into one line. The new line is computed, as in the neighborhood method.

The above process iterates until no lines can be merged.

### 4.1.3 Summary

We presented two methods for extracting linear structures from line segments. Both methods are based on the Gestalt grouping rules of proximity, continuity, and similarity. However, their grouping procedures are quite different. The neighborhood method starts grouping from local areas and extends to longer regions, whereas the classification method starts from the mass of lines and gradually focuses to local groupings. Both algorithms perform folding as well as line extension. The first method emphasizes folding and the second one line extension. Both algorithms run in an iterative fashion. They always terminate after a finite number of iterations, since there is a finite number of lines and since this number declines in each iteration. The LS Extraction module implements these two methods one after another such that both local and global information are properly used. The output of the LS Extraction module is a set of lines including the representative lines of the grouped line segments and the un-grouped line segments.

## 4.2 Primitive Structure Formation

This module finds primitive structures from the line segments. Currently, we have only implemented the extraction of parallel PSs. Previous research [3, 8, 23] also used parallel lines for perceptual grouping. However, the definition of the parallel PS here is different from the definition of parallel lines in [3, 8, 23]. The most distinctive difference is the requirement for overlapping between parallel lines. Usually, only a certain overlapping using an orthogonal projection is required [3, 8, 23]. For example, in Figure 5-(a), the

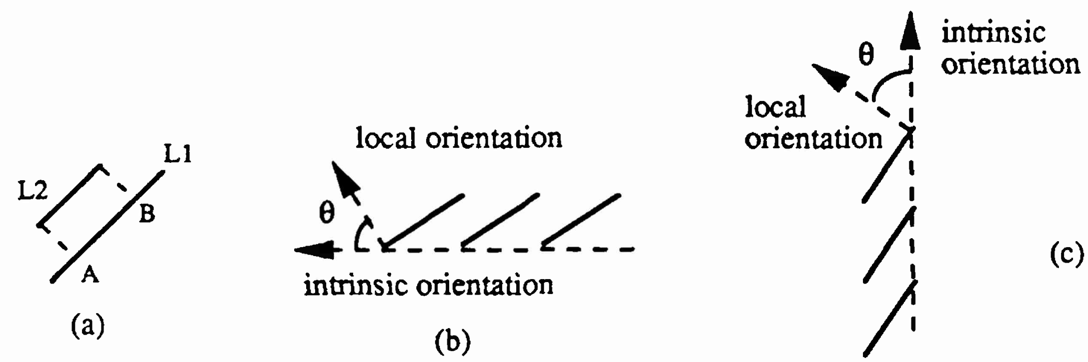



Figure 5: Lines with overlapping in different directions.

overlapping between the line segments $L_1$ and $L_2$ using orthogonal projection is the line segment $AB$. However, according to such a definition, when the ***intrinsic orientation*** of a set of similarly oriented lines [32] is different from the local orientation of each line, these lines may not be grouped. As a result, many apparent parallel lines, such as those shown in Figure 5-(b) and (c), will not be identified when there is a difference $\theta$ between the intrinsic and local orientations and the overlapping between lines is small using the orthogonal projection. Such situations arise very often in practice. For example, a set of 3D parallel lines may fall into this situation under the 2D projection of the imaging process. To solve this problem, we define two additional overlapping conditions in two perpendicular directions. The idea here is somewhat similar to the ***θ-aggregation*** of Marr's earlier work [32]. But the concept in [32] has not been pursued further.

The parallel PS is a set of lines, $S = \{L_1, L_2, ..., L_m\}, m \geq 2$, that have ***similar*** orientations. In addition, for each line $L_i \in S$, there exists a line $L_j \in S$ such that

1. $L_i$ and $L_j$ have *similar lengths.*

2. $L_i$ and $L_j$ have a *sufficient overlap* in one of the three projections, i.e.,

$$\frac{OL(Proj_x(L_i), Proj_x(L_j))}{Len(Proj_x(L_k))} > \delta$$

   where $Proj_x(L_i)$ is the projection of $L_i$ onto the $x$-axis, $OL(L, L') = Len(L \cap L')$ is the length of the overlap, and $L_k$ is the shorter line of the $L_i$ and $L_j$; or

$$\frac{OL(Proj_y(L_i), Proj_y(L_j))}{Len(Proj_y(L_k))} > \delta$$

   where $Proj_y(L_i)$ is the projection of $L_i$ onto the $y$-axis; or

$$\frac{OL(Proj_o(L_i), L_j)}{Len(Proj_o(L_i))} > \delta$$

   where $L_i$ is assumed to be the shorter line, and $Proj_o(L_i)$ is the orthogonal projection of $L_i$ onto $L_j$.

3. $L_i$ and $L_j$ are *relatively close.*

The above conditions are based on perceptual organization rules, such as proximity, parallelism, and similarity. Condition 1 requires that two line segments have similar a length. Two lines with very different lengths are unlikely to come from a parallel structure and are unlikely to be perceptually grouped. Condition 2 indicates that two line segments in a PS should have a sufficient overlap. With this set of overlapping conditions, lines in

Figure 5-(a), (b), and (c) can all be properly grouped. Condition 3 restricts the relative distance between the two line segments.

The PS Formation module groups line segments satisfying the above conditions into parallel PSs. To avoid the brute-force search, we adopt a strategy similar to the classification method. Line segments with similar orientations are first classified into clusters. A further grouping is performed within each cluster to find all parallel PSs.

## 4.3 Region-of-Interest Location

Object detection is actually a process of evidence collection. As we discussed earlier, each PS may be evidence of man-made objects. Hence, this module collects PSs by grouping spatially-closed PSs and locates the image regions occupied by these grouped PSs.

The rationale of this level of grouping is the following: (1) Spatially-closed PSs are likely to be related and to reflect meaningful structures. For example, an electric transmission tower is a connected entity and, hence, the PSs resulting from the image of the tower are spatially-closed. On the other hand, spatially closed PSs are more likely to be perceptually grouped according to the *proximity* grouping rule. (2) Some PSs may be caused by the accidental image relations [17] of natural objects. For example, line segments extracted from a cluster of tree leaves may accidentally form a parallel PS. Such PSs tend to be randomly and sparsely distributed and are unlikely to form meaningful structures, since they arise accidentally and since most natural objects do not have regular patterns consisting of straight lines. Hence, grouping spatially-closed PSs tends to eliminate isolated PSs caused by the accidental image events. (3) This process locates the most likely man-made object region in the image, again, since man-made objects usually consist of spatially closed PSs.

Each PS occupies a region in the image. For example, a PS containing two parallel lines occupies a trapezoidal region. The regions of spatially closed PSs tend to overlap, touch, or be close, whereas the regions of sparsely located PSs are usually isolated. Therefore, this module groups PSs whose regions overlap or touch. The largest image region the grouped PSs occupy is selected as the region of interest.

The technique used in this level of grouping is primarily based on computational geometry [33]. The region of a PS is defined as the convex hull of the line segments in the PS. Let $C(x)$ represent the convex hull of a set $x$. First, find the convex hulls of each PS. Let $P_i = C(PS_i)$ be the convex hull for the $i$th PS and $V$ be the set of all $P_i$s. Each $P_i$ has a flag indicating its status. Set $Flag(P_i) = active$ for $\forall P_i \in V$.

Then, iteratively merge PSs whose regions overlap or touch by merging the intersecting convex hulls. For each $P_i \in V$ with $Flag(P_i) = active$, find a set $W$ of all convex hulls intersecting with $P_i$:

$$W = \{P_j | P_j \in V \text{ and } Flag(P_j) = active \text{ and } P_i \cap P_j \neq \emptyset\}.$$

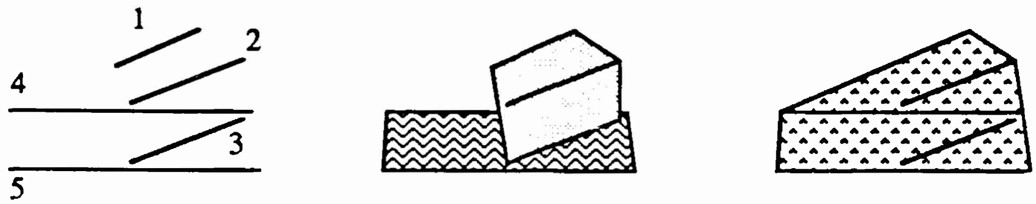
If $W$ is not empty, a new convex hull is found which is the convex hull of $P_i$ and $W$:

$$P_i' = C(P_i \cup (\cup_{\forall P_j \in W} P_j)).$$

Then we set $Flag(P_j) = inactive$, $\forall P_j \in W$, and let $P_i = P_i'$ and $Flag(P_i) = active$. The process continues until no new convex hulls can be formed.

Figure 6 shows an example of regions of two PSs overlapping. In Figure 6-(a), lines 1, 2, and 3 form a PS, and lines 4 and 5 form another PS. Figure 6-(b) shows the convex hulls of these PSs. The two convex hulls intersect and, hence, are merged. Figure 6-(c) shows the new convex hull containing the two PSs.

Currently, the largest resulting convex polygon is considered the region of interest. Since lines are represented by their end points, the convex hull of a PS or a set of PSs can be easily found by an existent algorithm, such as *Jarvis' march* [33].




(a) (b) (c)

Figure 6: Two primitive structures whose regions overlap.

The next operation is to determine whether two convex polygons, $P$ and $Q$, intersect. If $P$ and $Q$ intersect, then either $P$ contains $Q$, $Q$ contains $P$, or some edge of $P$ intersects some edge of $Q$ [33]. It is straightforward to prove the following two sufficient conditions for polygon intersection: (1) Two polygons intersect if some vertex of one polygon is inside the other polygon. (2) Two polygons intersect if some edge of one polygon intersects an edge of the other polygon. In determining if two given convex hulls overlap, we first use condition (1). If no decision can be made, we then use condition (2). The reason for this sequence is that for convex polygons, point inclusion is easy to determine [33], and condition (1) actually covers many cases. There is also an efficient way to check the intersection of line segments [34]. Therefore, the grouping of PSs using convex hulls can be implemented efficiently.

Reynolds and Beveridge [23] present a method for grouping significant geometric structures in an image using a graph. The grouping mechanism used in [23] is different from the one developed in this work. In [23], a graph is built to represent geometric relations among all the lines. The largest component of the graph under certain relations is illustrated to represent significant geometric structures. The grouping procedure presented in this report is hierarchical in nature. Lines are grouped into primitive structures, which are then grouped into a region of interest. Defining primitive structures enables us to establish higher level relations among image features, such as relations among primitive structures; to extract a variety of geometric structures; and to use area information. Involving the area has the potential advantage of incorporating other information, such as color and texture, in the grouping process. In addition, certain simple geometric structures which may not be extracted using the method in [23] can be easily found using our method. For example, in Figure 7, four lines form a parallelogram. Lines 1 and 3, and lines 2 and 4 have parallel relations, respectively. But these two sets of parallel lines belong to two different connected components, using the method in [23] if no other lines exist having relations that could connect these two sets. Since the regions of these two parallels overlap, our method groups them, although we have not represented them explicitly.

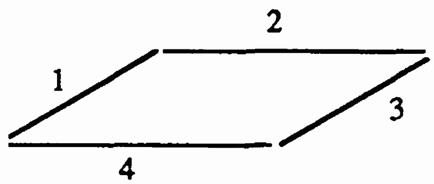



Figure 7: Four lines form a parallelogram.

In summary, this section presents various techniques to group low level image features hierarchically into a region of interest likely to enclose man-made objects. These techniques include linear structure extraction, primitive structure formation, and region of interest location. Each of these techniques presents its own unique property and advantage compared to previous related work, as we discussed in each of the subsections.

# 5 Experimental Results

This section presents several examples of finding regions of interest for the detection of large man-made objects in non-urban scenes. In the following examples, the *similarity-angle* is 5 degrees; the *half-width* of the neighborhood (in the neighborhood method) or the strip (in the classification method) of a line segment is 2 pixels wide; and $\delta$ is 60%. A single monochrome image is used for each of the examples. All the image sizes are $512 \times 512$.

Figure 1 shows an image containing an electric transmission tower. The image is processed by the Burns' algorithm generating line segments (Figure 2). These lines enter the LS Extraction module where very short lines (less than 4 pixels long) are eliminated and lines likely to come from the same linear structures are merged into one line. Figure 8 shows the resulting lines after applying this module. These straight lines then enter the PS Formation module for identifying parallel primitive structures. Figure 9 illustrates the parallel groups thus obtained. These parallel PSs are the only information used to find the region of interest. Figure 10 shows the located region of interest overlapped on the line image (Figure 8). The region of interest is bounded by a polygon displayed with a bold outline. From Figure 10, we see that the tower and most of the transmission lines are properly included inside the polygon.

Comparing the original image (Figure 1) with the edge image (Figure 2), we notice that many linear structures in the image correspond to sets of similarly oriented lines. This is most obvious for the transmission lines on the tower's left side. This phenomenon is caused by the nature of the Burns' algorithm, since the gradient of the intensity image changes rapidly on both sides of a thin linear structure, such as a transmission line. From Figure 8, we see that many of the linear structures are recovered by the LS Extraction module, especially the transmission lines on the tower's left side.

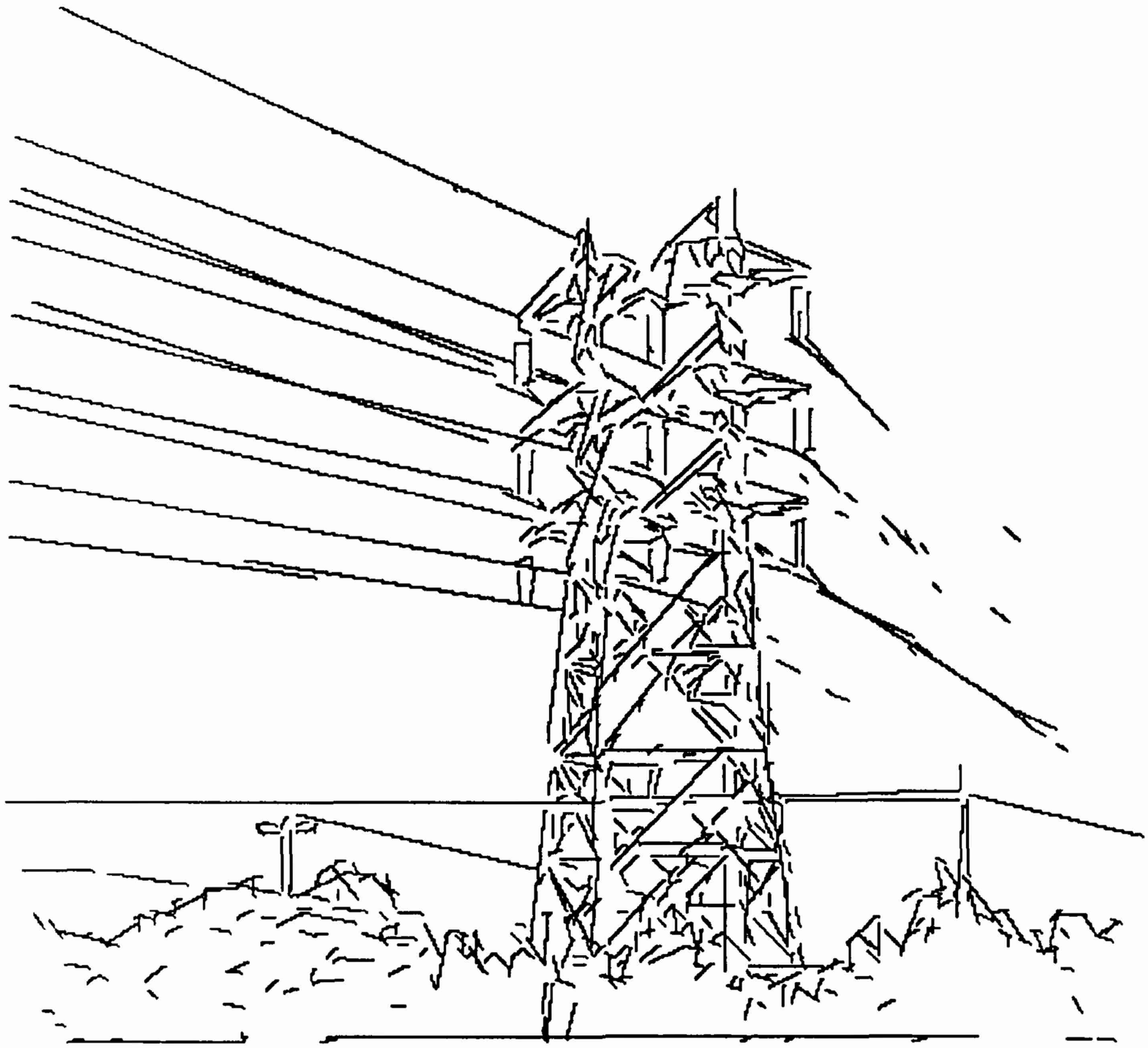


Figure 8: Line segments after the LS extraction for the tower image.

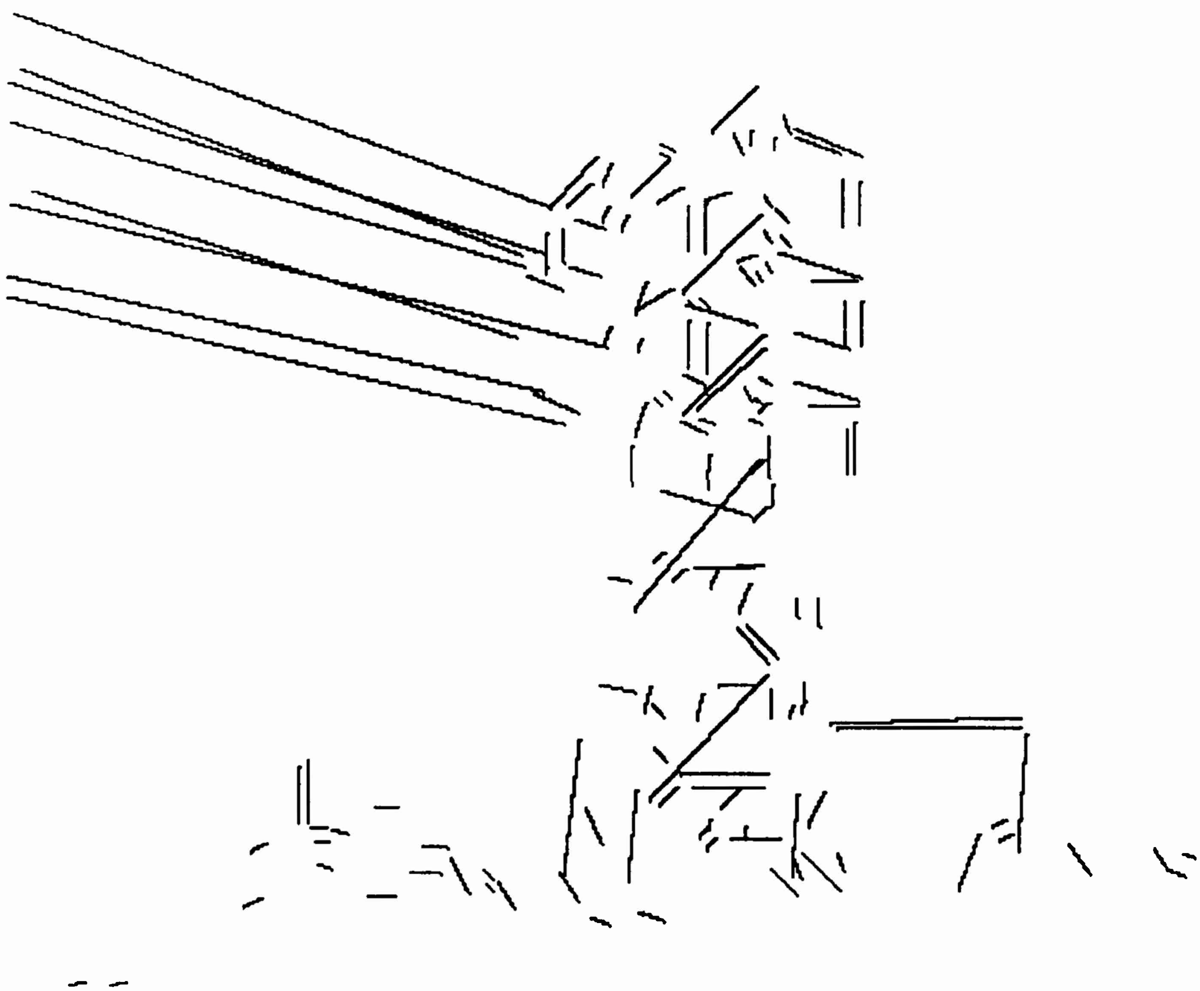


Figure 9: Parallel lines of the tower image.

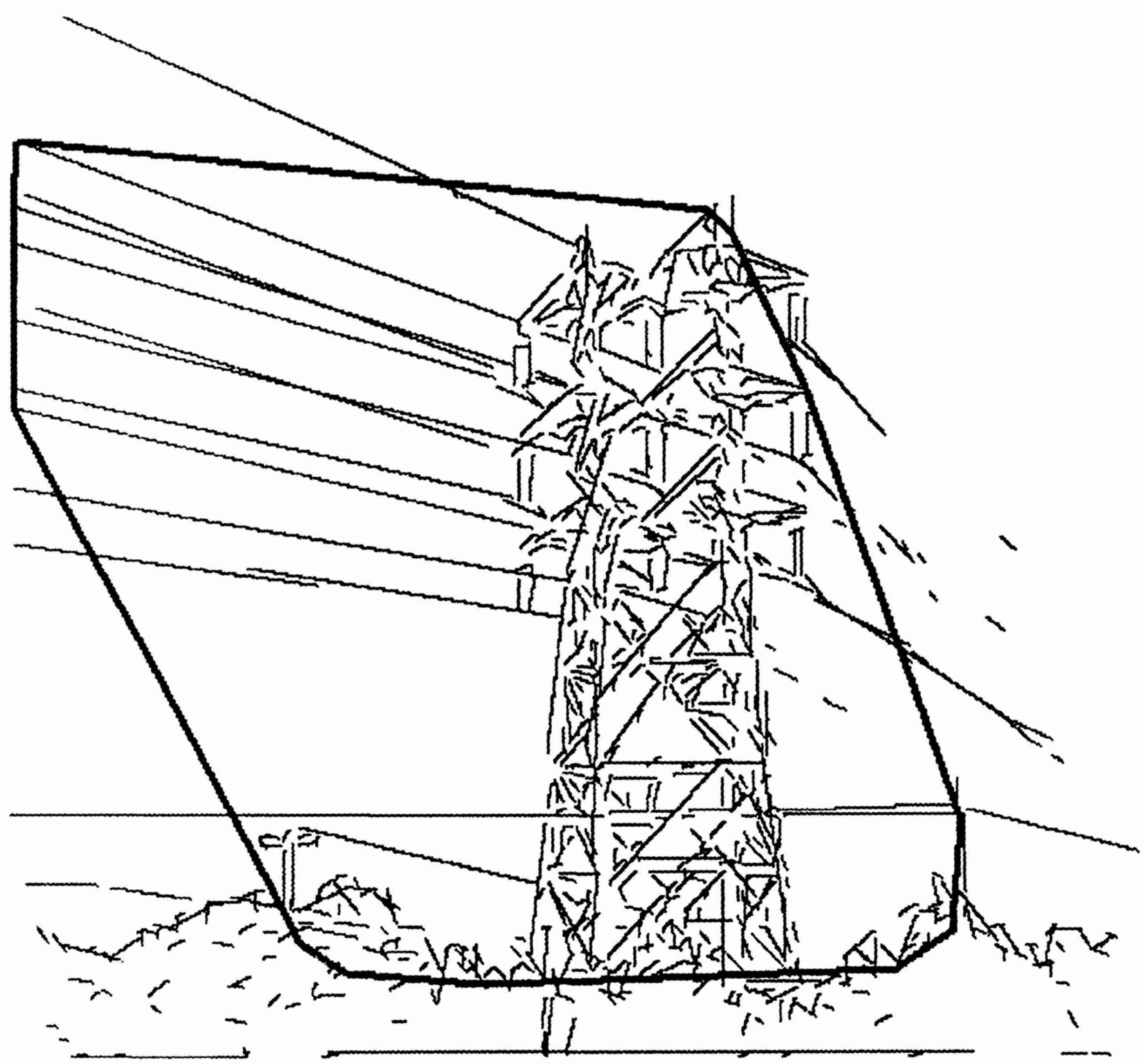


Figure 10: The region of interest for the tower image.

Figure 11 shows the second image, which contains a concrete bridge, trees, and a river. The result of locating the region of interest is shown in Figure 12. Lines in Figure 12 are the output of the LS Extraction. The polygon with the bold outline is the ROI. From Figures 11 and 12, we can see that most part of the bridge is enclosed inside the region of interest except for the bridge's far end near the image boundary. Trees and the river are properly excluded from the region of interest.

The third image, Figure 13, contains a tank. (This image was obtained from General Dynamics.) In this image, the tank is surrounded by complex background and foreground. Figure 14 shows the located region of interest. From Figure 14, we find that most of the tank is included inside the region of interest. The tail part of the tank is not included, since no parallel primitive structures are extracted from that part. Some foreground is improperly included in the ROI because of the parallel lines extracted in the foreground.

We have shown the significance and effectiveness of this approach through the above examples. These examples consist of different man-made objects in natural scenes. The approach found the ROI in each image, and these regions properly enclosed the man-made objects in the scenes.

# 6 Conclusion

This report presents a new approach for the detection of large man-made objects in a rural area using a single monochrome image. The research is a new experiment investigating how minimal knowledge and information about the domain can best be used for the vision task. Prominent features discriminating man-made objects from natural objects are identified. We propose a computational framework applying perceptual organization and using the prominent features to locate a region of interest, which is likely to enclose man-made objects, in a natural scene. Several techniques are developed to group low level image features

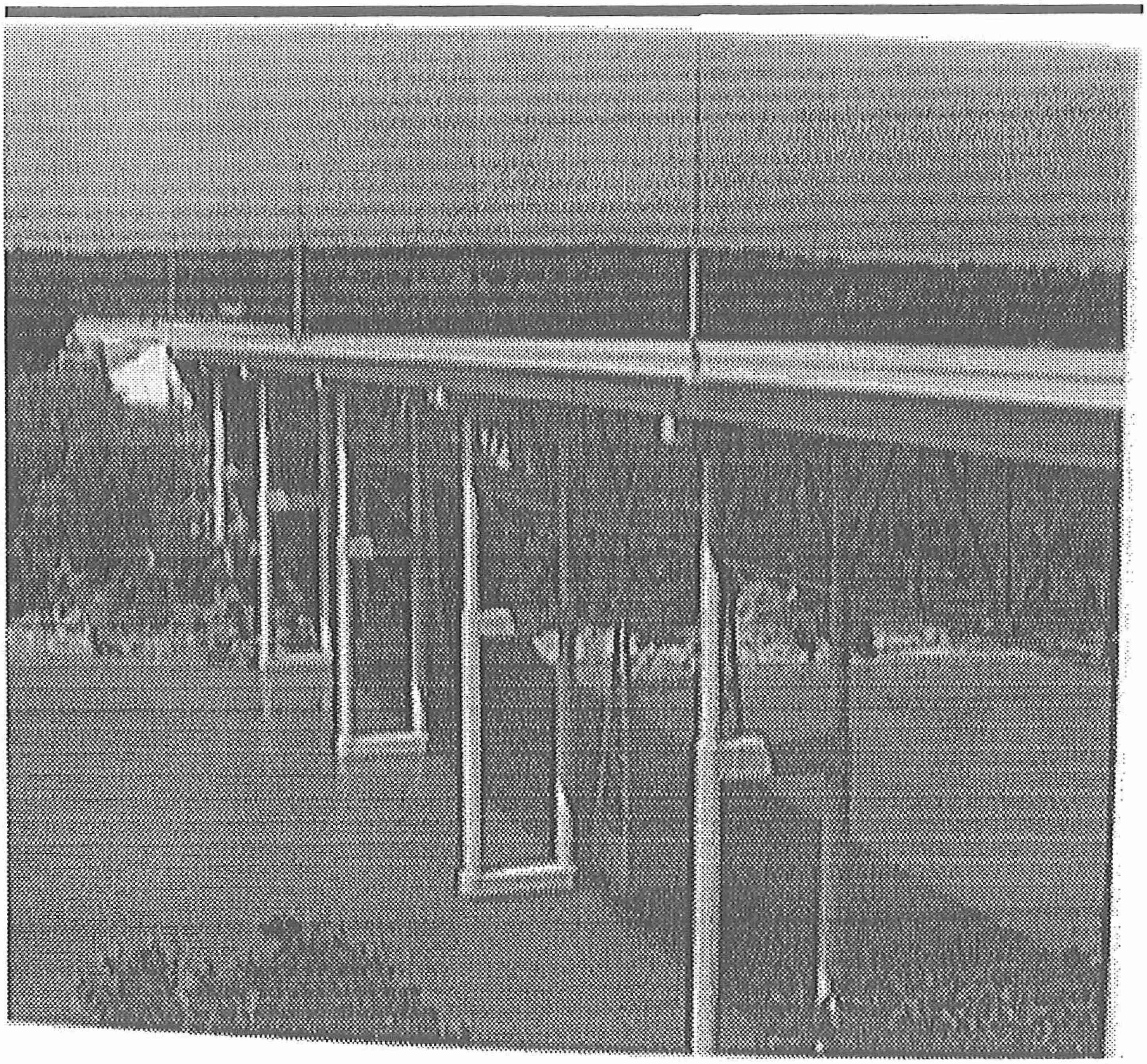


Figure 11: An image with a bridge.

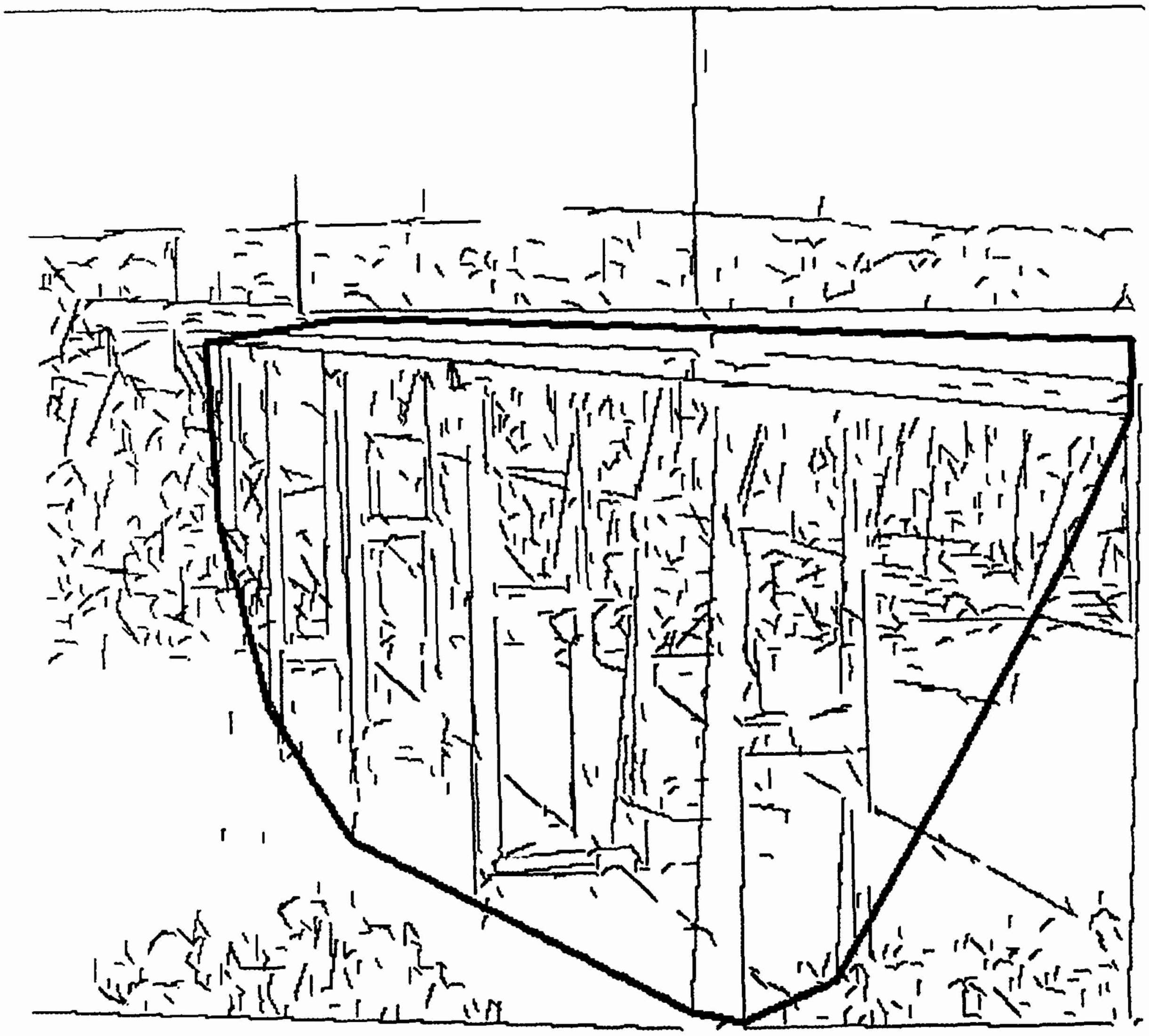


Figure 12: The region of Interest for the bridge image.

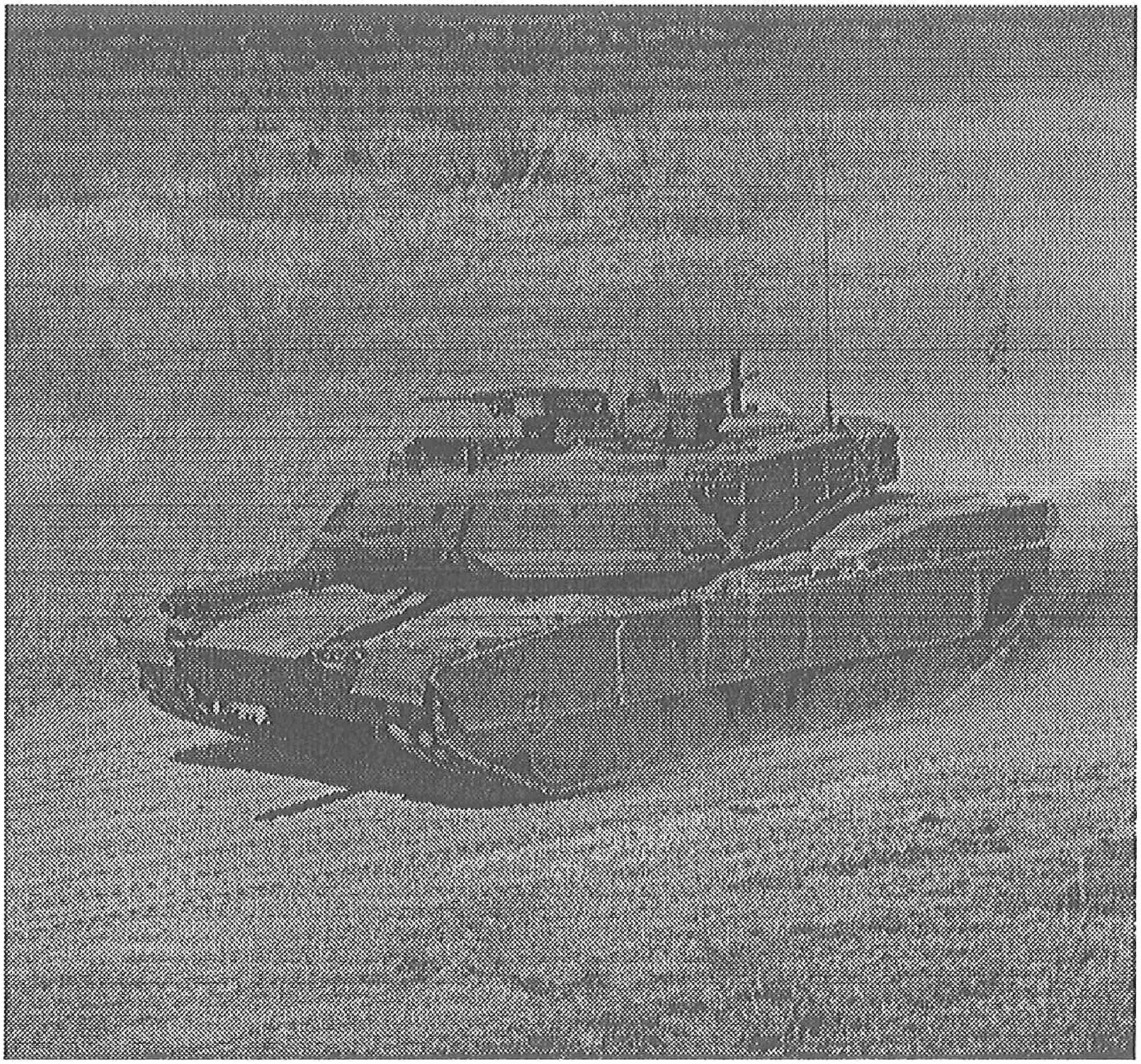


Figure 13: An image with a tank.

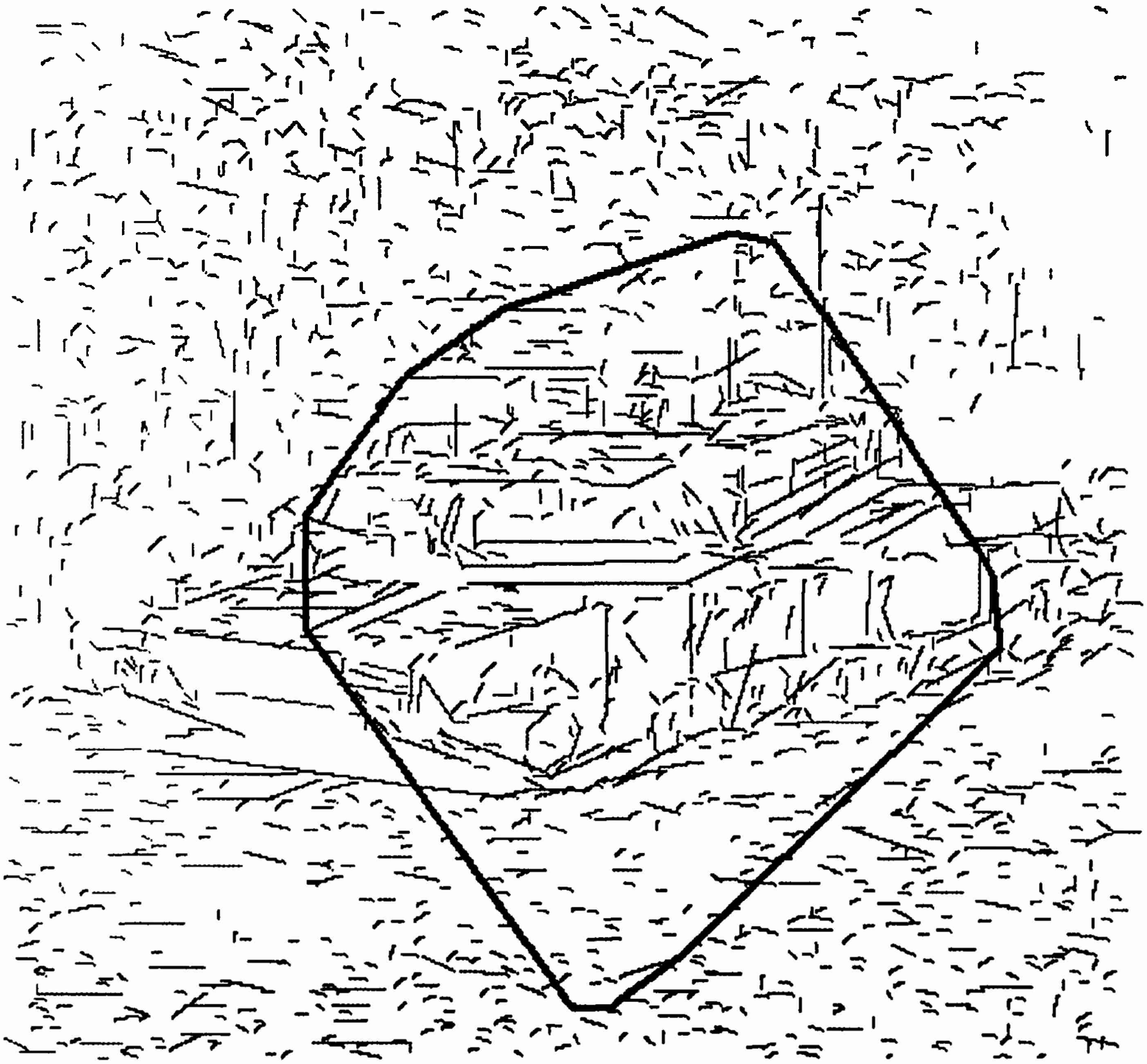


Figure 14: The region of Interest for the tank image.

hierarchically into the ROI. The linear structure extraction method performs line folding as well as line extension simultaneously and, hence, is different from existent collinearization methods. A technique is proposed to group line segments into parallel primitive structures. This technique considers more general situations for grouping parallel lines than most of the previous relevant work. A method of grouping PSs is presented. The PSs more likely to be related are grouped and those that may be caused by accidental image relations are eliminated. Each of these methods presents its own unique property and advantage compared with previous related work. Various examples, including different kinds of man-made objects and complex background, are illustrated to show the approach's effectiveness.

As the examples show, the presented approach is capable of locating a useful region of interest in complex real images. The extracted regions of interest properly enclose the man-made objects in the images. Hence, the search space is reduced from the whole image to the ROI. The ROI hypothesizes man-made objects. Further analysis of the ROI may lead to the identification of an object or the rejection of the hypothesis. Therefore, this technique of locating the ROI can be used for the initial screening of a large environment or a large number of images for automatic object recognition or for a human-machine system. For an automatic system, when specific object classes are given and models are established, primitive structures composing the ROI can be matched to object models instead of matching individual features. This will considerably reduces the search space for matching, since more constraints are applied. For a human-machine system, the ROI can be used as a focus-of-attention for human operators to further examine the image.

We have currently used only parallel primitive structures in the grouping process. Hence, man-made objects without parallel line structures are unable to be detected. Therefore, other types of primitive structures, such as arches, polygons, and junctions, should be considered. Verifying the existence of man-made objects in the isolated region of interest will also be investigated. Through this research, we hope to develop an image understanding method for grouping image events into meaningful structures that represent the objects in the scene.


## Acknowledgment

The authors want to give special thanks to Mr. Chen-Chau Chu at the Computer and Vision Research Center, The University of Texas at Austin, for his valuable suggestions and helpful discussions in preparing this manuscript.


# References


[1] R. T. Chin and C. R. Dyer, "Model-Based recognition in Robot Vision," *ACM Computing Surveys*, Vol. 18, No. 1, March, 1986, pp. 67-108.

[2] M. A. Fischler and T. M. Strat, "Recognizing Objects in a Natural Environment: A Contextual Vision System (CVS)," *The Proceedings of DARPA Image Understanding Workshop,* Palo Alto, California, May 1989, pp. 774-796.

[3] D. C. Baker, S. S. Hwang, and J. K. Aggarwal, "Detection and Segmentation of Man-Made Objects in Outdoor Scenes: Concrete Bridges," *Journal of the Optical Society of America*A, Vol. 6, No. 6, June 1989, pp. 938-950.

[4] D. C. Baker, *Incremental Segmentation for Interpretation of Man-Made Objects in Color Images of Non-Urban Scenes: Concrete Bridges*, Ph.D. Dissertation, The University of Texas at Austin, May, 1989.

[5] R. A. Brooks, "Model-Based Three-Dimensional Interpretations of Two-Dimensional Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 5, No. 2, March 1983, pp. 140-150.

[6] C. C. Chu and J. K. Aggarwal, "Interpretation of Laser Radar Images by a Knowledge-Based System," in press, accepted by *Machine Vision and Applications.*

[7] C. C. Chu, N. Nandhakumar, and J. K. Aggarwal, "Image Segmentation Using Laser Radar Data," *Pattern Recognition*, Vol. 23, No. 6, 1990, pp. 569-581.

[8] R. Mohan and R. Nevatia, "Using Perceptual Organization to Extract 3-D Structures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-11, no. 11, November 1990, pp. 1121-1139.

[9] R. Nevatia and K. R. Babu, "Linear Feature Extraction and Description," *Computer Graphics and Image Processing*, Vol. 13, 1980, pp. 257-269.

[10] P. Fua and A. J. Hanson, "Objective Functions for Feature Discrimination: Applications to Semiautomated and Automated Feature Extraction," *The Proceedings of DARPA Image Understanding Workshop,* Palo Alto, California, May 1989, pp. 676-694.

[11] P. Fua and A. Hanson, "Using Generic Geometric Models for Intelligent Shape Extraction," *The Proceedings of DARPA Image Understanding Workshop,* Los Angeles, California, February 1987, pp. 227-233.

[12] P. Fua and A. J. Hanson, "Resegmentation Using Generic Shape: Locating General Cultural Objects," *Pattern Recognition Letters* , Vol. 5, No. 3, March, 1987, pp. 243-252.

[13] D. Kuan, G. Phipps, and A.C. Hsueh, "Autonomous Robotic Vehicle Road Following," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 10, No. 5, September, 1988, pp. 648-658.

[14] C. Thorpe, M. H. Hebert, T. Kanade, and S. A. Shafer, "Vision and Navigation for the Carnegie-Mellon Navlab," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 10, No. 3, May, 1988, pp. 362-373.

[15] R. Kumar, "Determination of Camera Location and Orientation," *The Proceedings of DARPA Image Understanding Workshop,* Palo Alto, California, May 1989, pp. 870-881.

[16] R. Talluri and J. K. Aggarwal, "Positional Estimation for a Mobile Robot in an Outdoor Environment," *IEEE Workshop on Intelligent Robots and Systems'*, Tsuchiura, Ibaraki, Japan, July, 1990, pp. 159-166.

[17] D. G. Lowe, *Perceptual Organization and Visual Recognition*, Kluwer Academic Publishers, Hingham, MA, 1985.

[18] J. D. McCafferty, *Human and Machine Vision - Computing Perceptual Organisation*, Ellis Horwood Limited, West Sussex, England, 1990.

[19] D. W. Jacobs, "GROPER: A Grouping Based Recognition System for Two Dimensional Objects," *IEEE Computer Society Workshop on Computer Vision,* Miami Beach, Florida, 1987, pp. 164-169.

[20] R. Mohan and R. Nevatia, "Perceptual Organization for Segmentation and Description," *The Proceedings of DARPA Image Understanding Workshop,* Palo Alto, California, May 1989, pp. 415-424.

[21] R. Horaud, F. Veillon, and T. Skordas, "Finding Geometric and Relational Structure in an Image," *First European Conference on Computer Vision,* Autibes, France, April, 1990, pp. 374-384.

[22] J. Dolan and R. Weiss, "Perceptual Grouping of Curved Lines," *The Proceedings of DARPA Image Understanding Workshop,* Palo Alto, California, May 1989, pp. 1135-1145.

[23] G. Reynolds and J. R. Beveridge, "Searching for Geometric Structure in Images of Natural Scenes," *The Proceedings of DARPA Image Understanding Workshop,* Los Angeles, California, February 1987, pp. 257-271.

[24] M. Wertheimer, "Untersuchungen zur Lehe von der Gestalt II," *Psychol. Forsch.* , Vol. 4, 1923. Translated as "Principles of Perceptual Organization," in *Readings in Perception*, David Beardslee and Michael Wertheimer, Eds. Princenton, N. J. , 1958, pp. 115-135.

[25] A. P. Witkin and J. M. Tenenbaum, "On Perceptual Organization," *From Pixels to Predicates*, Edited by A. P. Pentland, Ablex Publishing Corporation, Norwood, NJ, 1986, pp. 149-169.

[26] J. Princen, J. Illingworth, and J. Kittler, "A Hierarchical Approach to Line Extraction," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* San Diego, California, June 4-8, 1989, pp. 92-97.

[27] R. Weiss and M. Boldt, "Geometric Grouping Applied to Straight Lines," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* Miami Beach, Florida, June 22-26, 1986, pp. 489-495.

[28] P. Fua and A. J. Hanson, "Objective Functions for Feature Discrimination: Theory," *The Proceedings of DARPA Image Understanding Workshop,* Palo Alto, California, May 1989, pp. 443-460.

[29] J. R. Beveridge, R. Weiss, and E. M. Riseman, "Optimization of 2-Dimensional Model Matching," *The Proceedings of DARPA Image Understanding Workshop,* Palo Alto, California, May 1989, pp. 815-830.

[30] J. B. Burns, A. R. Hanson, and D. M. Riseman, "Extracting Straight Lines," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 8, No. 4, July, 1986, pp. 425-455.

[31] D. H. Ballard and C. M. Brown, *Computer Vision*, Prentice-Hall, Inc. Englewood Cliffs, NJ, 1982.

[32] D. Marr, "Early Processing of Visual Information," *Philosophical Transactions of the Royal Society of London*, Series B, Vol. 275, 1976, pp. 483-524.

[33] F. P. Preparata and M. I. Shamos, *Computational Geometry: An Introduction*, Springer-Verlag, New York, 1985.

[34] T. Pavlidis, *Algorithms for Graphics and Image Processing*, Computer Science Press, Rockville, MD, 1982.