SEARCHING FOR GEOMETRIC STRUCTURE IN IMAGES OF NATURAL SCENES

George Reynolds
J. Ross Beveridge

COINS Technical Report 87-03

January 1987

Abstract

In this paper we examine the problem of grouping tokens extracted from images of natural scenes into geometrically significant components useful for image interpretation. We propose an algorithm which hypothesizes groups based on the Gestalt laws of perceptual organization, and uses a notion of simplicity in order to resolve conflicting hypotheses. Initially we are examining the problem of grouping straight line segments into larger geometric structures using the geometric relations of collinearity, parallelness, relative angle and spatial proximity. Lines may be viewed as the nodes of a graph and the geometric relations between lines as links in this graph. Significant geometric structures then arise as connected components which our results show bear a close relation to interesting scene events.

This work has been supported by the following grants: DARPA N00014-82-K-0484 and DMA 800-85-C-0012.

1. Introduction

Interpreting an image involves many processes of description and explanation which transform the original intensity array into a form appropriate for the goals of the system (see [10,16,28,15]). Any interpretation system must deal with two central issues: (1) How are the semantics of the scene to be defined in terms of the description and explanation processes?, and (2), How is the enormous search space, which contains the projection of the scene events of interest, to be pruned to managable size? In this paper we discuss a class of algorithms for grouping collections of primitive image events, herein called "tokens", into *geometrically* significant components useful for image interpretation.

Crucial to the process of constructing an image interpretation is the generation of intermediate level tokens which *reduces* the amount of data which needs to be processed, hence cutting the search space, while *increasing* the amount of information carried with each token and providing the semantic primitives for the interpretation processes. These tokens fit into a variety of *abstraction hierarchies*, such as spatial resolution, part - whole and, important for this paper, image geometric - scene semantic.

The relationship between geometric events in the image and corresponding events in the scene has attracted the attention of the image understanding community for many years. Some work has been done recently in the context of aerial photographs, see for example [5,24,13,17]. The work done in the "blocks world" domain (see [8,26,23]) includes many analyses of the relation between image geometry and scene semantics. It is interesting to note that many of the "blocks world" algorithms are good examples of the explanation processes which Witkin and Tennanbaum ([28]) assert are important to the process of image interpretation. Events such as T-junctions have only a fixed number of "explanations" within a blocks world scene, and spatially propagating these events and taking advantage of the mutual constraints between pairs of events results in a single "explanation" or 3-D structure consistent with the constraints.

One reason that this work did not generalize to natural scenes is that the events which occur in natural scenes could not be represented in terms of the primitives which the blocks world algorithms assumed as input. The work presented in this paper builds the structures which, for images of natural scenes, *require explanation* in terms of scene events.

In the context of natural scenes Hanson and Riseman ([10]) proposed a hierarchical decomposition of declarative knowledge (see Figure 1). This decomposition involved placing explicit representions of geometric events in the image at different levels of the hierarchy and linked them via a knowledge base to scene events. The bottom three levels of this hierarchy involve geometric relations in the image The algorithms presented in this paper can be viewed as manipulating the representations

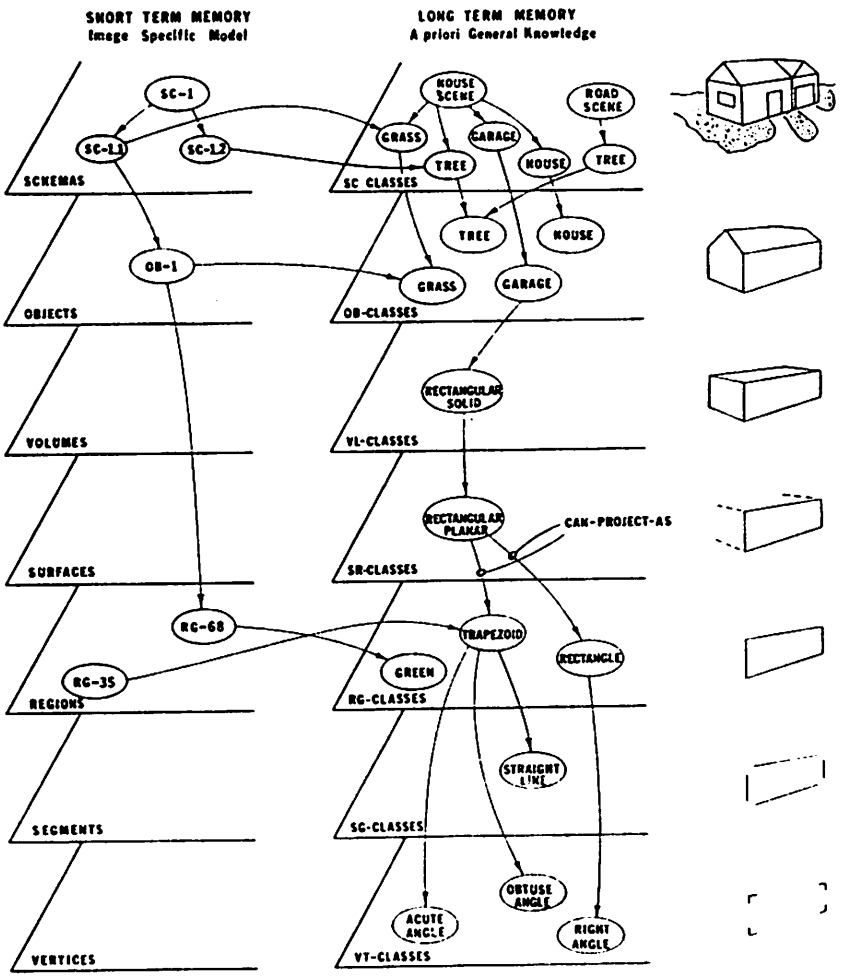

Figure 1: Hierarchical decomposition of geometric knowledge. From Hanson and Riseman 1978.

residing at these levels.

Of course natural scenes are rich in geometric structure, and we present here a methodolgy for constructing tokens whose features and descriptions provide the necessary cues for infering the scene structure. We will argue that this requires complex hypothesis generation and resolution strategies at the image description stage of the processing (see also [27]). The initial results reported on in this paper deal with the problem of grouping line collections of the type typically seen in road scenes and aerial photographs.

Generally scenes of this nature will require that the geometric relations between lines and regions be combined in different ways in different parts of the image, and a general system will require a family of grouping strategies guided by a knowledge-directed or schema-based interpretation system (see [25,7]). We present some preliminary results on natural outdoor scenes and aerial photographs in which lines are grouped based rectilinearity. The results demonstrate that our system can successfully extract geometric structures which closely correspond to important semantic features in these domains.

Figure 2: Natural Scene, an Aerial Photograph.

Let us review briefly the kind of primitives which we can assume our system will have as input. In Figure 2 we see an aerial view of and urban area and in Figure 3 we see a road scene.

Figure 4 and Figure 5 show the results of a local histogram based segmentation algorithm on the images in Figure 2 and Figure 3 respectively (see [3]). Here the input is some collection of pixels and the output is a collection of regions each of which is homogeneous with respect to some (possibly complex) feature. Figure 6 and Figure 7 show the results of a straight line extraction algorithm (see [6]) where the output lines are formed from a set of pixels of approximately uniform gradient direction. These results, although in some ways quite good, are typical of the output of complex low level algorithms. These algorithms often produce fragmented and in some cases (from the point of view of the human obeserver) incorrectly located region boundaries and straight lines.

Moreover, neither of these image abstractions alone contains the information needed to capture the semantic richness of the scene. This is true for two reasons. First, in the case of the region segmentation algorithm, although it has made explicit some semantically important regions of the image, important geometric structure is only implicit in the boundaries of the regions. In the case of the line data, some of the important geometic structure has been made explicit, but more complex

Figure 3: Natural Scene, a Rural Road.

Figure 4: Region segmentation of the aerial photograph.

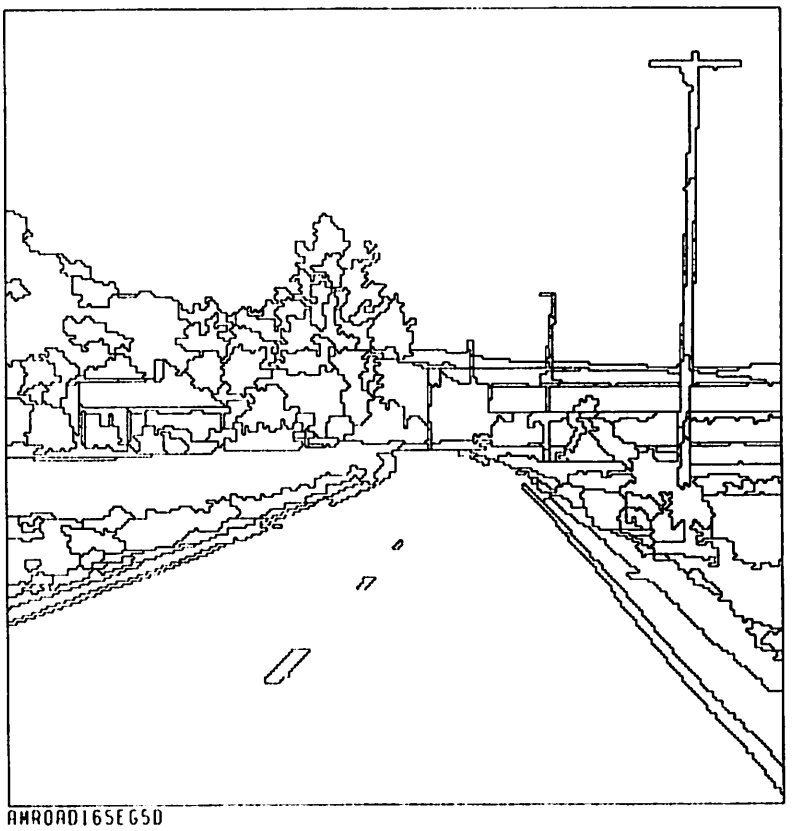

Figure 5: Region segmentation of the road scene.

geometric structure such as closed regions and collections of parallel lines though immediately obvious to the human observer, are again only implicit in the lines. Secondly, regions and lines are simply not the primitives with which to represent many of the events of the scene. In the case of the road scene, the road, barn, trees and other structures are composed of groupings of these (and other) primitives. Therefore the primitves seen in Figure 4 and Figure 6 for the aerial photograph, and Figure 5 and Figure 7 for the road scene, need to be fused and manipulated in such a way that the primitives required to represent 'objects' in the scene are explicit in the resulting structres.

What is needed then are algorithms which group tokens of each type of segmentation separately and simultaneously in order to generate a more complete set of image based tokens with which to represent the image and scene events. Specifically we are interested in *geometric segmentation and grouping algorithms* where the input is some collection of tokens (regions,lines...), and the output are collections of tokens satisfying some (possibly complex) geometric relation between the tokens, for example: relations of adjacency, similar orientation, T-junction and arbitrary combinations of these involving many lines and regions.

Figure 6: Straight lines found in the Aerial Photograph.

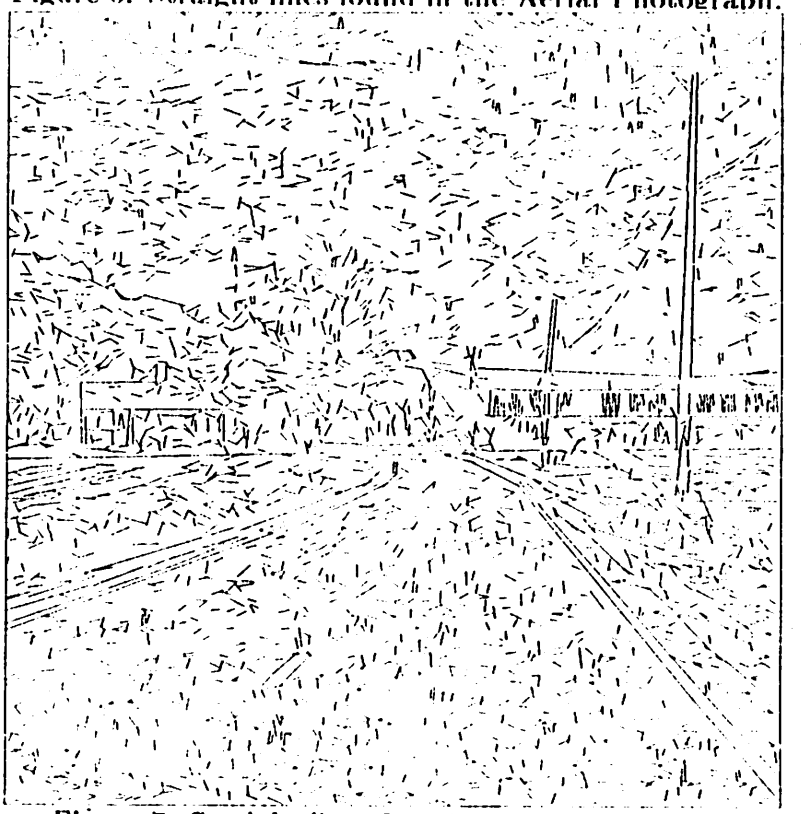

Figure 7: Straight lines found in the Road Scene.

2. Generating and Resolving Multiple Hypothesis

One difficuty in formulating any grouping algorithm or strategy is that a single token may have many possible interpretations, and in order to account for its occurence in the image, it is sometimes necessary to observe the token in more than one larger context. Thus, it is necessary to generate hypotheses which suggest possible explanations and then determine which of those hypotheses (or contexts) serves as the "best" explanation. The criteria by which such a determination is made will depend on the domain, top-down information available at the time such a determination is being made, bottom up information in the local area about the line, and various factors involving the "simplicity" (see the next section) of the structure being hypothesized.

Let us start with a simple example. Consider the image contained in Figure 8. If one is presented with a list of the lines bounding the dark region (shown in bold) and asked to organize them, and return a description of that organization, there are many possibilities. Two are shown in Figure 9 and Figure 10. Figure 9 shows one natural organization of the lines using the Gestalt Law of "good continuation" as the primary organizational principle, and suggests an addmittedly incomplete description of Figure 8 such as: "a dark figure consisting of a diamond and a rectangle, 'transparently' overlapping".

Figure 10 shows another natural organization using the Gestalt Law of "closure" as the dominant organizational principle and suggests another also incomplete description of Figure 8 such as: "a dark figure composed of two pentagonal figures and two triangles with a hexagonal hole in the middle".

From a psychological point of view the descriptions suggested in Figure 9 and Figure 10 are very different, and one might be able to determine which of these organizations (or even some other) is more "natural" to a human observer. However from the point of view of computer vision, each of these descriptions, and many others for that matter, are equally reasonable. Indeed there is no reason to prefer one over the other, *unless* there are scene (general or specific) constraints which guide the selection.

Indeed it is exactly such constraints which are at work in the human visual system. The question is, how do we translate these constraints into the computational processes of a computer vision system? In natural scenes the situations in which region and line relations allow for multiple hypotheses of this nature to be generated expand exponentially with the number of tokens. Methods for translating scene knowledge and constraints into constraints on the grouping and hypothesis generation processes are crucial to limiting the number of either top-down or bottom-up hypotheses actually formed. It is natural for the bottom-up processes

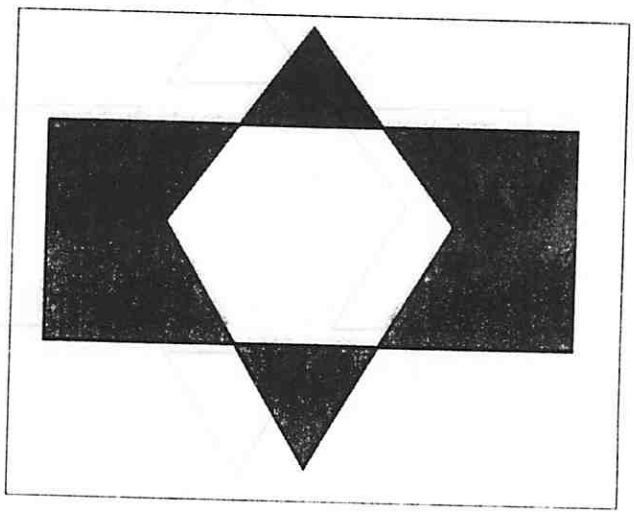

Figure 8: A grouping description problem.

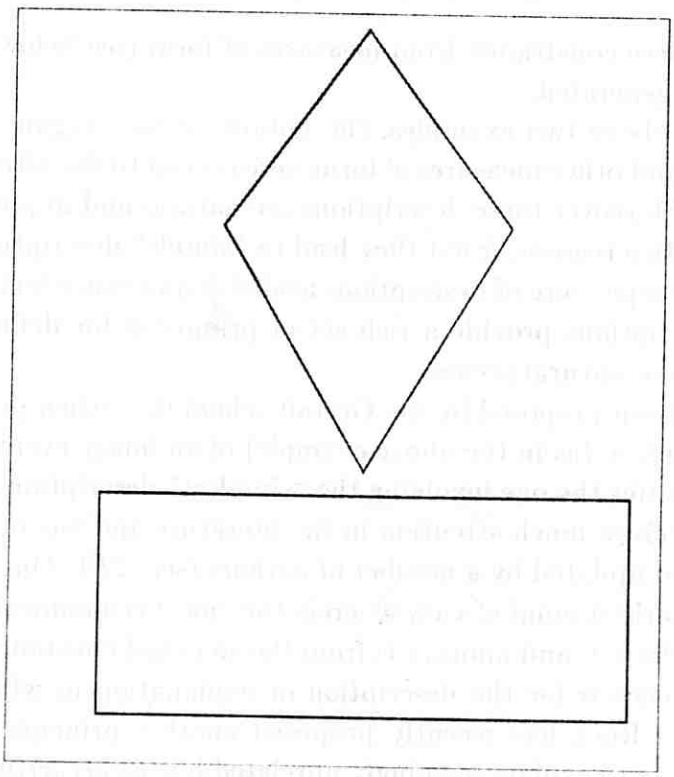

Figure 9: Description 1

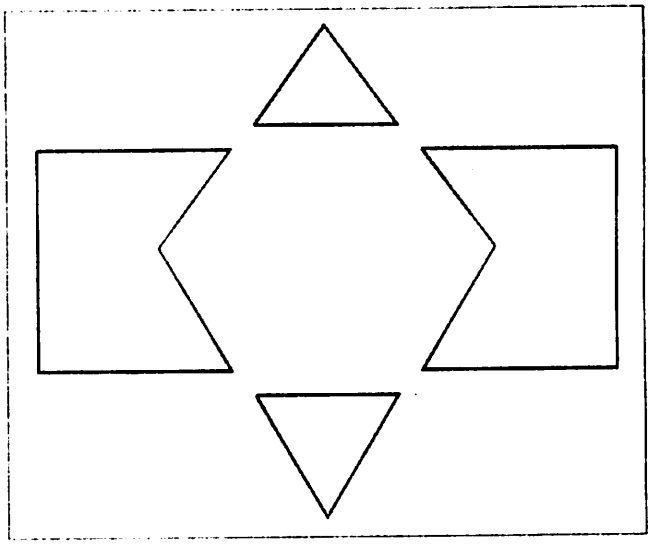

Figure 10: Description 2

to draw these constraints from measures of form (see below) applied to the groups which are generated.

In the above two examples, the notions of line, region, closure, angle, square, rectangle and other measures of form were crucial to the ultimate descriptions which resulted. Moreover these descriptions are natural and important to image interpretation for two reasons. First they lead to "simple" descriptions of the image events, within the repertoire of descriptions available and consistent with the data. Second, these descriptions provide a rich set of primitives for defining a representation of the events in natural scenes.

It has been proposed by the Gestalt school that when presented with more than one description (as in the above example) of an image event, the perceptual system tends to perfer the one involving the "simplest" description. This notion of simplicity has received much attention in the literature and has in more recent years been revised and updated by a number of authors (see [22]). One variation is from the so called empiricist point of view wherein the "most economically encoded" representation is preferred; and another is from the so called constancy point of view wherein the preference is for the description or explanation in which the "object remains constant". Rock has recently proposed another principle wherein "an executive agency seeks to explain seemingly unrelated but *co-occurring* stimulus variations on

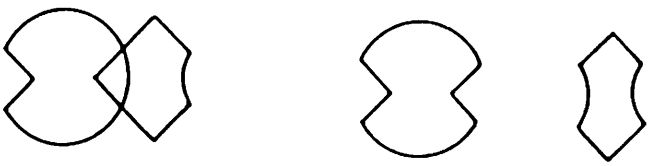

Figure 11: It is difficult to perceive (describe) the figure on the left as composed of two symmetrical figures, even though it seems to be a "simpler" description. After Rock 1983.

the basis of a common cause or in the case of stationary configurations, seeks solutions that explain seeming coincidences and unexplained regularities that otherwise are implicit in the nonpreferred solution" (see |Rock| page 133). For example in Figure 11 we see an example where the "executive agency" prefers the perception (description) which explains the occurence of the collinear lines over the perception (description) which yeilds (the "simpler") two symmetrical objects. Apparently the "executive agency" forces a perception explaining the collinearity over one explaining the symmetry. Perhaps the elements requiring explanation do not even include the two symmetrical objects.

These proposals are interesting not only from the point of view that it helps to explain observed phenomena in human perception, but also that they provide a preliminary framework for constraining the hypothesis generation processes. For example, Julian Hochberg |1| suggested a notion of simplicity or "figural goodness" (in the context of economic encoding) which he proposed explains our perception of line drawings in three dimensions. Consider the following three features of a line drawing:

1. The number of angles enclosed within the figure.

2. the number of different angles divided by the total number of angles.

3. The number of continuous lines.

Hochberg proposed that minimizing a quantitative measure defined in terms of these features yielded the structure most likely to be perceived. Thus we perceive the object to the left in Figure 12 as a three dimensional cube "since" it minimizes Hochberg's measure. Moreover when the cube is viewed from a particular angle (the

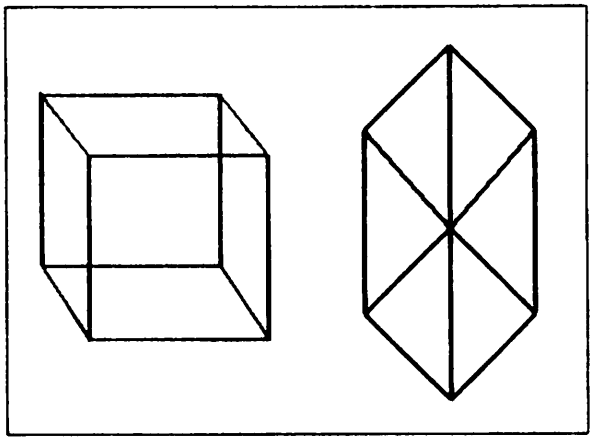

Figure 12: The Necker cube.

object to the right in Figure 12) we no longer perceive the three dimensional structure, "since" there is a simpler description (with respect to Hochberg's measure) which yields a two dimensional perception.

We are *not* interested here in the validity of Hochberg's measure for human perception. We are interested though, in the notion of finding the simplest description for the maximum amount of structure as a possible model for creating bottom-up organizational processes. These processes will require measures of "simplicity" guided by both top down and bottom-up information in order to significantly reduce the search space and build appropriate semantic primitives. In this paper we will develop an algorithm which uses this paradigm for line organization and apply it to line grouping in the context of natural scenes.

3. Requirements for a Geometric Grouping System

There are four essential requirements on any geometric grouping system. First is the representation of primitive tokens (in our case lines) and the geometric relations between tokens (in our case collinearity, parallelness, relative angle, and spatial proximity. Our choice of a representation will be discussed in next section.

Second, it must identify the relationship between the groups of tokens satisfying these primitive relations, and knowledge about the domain being interpreted. That is, one must select primitives which form the basis of a language for representing objects and knowledge about the scene.

Third, it must develop grouping strategies, based on the objects of interest in the scene, knowledge of the domain and the current state of the system. That is, one must build a language of *grouping strategies* for representing procedural knowledge about the objects of interest in the domain. Finally, at all stages of grouping the system must deal explicitly with the problem of search, and its relation to the objects in the domain which are to be hypothesized.

The algorithm presented in this paper is part of a larger computational framework under development at UMASS (see also [27]) for confronting the issues described above. We view the grouping and search processes as part of a 4 stage iterative grouping and extraction strategy which can be summarized as follows:

- *Primitive Structure Generation*: These processes provide the primitives (regions, lines, more complex tokens) which are the input to the grouping and hypothesis generation process described next.

- *Linked Structure Generation*: This step applies very general geometric constraints to obtain graphs within which search processes can be applied to identify specific structures of interest. For example rectilinear structures which would contain rectangles or other simple geometric structures. This is essential for generating search spaces of reasonable size.

- *Subgraph Extraction*: This step involves the extraction of specific structures "one step up" the abstraction hierarchy, and uses the linked structures to constrain the search.

- *Replacement and Iteration*: Having extracted more abstract tokens, these can now play the role of primitives in another pass of grouping and extraction.

The algorithm presented in this paper is at the *Linked Structure Generation* stage of the strategy and is applied here only to straight lines for the purpose of building rectilinear structures. In [27], a similar strategy is applied with striking results for the purpose of extracting straight lines. In general each step of the grouping and extraction strategy must apply constraints which either significantly reduce the search space and/or add important information to the descriptive power of the system.

4. Identifying Rectilinear Structure

In this section we review the system under development for the extraction of rectilinear structure and our approach to limiting the search for both general and specific geometric structure. The process starts with a set of primitives which are the the output of a straight line extraction algorithm, in this case the Burns algorithm was used (see [6]). These lines are viewed as nodes in a graph and the geometric relations of

- spatially proximate collinear
- spatially proximate parallel
- spatially proximate orthogonal
- and any subset of the above,

as relations between the nodes. These specific relations will be defined below.

Hypotheses of line groups are then generated using a connected components algorithm based on the chosen geometric links. These components, which are called *Rectilinear Line Groups*, form a new class of tokens which have emergent features and form a new level of the Geometric-Semantic abstraction hierarchy. Because different choices of the primitive geometric relations yield different rectilinear groups, a given line may participate in more than one group and so we are begining to explore various notions of "simplicity" or "preference" to resolve conflicting or competing hypotheses. Finally, specific geometric structure may be identified (rectangles, collinear and parallel structure) as subgraphs of these connected components. We refer the reader to [7] for examples of the extraction of specific geometric structure corresponding to the projection of scene events.

In defining spatial proximity one must recognize that there are essentially three types of proximity to choose from. First, there is a notion of *image-dependent proximity* wherein the metric of the image itself is used to define the distance between two tokens independent of the size of the tokens. The second is *token-dependent proximity* where the distance between two tokens is a function of the size of the tokens. For example, the length of a line might be a factor in defining a notion of parallel. Finally, there is *scene-dependent proximity* where, for example in a road scene with a camera in standard position, distance between two objects at the bottom of the image might mean something very different than for two objects at the top of the image. For the results presented in this paper we have used *token-dependent proximity* for defining orthogonal, parallel and collinear relations. Clearly this is only a start, and indeed there are many issues that we have only begun to address with regard to the relations between these three types of proximity.

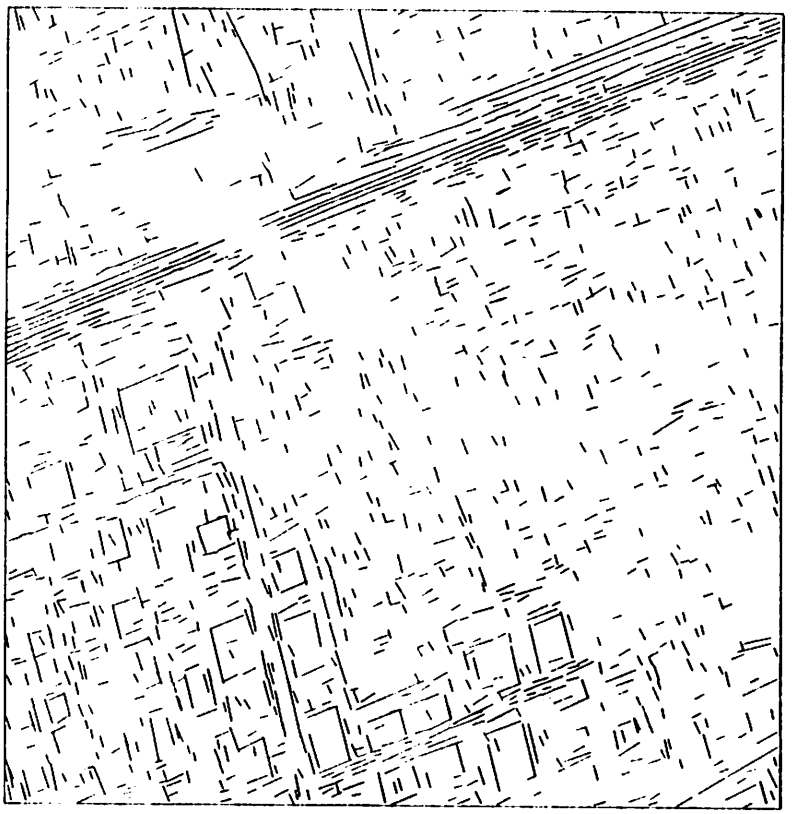

Figure 13: Lines in the Orthogonality Range for the Aerial Scene containing the greatest number of lines.

The process of finding rectilinear line groups involves partitioning the lines into overlapping groups filtered with respect to orientation. These groups, which we call orthogonality ranges, contain all lines within $\pi/16$ radians (11.25 degrees) of some orientation or $\pm\pi/2$ radians of that orientation. The choice of $\pi/16$ radians results in 8 distinct orthogonality ranges. By restricting the connected components algorithm to lines from a single orthogonality range the rectilinearity of the group as a whole is guaranteed. In addition, considering each of the orthogonality ranges independently reduces the search space without limiting the resulting geometric structures and lends a degree of parallelism to the process. In Figure 13 we see the lines in the orthogonality range from Figure 6 containing the the greatest number of lines. In Figure 14 we see the lines from Figure 7 in the orthogonality range about the horizontal orientation.

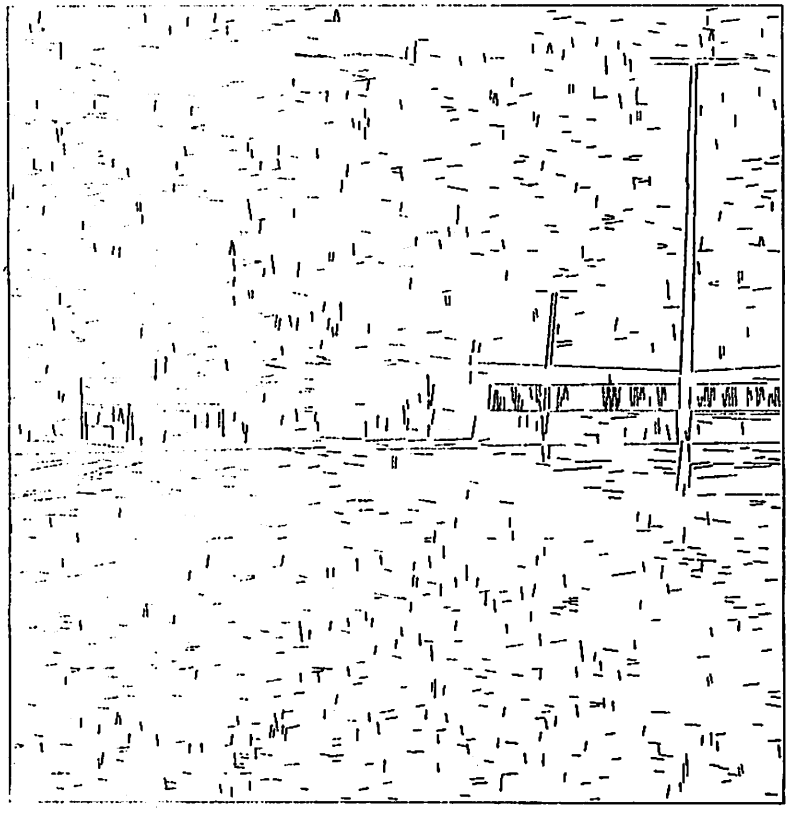

Figure 14: Lines in the Horizontal and Vertical Orthogonality Range for the Road Scene.

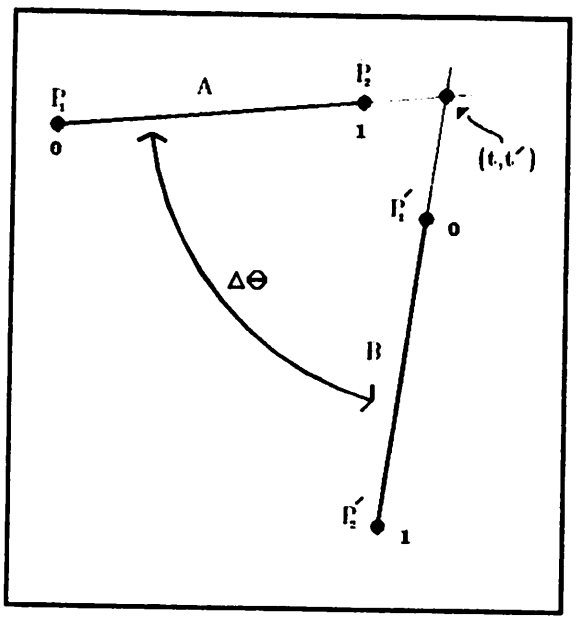

Figure 15: Illustrating Comparison of two lines for Spatially Proximate Orthogonality.

4.1 Spatially Proximate Orthogonal Line Segments

Figure 15 illustrates our use of token dependent proximity to determine whether two line segments **A** and **B** are spatially proximate orthogonal. Three measures contribute to determining spatially proximate orthogonality, $\Delta\theta$, t and t'. If one thinks of line segment **A** as the unit vector of a coordinate system obtained by extending **A** infinitely in both directions, then t is the value where the extension of line segment **B** intersects it. The value t' is analogus to t, except the roles of lines **A** and **B** are reversed. The value $\Delta\theta$ is simply the relative orientation between **A** and **B**.

Definition: Two lines are *spatially proximate orthogonal* if the following conditions are satisfied:

- $-\delta < t < 1 + \delta$,
- $-\delta < t' < 1 + \delta$,
- $\pi/2 - \epsilon < \Delta\theta < \pi/2 + \epsilon$.

The terms ϵ and δ are thresholds which constrain the definition. For smaller ϵ the two lines must be closer to orthogonal to be related. For smaller δ the lines must be spatially closer with resepect to their endpoints. Note this definition makes no distinction between corners and T-junctions. Values of ϵ 0.17 radians (10

degrees) and δ 0.5 were used for the results in this paper and an investigation of this parameter space is currently underway.

4.2 Line Overlap and Displacement

Spatially proximate parallel and spatially proximate collinear are defined in terms of two more primitive relations which we call symmetric lateral displacement (DIS_{sym}) and symmetric overlap (OV_{sym}). The basis for these two measures are more primitive relations shown in Figure 16 where
$DIS(A, B, P_1)$ measures the displacement *from* point P_1 on **A** *to* **B** and $OV(A, B)$ measures the overlap *from* **A** *to* **B**. Symmetric measures DIS_{sym} and OV_{sym} result essentially by taking averages of measures from **A** to **B** and **B** to **A** with these and the same measures with the roles of **A** and **B** reversed. A detailed explanation of how displacement and overlap are calculated can be found in [20].

The resulting symmetric measures satisfy the following conditions

$$0 \quad DIS_{sym}(A, B) \quad \infty$$

$$\infty \quad OV_{sym}(A, B) \quad 1.$$

These measures are intended for use between lines already known to be roughly of the same orientation. The lateral displacement measure captures the distance in the direction orthogonal to the orientation of the line. The overlap measure captures the distance along the projection of the line. A negative overlap is used to measure the distance between two lines along their projection. For example a line completly overlaps itself ($OV_{sym}(A, A)$ 1) while there is no lateral displacement between a line and itself ($DIS_{sym}(A, A)$ 0). For two parallel sides of a square, $DIS_{sym}(A, B)$ 1 and $OV_{sym}(A, A)$ 1. For two collinear lines lying end to end, $DIS_{sym}(A, A)$ 0 and $OV_{sym}(A, A)$ 0.

4.3 Spatially Proximate Collinear and Parallel

We are now in a position to define spatially proximate collinear using the definitions of displacment and overlap.

Definition: Two lines are *spatially proximate collinear* if the following conditions are satisfied:

- c' $OV_{sym}(A, B)$ c
- $DIS_{sym}(A, B)$ δ
- $|\Delta\theta| \leq \alpha$

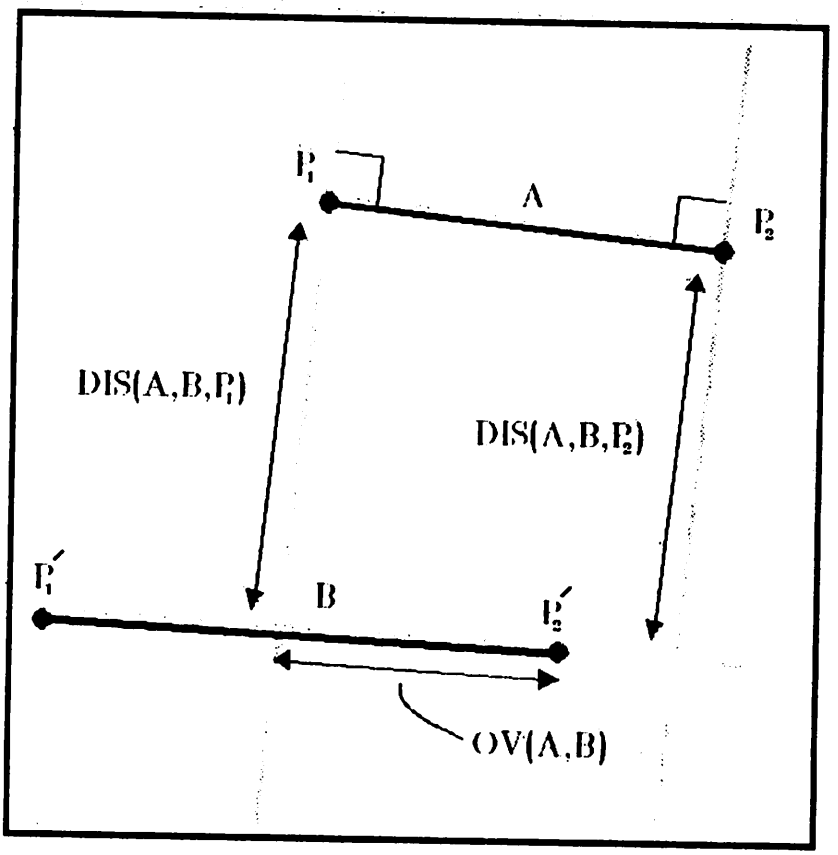

Figure 16: Illustrating Displacement and Overlap Between Two Lines.

For the results shown in this paper we have used the values $\epsilon' = .5$, (separated by at most 50 percent of their average length), $\epsilon = .15$, (having at most 15 percent overlap), $\delta = .15$, (displaced by at most 15 percent of their average length) and $\alpha = 0.17$, (10 degrees).

Similarly we have used displacement and overlap to define spatially proximate parallel.

Definition: Two lines are *spatially proximate parallel* if the following conditions are satisfied:

- $OV_{sym}(A, B) > \epsilon$
- $DIS_{sym}(A, B) < \delta$
- $|\Delta\theta| < \alpha$

For the results shown in this paper we have used the values $\epsilon = .5$, (at most 50 percent overlap), $\delta = .5$, (at most 50 percent of the average length apart), and $\alpha = 0.17$ (10 degrees).

In summary these definitions provide measures of token dependent proximity with respect to the relations of collinear, parallel and orthogonal. There is much work to be done in exploring the parameter spaces employed in these definitions. In Figures 17, 18 and 19 we see the line pairs satisfying the relations of spatially proximate orthogonal, spatially proximate collinear and spatially proximate parallel for the aerial photograph. Figures 20, 21 and 22 show the same relations for the road scene. The pairs shown in these figures are the in input to the connected components phase of the grouping algorithm which we describe in the next section.

4.4 Line Group Generation and Competing Hypothesis Resolution

We are now in a position to describe the creation of rectilinear groups. Their generation really involves two processes:

1. *Connected Components:* which forms an hypothesis of a spatially proximate rectilinear structure.
2. *Voting Processes, Competing Hypothesis Resolution:* which tries to find the simplest explanation for the largest collection of lines.

The connected components algorithm uses any subset of the three relations, spatially proximate orthogonal,spatially proximate collinear and spatially proximate parallel. These different subsets make up a family of possible relations upon which the connected components can be based. For example we might choose to select conditions only from the rules for collinearity and parallelness and ignore

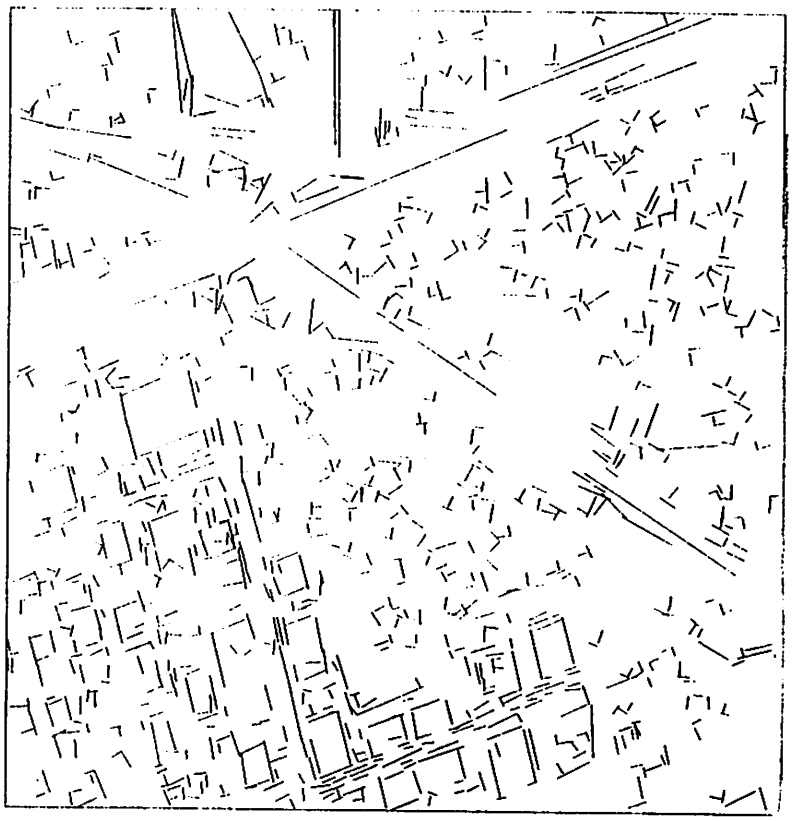

Figure 17: Lines Belonging to Spatially Proximate Orthogonal Pairs for the Aerial Photograph.

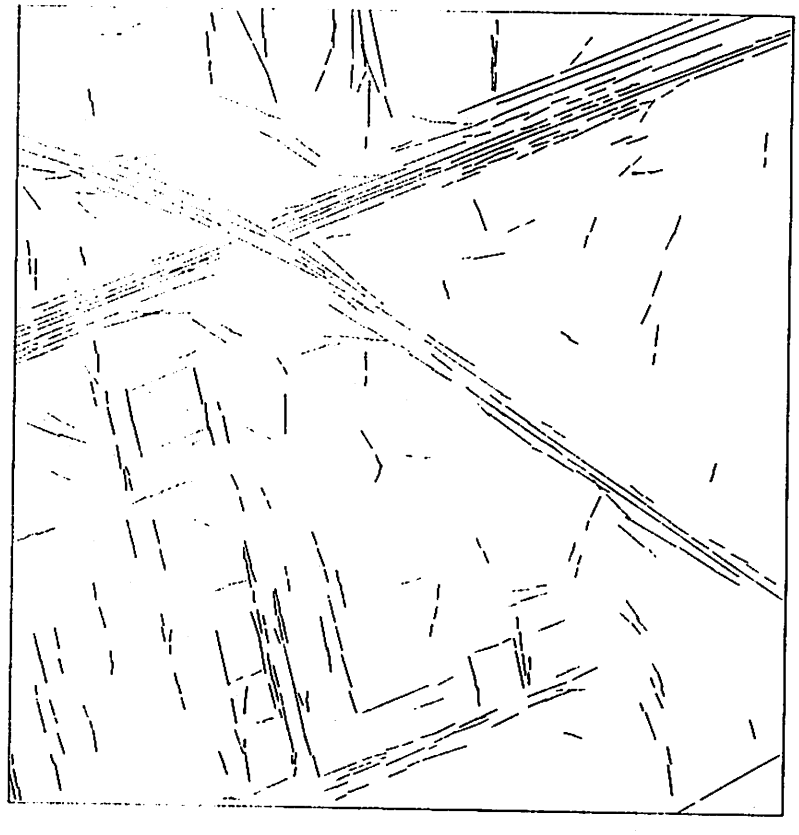

Figure 18: Lines Belonging to Spatially Proximate Collinear Pairs for the Aerial Photograph.

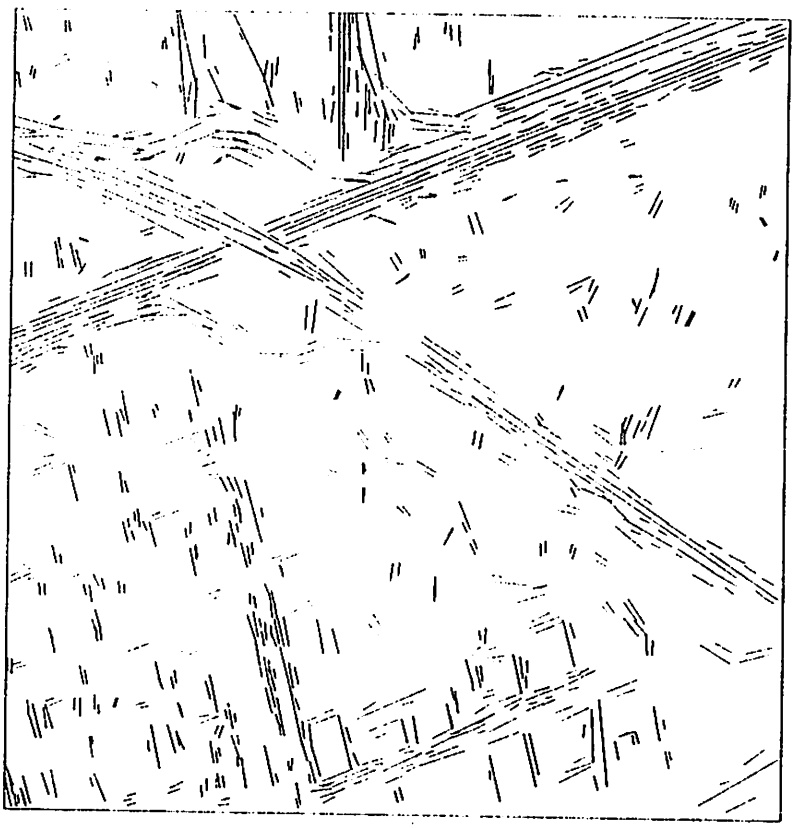

Figure 19: Lines Belonging to Spatially Proximate Parallel Pairs for the Aerial Photograph.

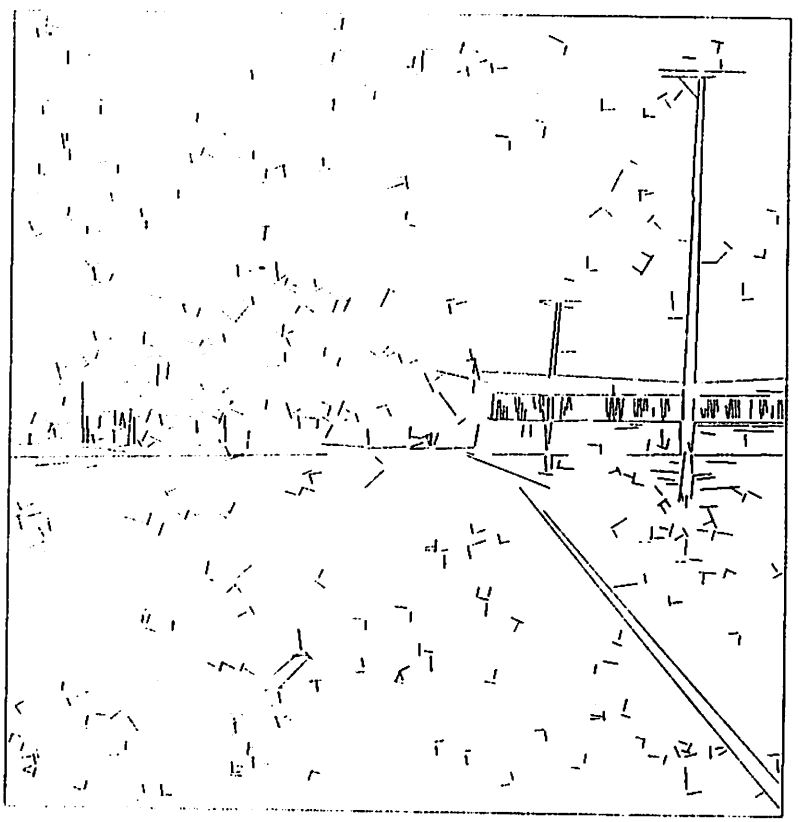

Figure 20: Lines Belonging to Spatially Proximate Orthogonal Pairs for the Road Scene.

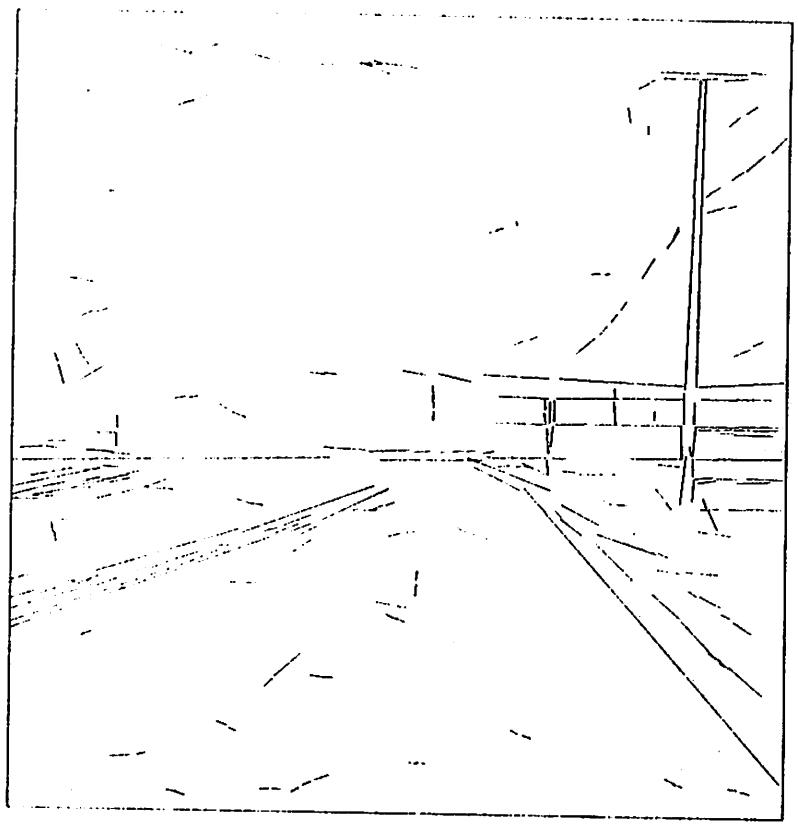

Figure 21: Lines Belonging to Spatially Proximate Collinear Pairs for the Road Scene.

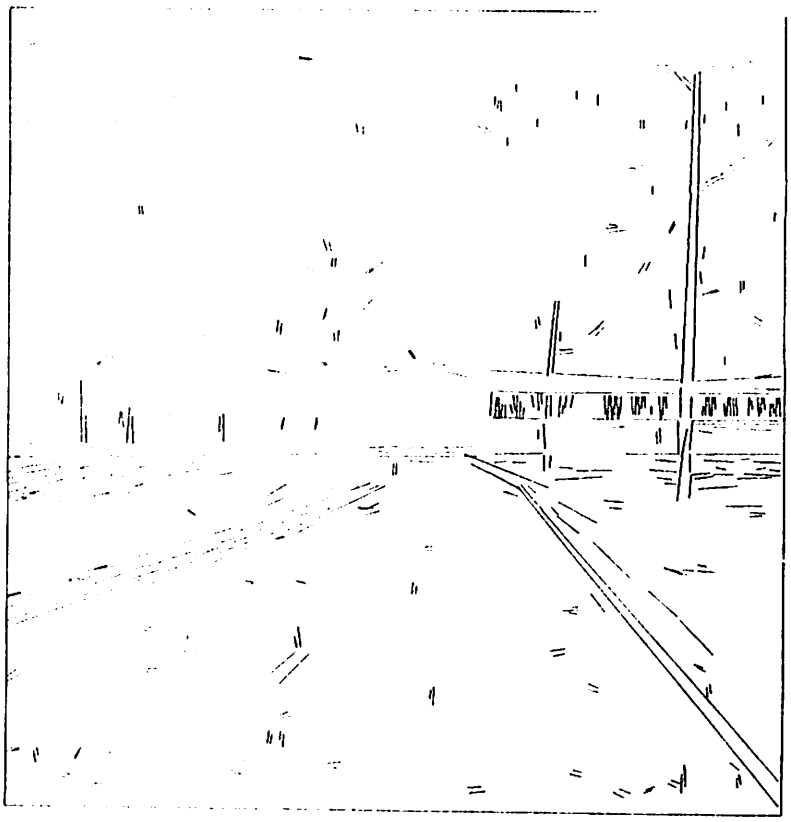

Figure 22: Lines Belonging to Spatially Proximate Parallel Pairs for the Road Scene.

the orthogonality conditions. The output of the connected components algorithm would be structures all of whose elements are parallel or collinear with the property that any line in the group is spatially proximate (in the sense described above) to at least one other line in the group.

Given that the the connected components algorithm is run separately on each of the othogonality ranges and that the orthogonality ranges overlap, each line can belong to two connected components. In addition, groups based on different combinations of links are formed separately and a line can belong to groups of each type independently. Currently we have explored resolution within groups formed using a single combination of links. One means for accomplishing this resolution is to invoke a process which asks of each line which component it prefers according to various selection criteria. For example: Let each line vote for the group which has the most lines, or the group whose total length is longest, or the group which is "simplest" based on some more complex rule. Some preliminary results are shown in the next section.

5. Results

The Rectilinear Line Grouping System, as the system described above is called, has been run on roughly ten images drawn from different domains including road scenes, aerial photographs and house scenes. In each case connected components, or what we will simply call line groups, were formed using all combinations of the three basic relations, spatially proximate orthogonal, spatially proximate collinear and spatially proximate parallel. These line groups clearly represent a next step up the geometric-semantic abstraction hierarchy. The groups we have produced differ greatly in size and form, capturing a wealth of structure in the images. We are just begining the process of learning how best to utilize and characterize these groups. Presented here is a sampling of the groups generated for the images presented in Figures 2 and 3.

5.1 A Sample of the Connected Components Grouping Results

In Figure 23 we see the lines which belong to the group containing the largest number of lines formed using only the relation spatially proximate orthogonal. The source lines (Figure 6) are for the aerial photograph. This group contains 82 lines, the next largest group of the same type for this image contained only 19. An examination of Figure 17 shows that many spatially proximate orthogonal line pairs are identified around the buildings in the lower lefthand portion of the image. The connected components grouping finds a number of groups containing roughly 5 to 15 lines in this area.

The line groups shown in Figure 24 are the result of grouping lines from the same aerial photograph based on the relations spatially proximate collinear and spatially proximate parallel. The figure was generated by selecting the 4 largest groups and displaying 'best' 3. How the best 3 were chosen will be discussed below as an example of resolving multiple hypothesis.

The line groups shown in Figure 25 were produced using all three relations, spatially proximate orthogonal, spatially proximate collinear and spatially proximate parallel. Four groups are shown, selected out of the largest 5. What appears to be a large group resulting from buildings in the lower portion of the scene is actually two groups. The two groups contain 158 and 135 lines respectively and share 54 lines in common. The two groups arose out of the connected components grouping in adjacent orthogonality ranges. It is clear that the significant groups based on rectilinearity correspond to significant scene events.

In Figures 26 and 27 we see rectilinear line groups for the road scene image of Figure 3. The source lines are shown in Figure 7. The two groups shown in Figure 26 are the result of grouping based solely on spatially proximate orthogonality. They

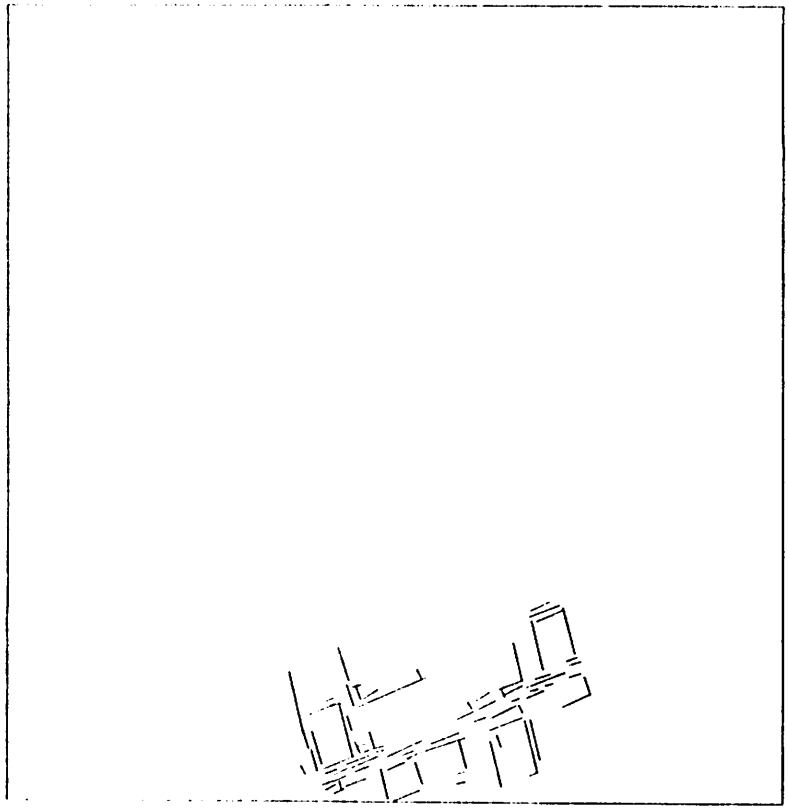

Figure 23: Largest Group Based on Spatially Proximate Orthogonality.

are the 2 'best' groups out of the largest 4. The 2 discarded each contain about half the lines in the largest group, and are the result of grouping the lines for the large group in the orthogonality ranges adjoining the one in which the large group is found. Figure 27 shows the single largest group produced by grouping based upon all three relations.

5.2 Illustrating the Resolution of Multiple Hypothesis

In selecting the groups shown in Figure 24 we said the 'best' three of the four groups were selected. Figure 28 shows the group the was rejected and Figure 29 the group that caused it to be rejected. As mentioned earlier, a simple means of resolving conflicting hypotheses is for each line to vote for the group it 'prefers'. One basis for preference is size. Reliance on size amounts to an elementary definition of 'simplicity' which in this case amounts to the selecting the group accounting for the maximum ammount of structure. Size may be measured in many ways, and two simple ways are in terms of the number of lines or the cumulative length of the lines. The group shown in Figure 28 contains 88 with a cumulative length of 1116. The group shown in Figure 29 contains 117 lines with a cumulative length of 1336. These 2 groups share 83 lines in common. Using either of these measures, the group

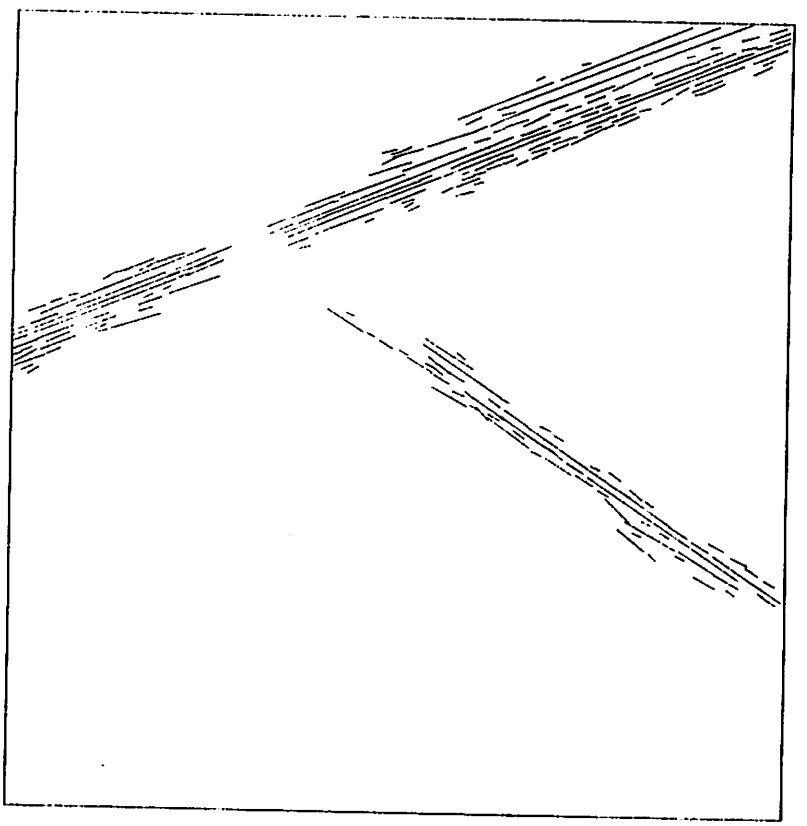

Figure 24: Largest Groups Based on the Relations Spatially Proximate Collinear and Spatially Proximate Parallel.

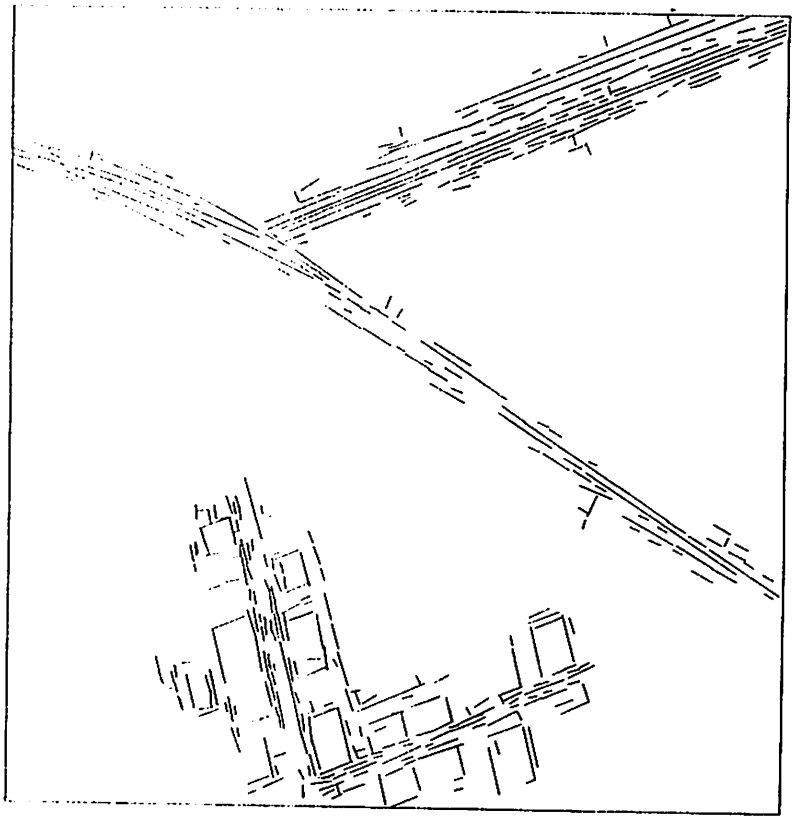

Figure 25: Largest Groups Based on all Three Relations: Spatially Proximate Orthogonal, Spatially Proximate Collinear and Spatially Proximate Parallel.

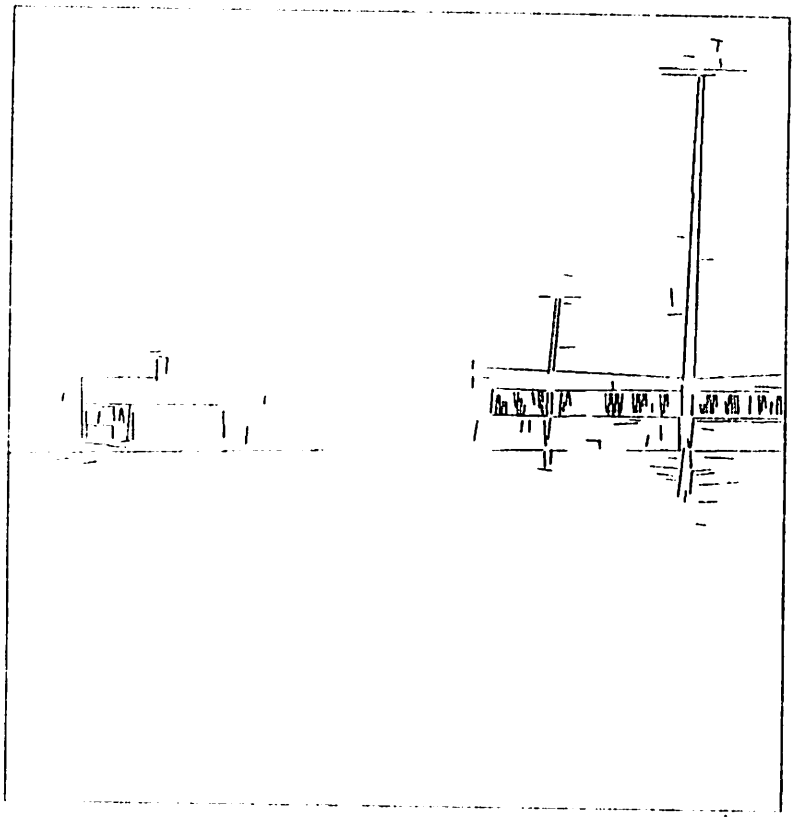

Figure 26: Largest Groups Based on Spatially Proximate Orthogonality.

in Figure 29 would clearly receive the majority of the votes.

In resolving the conflict between the groups in Figures 28 and 29 it made little difference what measure of size was used. This may not always be the case, however as the groups illustrated in Figures 30 and 31 the group in Figure 31 was chosen over the one in Figure 30 for inclusion in Figure 25 by hand Which of these two groups would be considered 'best' depends upon the measure of size. The group in Figure 30 contains 107 lines with a cumulative length of 902. The group in Figure 31 contains 102 lines with a cumulative length of 983. The two groups have 70 lines in common. Hence which group is chosen depends upon the measure used. Based on cumulative length, the group in Figure 31 wins. Based on the number of lines, the group in Figure 30 wins. This example gives some flavor for the types of difficulties surrounding the issue of multiple hypothesis resolution.

6. Conclusion and Future Directions

In this paper we have examined the problem of grouping tokens extracted from images of natural scenes into geometrically significant components useful for image interpretation. An implementation of a system for grouping straight lines into larger rectilinear configruations is described. In designing and implementing this system,

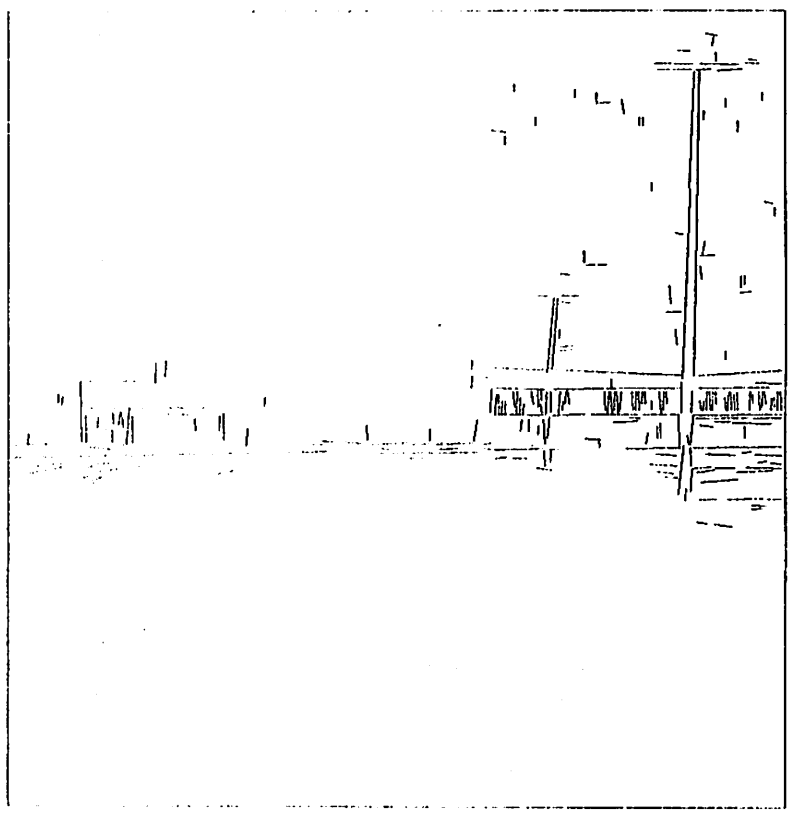

Figure 27: Largest Single Group Based on all Three Relations: Spatially Proximate Orthogonal, Spatially Proximate Collinear and Spatially Proximate Parallel.

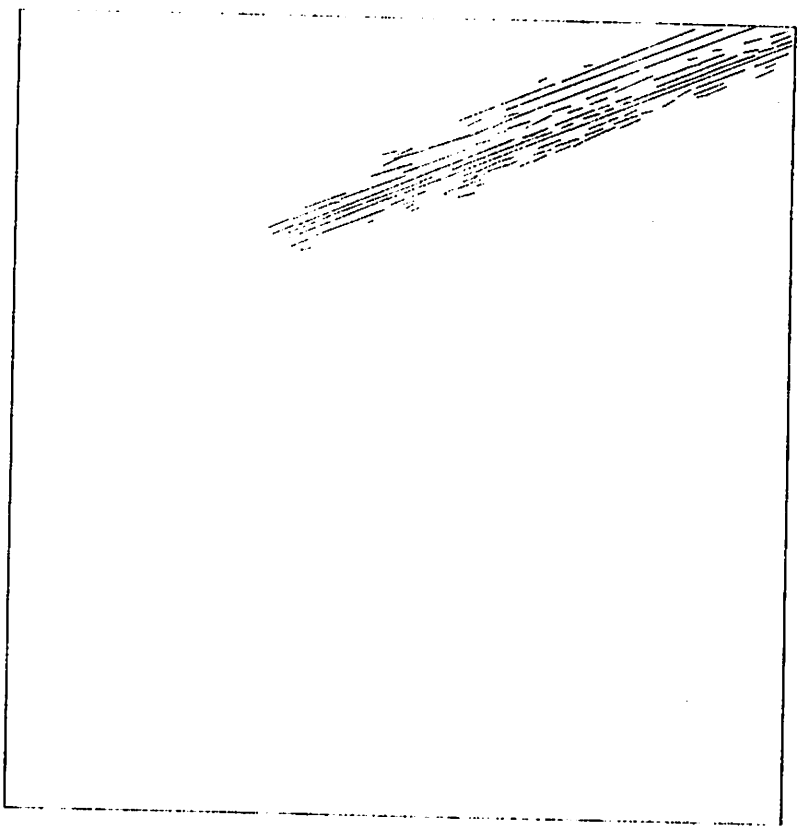

Figure 28: Line Group Based on Spatially Proximate Collinearity and Parallelness, rejected For Lack of Support.

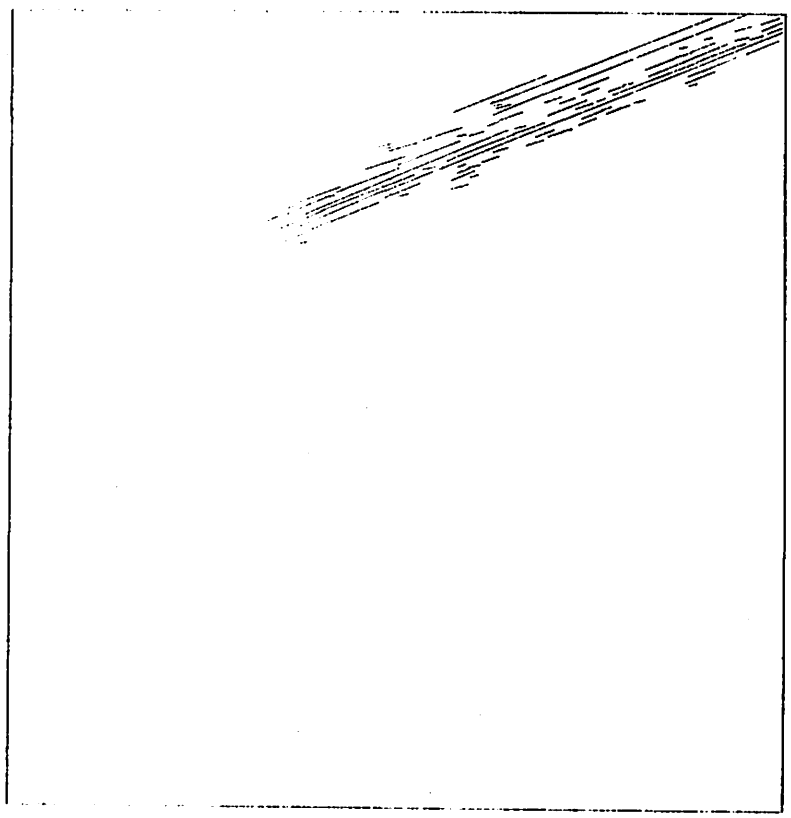

Figure 29: Line Group Based on Spatially Proximate Collinearity and Parallelness. Selected Over Group Shown the Previous Figure. Contains 117 Lines.

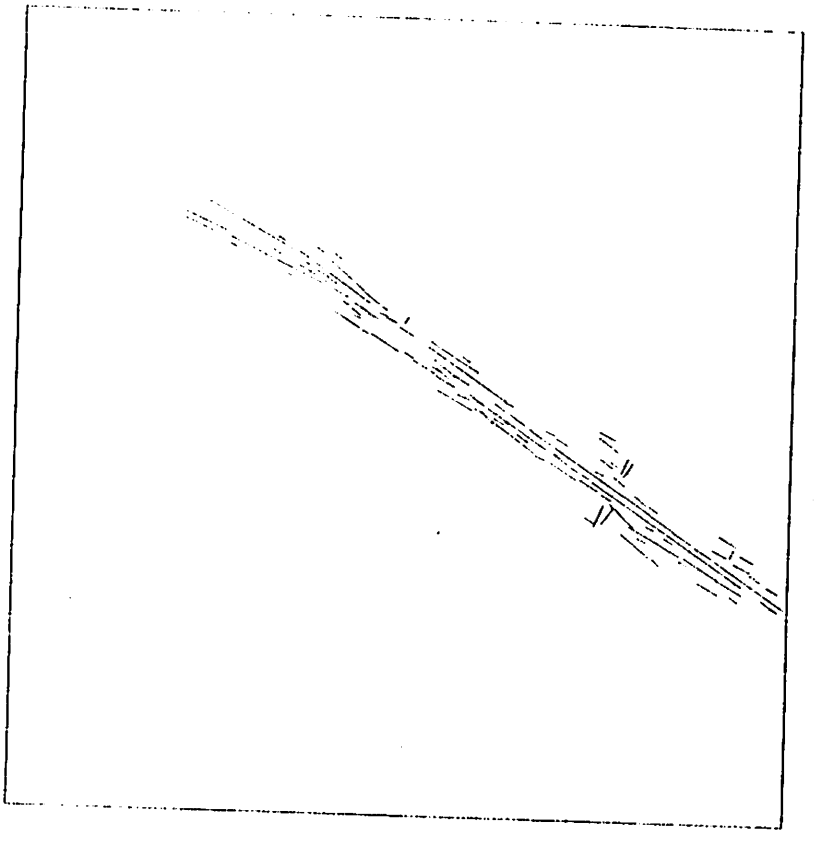

Figure 30: Line Group Based on Spatially Proximate Othogonality, Collinearity and Parallelness. Group Contains 107 lines with a cumulative length of 902.

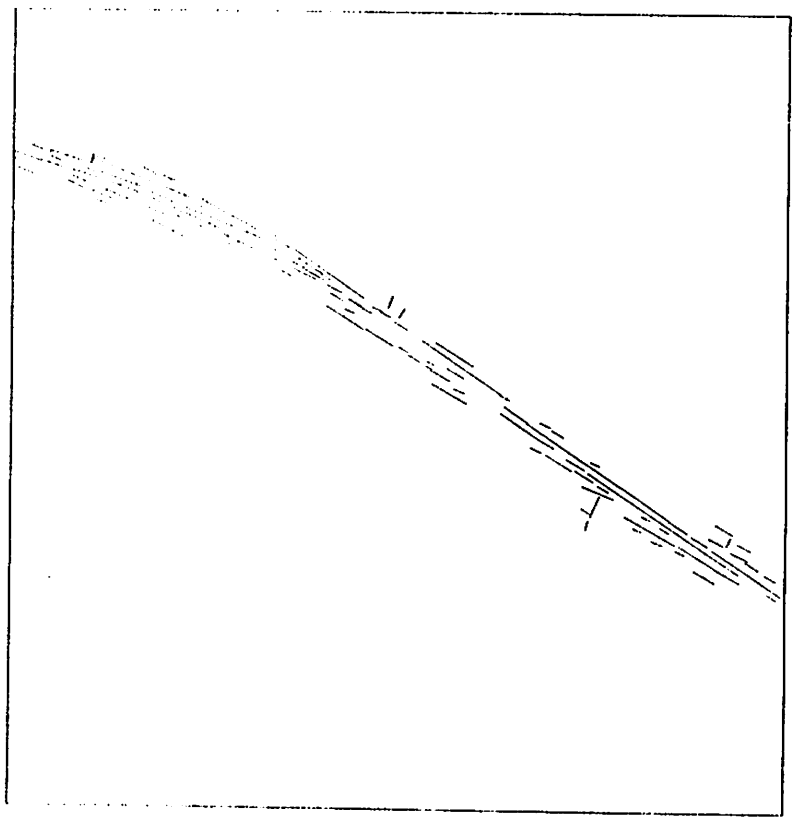

Figure 31: Line Group Based on Spatially Proximate Othogonality, Collinearity and Parallelness. Group Contains 102 lines with a cumulative length of 983.

two central issues in geometric grouping had to be addressed. First, the importance of building structures which can serve as primitives for defining scene events and second, the importance of pruning the enormous search spaces which contain the projections of the scene events of interest. The Gestalt Laws of perceptual organization and in particular many of the rules of simplicity and economic encoding provide a framework for developing descriptions and algorithms which constrain the number of hypothesis generated.

In this paper we have dealt only with line organization, and our future work includes extending the algorithms we describe to include line and region relations. For example in the algorithm described in this paper we have not used the direction of the intensity gradient across the line, color information in a neighborhood about the line or information about regions of a segmentation to which the line is adjacent, to constrain the grouping processes. Each of these additional constraints will allow the generated structures to be more complete descriptions of some local area in the image, allow further pruning of the search space, and provide a richer set of primitives for higher level processes.

Acknowledgements

We would like to thank Al Hanson, Bruce Draper and Michael Boldt for many valuble discussions during the development of this work.

References

[1] R. Arnheim, *Art and Visual Perception, A Psychology of the Creative Eye*, University of California Press, Berkeley, 1974.

[2] R. Belknap, E. Riseman, and A. Hanson, "The Information Fusion Problem and Rule-Based Hypotheses Applied To Complex Aggregations of Image Events," *Proc. DARPA IU Workshop*, Miami Beach, FL, December 1985.

[3] R. Beveridge, A. Hanson, and E. Riseman, "Segmenting Images using Localized Histograms," COINS Technical Report, University of Massachusetts at Amherst, in preparation, 1987.

[4] I. Biederman, A. Glass, and E.W. Stacy, "Searching for Objects in Real-World Scenes", *Journal of Experimental Psychology* Vol. 97, No. 1, 1973, pp. 22-27.

[5] R. Brooks, "Symbolic Reasoning Among 3-D Models and 2-D Images," *STAN-CS-81-861*, and **AIM-343**, June 1981, Department of Computer Science, Stanford University.

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

[7] B. Draper, R.Collins, J Brolio, J. Griffith, A. R. Hanson, E. Riseman, "Tools and Experiments in the Knowledge Directed Interpretation of Road Scenes", this proceedings.

[8] A. Guzman, "Decomposition of a Visual Scene into three-dimensional bodies", in A Grasselli, ed., *Automatic Interpretation of and Classification of Images*, Academic Press, 1969.

[9] A.R. Hanson and E.M. Riseman (Eds.), *Computer Vision Systems*, New York, Academic Press, 1978.

[10] A. Hanson, and E. Riseman, "VISIONS: A Computer System for Interpreting Scenes, in *Computer Vision Systems* (A. Hanson and E. Riseman, eds.) pp. 303-333, Academic Press, 1978.

[11] A. Hanson and E. Riseman, "Segmentation of Natural Scenes," in *Computer Vision Systems,* (A. Hanson and E. Riseman, Eds.), Academic Press 1978, pp. 129-163.

[12] A.R. Hanson and E.M. Riseman, "A Methodology for the Development of General Knowledge-Based Vision Systems," to appear in *Vision, Brain, and Cooperative Computation*, (M. Arbib and A. Hanson, Eds.) 1986, MIT Press, Cambridge, MA.

[13] M. Herman and T. Kanade, "The 3D MOSAIC Scene Understanding System: Incremental Reconstruction of 3D Scenes from Complex Images", *Proceedings of the DARPA IU Workshop*, October 1984, pp. 137-148.

[14] G. Kaniza, W. Gerbino, "Convexity and Symmetry in Figure Ground Organization", in M. Henle, ed., *Vision and Artifact*, Springer, New York, 1976.

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

[16] D. Marr, *VISION*, W.H. Freeman and Company, San Francisco, 1982.

[17] D. McKeown and Pane, "Alignment and Connection of Fragmented Linear Features in Aerial Imagery", CMU Tech Report, 1985.

[18] M. Nagao and T. Matsuyama, "A Structural Analysis of Complex Aerial Photographs," Plenum Press, New York, 1980.

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

[20] G. Reynolds, J. Ross Beveridge, "Geometric Line Organization Using Spatial Relations and a Connected Components Algorithm", COINS Technical Report, University of Massachusetts at Amherst, in preparation, 1987.

[21] G. Reynolds, N.Irwin, A. Hanson and E. Riseman, "Hierarchical Knowledge-Directed Object Extraction Using a Combined Region and Line Representation," *Proc. of the Workshop on Computer Vision: Representation and Control*, Annapolis, Maryland, April 30 - May 2, 1984, pp. 238-247.

[22] I. Rock, *The Logic Of Perception*, MIT Press, Cambridge, 1983.

[23] K. Sugihara, *Machine Interpretation of Line Drawings* MIT Press, Cambridge, Mass 1986.

[24] M. Tavakoli, A. Rosenfeld, "Building and Road Extraction from Aerial Photographs", *IEEE Transactions on Systems, Man and Cybernetics* vol 12, 1982.

[25] T.E. Weymouth, "Using Object Descriptions in a Schema Network For Machine Vision," Ph.D. Dissertation, Computer and Information Science Department, University of Massachusetts at Amherst. Also COINS Technical Report 86-24, University of Massachusetts at Amherst, 1986.

[26] D. Waltz, " Generating Semantic Descriptions from Drawings of Scenes with Shadows", in P. Winston ed., *The Psychology of Computer Vision*, McGraw-Hill, New York, 1975.

[27] R. Weiss, M. Boldt, "Geometric Grouping of Straight Lines", Proc. IEEE Conf. Computer Vision and Pattern Recognition, Miami Beach, June 1986.

[28] A. P. Witkin and J.M. Tenenbaum, "What Is Perceptual Organiztion For?", *IJCAI*, 1983, pp. 1023-1026.