

Figure 25 Fort Hood - Scene 2

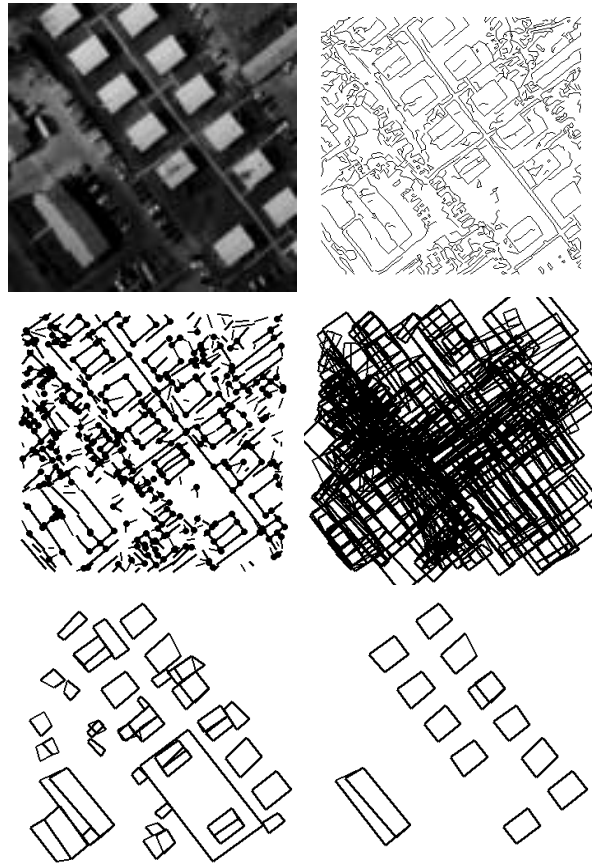


Figure 28 Fort Hood - Scene 5

Figure 29 Modelboard Oblique scene 1

Figure 30 Modelboard Oblique scene 2

Figure 31 Modelboard Oblique scene 3

Figure 26 Fort Hood - Scene 3

- Orthogonal Trihedral Vertices or OTVs will now be detected as for many views both the roofs and the walls will be visible.
- U-contours will be replaced by *skewed* U-contours. The angle between the base and the parallel sides will no longer be restricted to right angles. However, since parallel lines still map to parallels, the sum of the angles between the base and the sides of a skewed-U will be constrained to lie near 180_{circ} .
- Rectangles will be replaced by parallelograms. Note that the parallelogram is composed of skewed-Us, parallels and lines in a fashion essentially the same as for rectangles.
- One possible new grouping to be considered is that of a “hinge.” A hinge would correspond to two parallelograms which share a side. This creates two OTVs at the ends of the shared line. Given one parallelogram that is the projection of a side of a solid rectangular object, there is high probability that another side of the object, sharing a boundary, is visible. However, there could be cases where no hinge is present with a parallelogram due to occlusion. Therefore, a hinge would be a useful but not necessary grouping for detecting buildings.

The detection process for the grouped features, would essentially be the same as in our current system. However, due to the removal of some structural regularity (namely, right angles) the use of some geometric constrains may not apply.

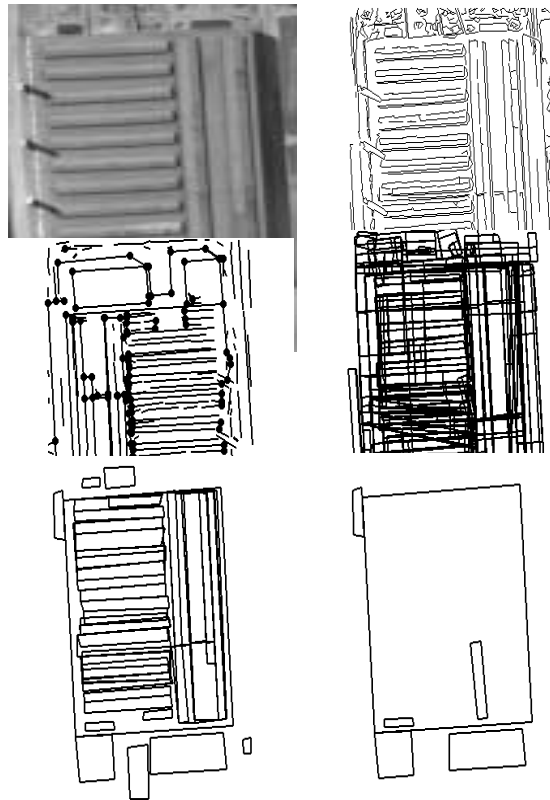


Figure 23 Modelboard - Scene 4

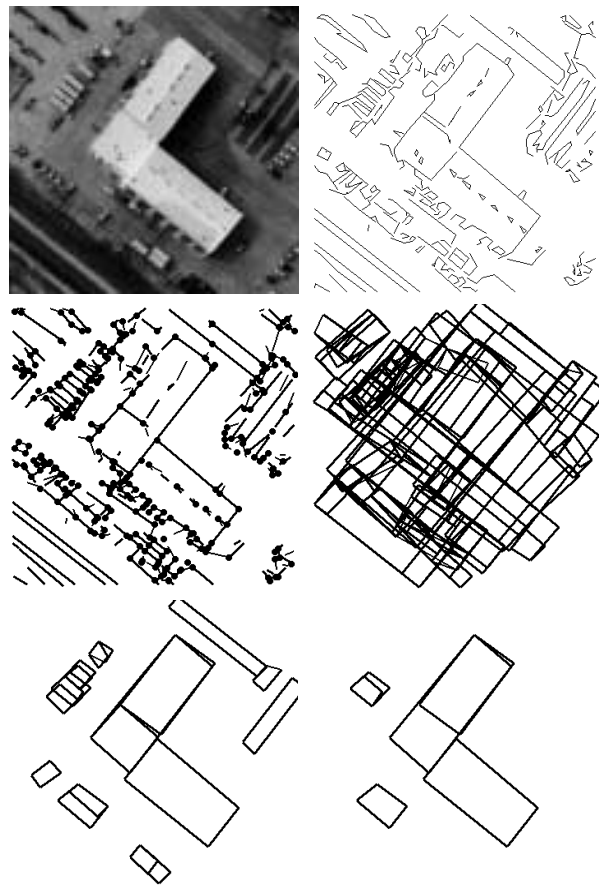


Figure 24 Fort Hood - Scene 1

Figure 27 Fort Hood - Scene 4

rectangles representing grassy areas and parking lots, with the appropriate sidewalks parallel to the building.

Figure [Ref: hood9] a group of small buildings is shown

Also, T-junctions for traditional overhead urban aerial imagery do not have the usual interpretations of occlusion, as some of them would for oblique views. When the image plane is nearly parallel to the ground plane, the buildings may have wings and nearby objects like roads, etc. that are aligned to the building sides. In a top view, the sides of two different structures can create T-junctions in which the top line belongs to two different objects and is not occluding the stem. Therefore, the T-junctions are used to break the line belonging to the top of the T into sections.

Figure 21 Modelboard - Scene 2

7 Future: Integration of Information from Multiple Viewpoints and Sources

In the scenes analyzed by the systems mentioned above, many of the objects have restricted shapes and often the viewpoint is restricted. For some applications, it is necessary to integrate information extracted from images of a scene acquired from various viewpoints or acquired through various types of sensors. At the present time we are not aware of complete systems that provide information integration for photointerpretation tasks, and the existing techniques would have to be reviewed to determine the feasibility of relaxing viewpoint restrictions.

In our systems for instance, we detect two types of corners between the lines, L- and T-junctions. We currently do not investigate orthogonal trihedral vertices (OTVs) as few walls are visible, and those that are appear highly foreshortened and have shadows etc. near them making the OTVs difficult to detect accurately.

Figure 22 Modelboard - Scene 3

If we continue to assume that we restrict the shape of the objects to rectangles, the most significant change is that right angles in the real world no longer necessarily project onto right angles in the image. The following changes to our systems, some already suggested in [Moha89], are currently being incorporated:

- L and T junctions will no longer be considered only for lines meeting at right angles. The lines may meet at any angle.

[Ref: hood3-r-ele] encodes the height computed for each rectangle for each pixel inside the rectangle. A 3-D rendered view of this model is shown in figure [Ref: hood3-r-ren]

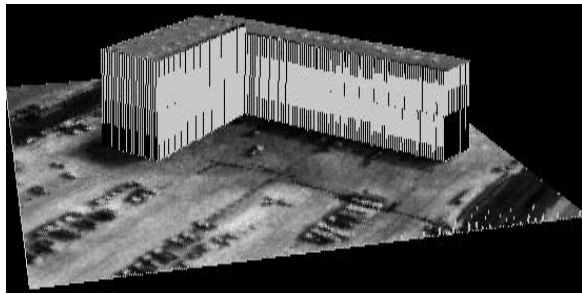


Figure 193-D view from arbitrary viewpoint

6 Results

We have tested our technique on many images from the Ft. Hood, Texas site, from Newport Beach, California, and from a modelboard site. We have selected a few here that demonstrates the performance of our system..

Figure [Ref: j3-b1] shows a set of four buildings and part of another. The difficulty here is with the building having a patterned arrangement of small objects on the roof. The shadows cast by these reach the bottom boundary of the building causing it to be fragmented. The shadow occluding the top left corner of the building and the poor boundary definition on the top right are also a source of difficulty. The strong shadow cues however help form rectangle hypotheses for most of the building

In figure [Ref: j3-b2] The small building on the top left corner must be detected separately from the large one.

Figure [Ref: j3-b3] shows two dark buildings. The boundaries between buildings and shadows in cases like this has low contrast and are difficult to detect.

Figure [Ref: j3-b4] shows a complex building with numerous rectangular components on the roof. We are able to exploit here the presence of strong shadow evidence in the form of junctions and shadow lines that allow the system to form a hypotheses for the entire building inspite of the broken and fragmented top and right-side boundaries. Note that the selection mechanism is able to select most of the rectangular components on the roof.

Figure [Ref: j3-w2] shows a number of buildings. The boundaries of some of them would be difficult to follow as they merge with shadow boundaries or have low contrast and become fragmented and distorted.. The rectangular area on the left appears to be a walled area with no roof and the walls cast shadows.

Figure [Ref: hood1] shows a building

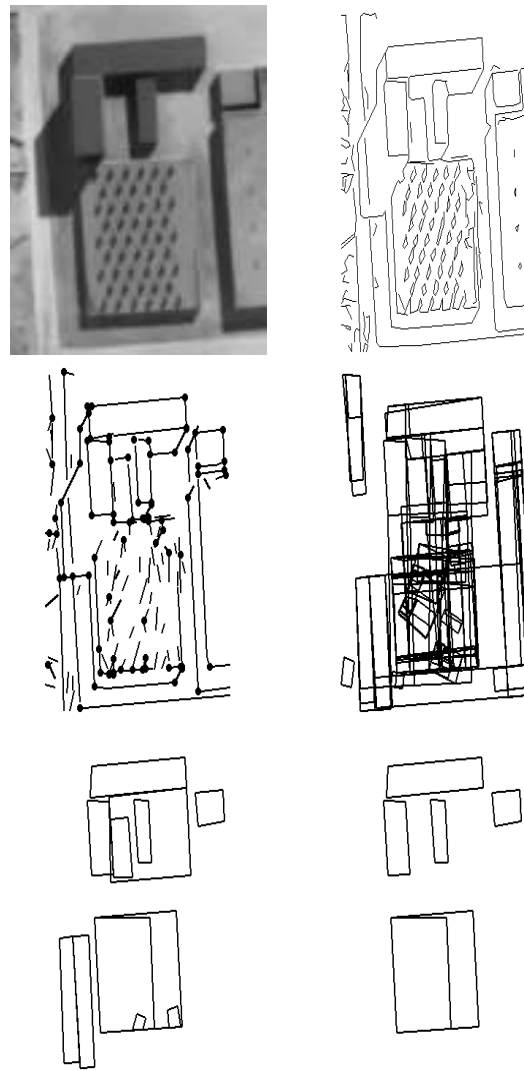


Figure 20 Modelboard - Scene 1

where some of the details of one of its sides is visible, apparently doors. These and the vehicles parked on the other side result in highly fragmented boundaries. The rectangles verified by shadows include one that is formed from various aligned parked trailers which collectively cast a shadow. The small rectangle on the bottom has a strong shadow junction corresponding to an actual shadow cast by a vehicle. This would give a very narrow shadow width and we do not filter these on shadow width in the current version. The lower wing of the building has a strong line and a corresponding medium junction. The rest of the shadow is diffused and only gives a “dark” region next to this and the upper wing.

In Figure [Ref: hood2] The I shaped building has no strong evidence of shadows. The validated rectangles are on the basis of a strong region which up a given maximum search distance remains “strongly” dark.

In Figure [Ref: hood5] we show a not uncommon situation. The building is surrounded by

the line and parallel to it, we sweep a search line along the direction of illumination for a determined distance (see figure [Ref: sha-search]). This distance is arbitrarily chosen as a function of the maximum expected building height and the sun incidence angle. We have assumed that the sun angles are available, as they are easily computed from the date of acquisition and the geographic location of the scene. Note from the figure that only the solid heavy black lines and junctions are part of the search space. The large windows on the other hand, denoted by dashed lines in the figure are designed to extend beyond the shadow boundaries. The possibility that other lines, not relevant to the current rectangle be included is high. The collection of lines and junctions detected for each rectangle side are evaluated (see below).

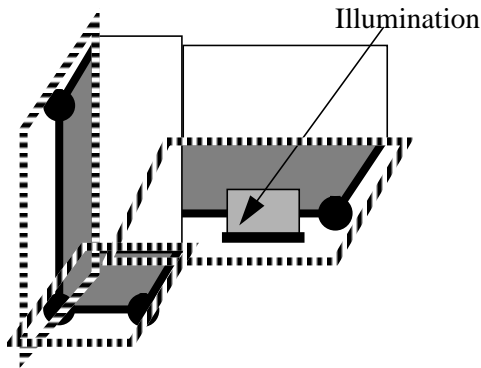


Figure 17 Windows to search for shadows

The following evidence is extracted if found: Strong lines, medium junctions (at ends of strong lines), strong junctions. From the medium lines we favor more those parallel to the rectangle side. In some cases there may be various sets of lines, all parallel to the building side but at various distances from the rectangle side. In this case we choose those shadow lines at the distance from the rectangle side such that the sum of their lengths is larger. This is actually a common occurrence since many side walks, grass areas, streets, vehicles and so on, will be found to be arranged or located parallel to building sides.

From the junctions and/or the medium shadow lines, and/or the lengths of the strong lines we determine the average width of the shadow for each side. Next we compute the mean intensity for the region adjacent to the side in question and up to the width estimated.

The evidence collected for both sides is combined to give the evidence for the rectangle.

5.3.3 Evaluation and use of Shadow Evidence

The evidence collected is evaluated as follows:

In order to evaluate the shadow evidence

and give a confidence value we assigned points to the various features: Strong junction (20), medium junction (15), Strong line (15), Weak lines (10), Strong region (15), Weak region (10). For a single rectangle with 2 sides casting shadows the total possible is 115. High confidence is 100, the extra 15 is a bonus from the strong region contribution, that is whether the dark region adjacent to the building is "dark enough" to be a shadow. We have considered methods to estimate this threshold from a scene [Huer83] automatically, mostly from histogram analysis or similar techniques, but consider these intensity-based measurements weak in general. In the current version we set the threshold for darkness to some arbitrary value. If the region is dark enough, the bonus is added.

We designated 5 levels of confidence as follows. More than a 100 points represents very high confidence. Everything shadow is there. More than 80 points denote high confidence but we require that certain evidence be present. There must be at least a strong or medium junction present, at least one strong line, a set, however small of medium lines, a strong region on either side of the rectangle, and the relative brightness requirement must be met. More than 60 points denote moderate evidence. Here we require that at least one strong line be present, a set of medium or weak lines, that a dark region be present, and that the shadow region be darker than the rectangle region. More than 40 points denotes fair confidence. We require that there be a set of medium lines at least, a dark region on either side of the rectangle. The brightness of the shadow should be lower than that of the rectangle. Above 20 points denotes low confidence. Since no evidence was found from the geometric constraints we at least require that a sample of the regions on both sides of the boxes denote either strong shadow be present, or at least that the relative brightness of the adjacent regions be lower than that of the rectangle.

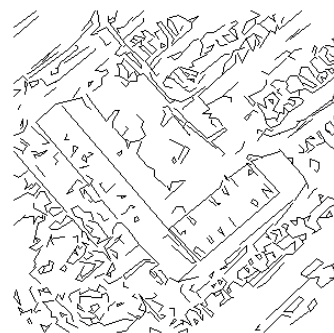


Figure 18 Elevation image. Pixels encode height

The rectangles validated by shadows are used as the footprint of buildings or portions of buildings. The shadow widths are used to estimate their height. The elevation image shown in figure

by smaller objects inside or near the shadow regions. They may also correspond to shadows cast by the buildings but laying on these objects surfaces. These lines are detected by the shadow process (see below) but are ignored in the current version of the system.

Strong Regions

The gray level mean of the intensity of a region surrounded by strong and medium junctions and lines denotes strong evidence of shadow region. The actual gray level is not important and contributes weakly to the shadow evaluation process, but we require that this region be darker than the surround, and that the region be darker than the rectangle region also..

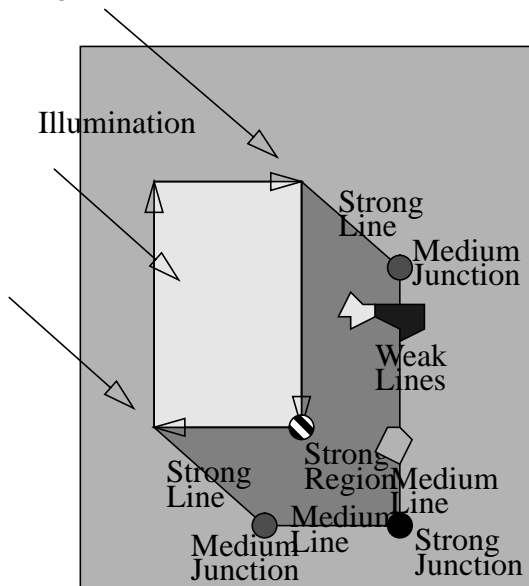


Figure 15 Shadow evidence

Weak Regions

In the absence of any geometric correspondences of junctions and lines, the presence of a dark region on the side of a rectangle consistent with the direction of illumination constitutes weak evidence of a shadow if the region is darker than the rectangle region. This weak evidence is used on cloudy days when shadow lines are not sharply defined, but are widened by the diffused illumination.

5.3 Shadow Process

The shadow process consists of three steps:

- Detection of potential shadow evidence: We detect for the entire scene the potential shadow lines and junctions.
- Search for shadow evidence: For each rectangle selected we search explicitly for nearby shadow evidence among the lines and junctions detected in the previous step.

- Evaluation of shadow evidence: The evidence collected is evaluated by a function that considers its combined strength and assigns a confidence value. It also computes shadow widths for height estimation.
- Use of shadow evidence: Validated rectangles are used to generate 3-D models of the buildings that can be used for site modelings, planning, change detection, 3-D maps, realistic 3-D renderings, and so on.

5.3.1 Detection of Shadow Evidence

Shadow evidence consists of lines, junctions and intensity statistics. Shadow lines are considered strong or weak evidence as follows:

- Lines parallel to the projection of the sun rays are potential shadow lines cast by vertical edges of 3-D structures in the image.
- Lines not cast by vertical edges but having their dark side on the side of the illumination source are potential shadow lines.
- Junctions among the lines above are computed in the same manner as those to compute junctions for hypotheses formation.
- Pixel statistics to compare relative brightnesses consists of the mean intensity only.

The potential shadow lines detected from the lines in our Ft.Hood example are shown solid in figure [Ref: hood3-r-sha-evi] . The underlying lines are shown in gray.

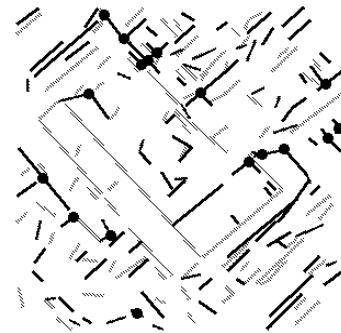


Figure 16 Potential shadow lines and junctions

5.3.2 Search for Shadow Evidence

We look in the set of potential shadows lines and junctions. For each rectangle to be validated, we determine the two (or one if there is alignment with the sun rays) lines representing the lines that should cast shadows if this rectangle represents a building or a portion of a building. Rectangle pixel statistics are collected at this time.

Next, for each of the sides, and starting at

we have found that perceptual organization gives us useful high level features. Even if we obtain multiple groupings, or even some wrong groupings, from the grouping process (in contrast to unique features as in edges or regions) it is still simple to perform correspondence and segmentation with the groupings computed in 2-D, or with additional information from stereo, to obtain unambiguous segmentation at object level.

5.2 Shadow Analysis

The basis for the verification of 3-D structures from monocular cues are the shadows that they cast on sunny days. We assume that the direction of illumination is available and that the ground surface in the immediate neighborhood of the structure is fairly flat and level. This assumption allows to compute accurate height information. When these surfaces are not flat or level, or are cast on surfaces of adjacent structures, the quality of the 3-D information suffers in the absence of terrain data but still provides a source of verification.

By shadow analysis we mean the establishment of correspondences between shadow casting elements and shadows cast. Given the direction of illumination we attempt to establish these correspondences between shadow casting lines and shadow lines, and between shadow casting junctions (or vertices) and shadow vertices. In addition, the foreshortened or non-visible vertical edges of buildings cast visible shadows in a direction parallel to the projection of the sun rays on the ground. These shadow lines start at the edge of a building and can be easily detected. Even when these lines are partially cast on other surfaces of nearby structures, portions of them satisfy the projection constraint and can be reliably detected and used.

There are a number of difficulties that may prevent the accurate establishment of correspondences however. Building sides are usually surrounded by a variety of objects such as loading ramps and docks, grass areas and sidewalks, trees, plants and shrubs, vehicles, light and dark areas of various materials. Nearby structures may reflect light into the shadowed areas making the objects in it more visible, and so on. To deal with these problems we have adopted the following criteria and geometric constraints to analyze the shadows adjacent to rectangles (see figure [Ref: sha-evidence] and also [Huer83,Huer-Neva88]):

Strong Junctions

A rectangular structure has two sides that cast shadows (except when the rectangle sides of the rectangle are coplanar with the sun rays and their projection). The lines representing these sides form a junction which casts a matching shadow junction on the first encountered surface along the sun ray. These matching junctions have a consistent *shape* and a consistent *attitude*. If the angles between the

lines forming the junctions are the same within a small tolerance then the junctions have the same shape. The angle tolerances in our system are 5° (to allow for inaccuracies in the line detection process and for small perspective effects). If the bisector of two junctions are parallel and oriented in the same direction then the junctions have the same attitude. The correspondences established between these junctions constitute the strongest monocular cue to the presence of a 3-D structure. As mentioned in an earlier section, we rely on this matches to form and select hypotheses.

Strong Lines

Vertical building edges cast shadow lines in a direction similar to the direction of the projection of the sun rays. These lines begin at the base of the vertical (and mostly non-visible) edges that are connected to the shadow casting lines described in the paragraph above, at the other end of the Strong Junctions. Since we assume that most buildings have vertical sides, these evidence when found is also very strong and reliable. We use it as well, as mentioned earlier, during hypotheses generation and selection.

Medium Lines

The rectangle lines that form strong junctions cast shadow lines, in the direction of illumination, to correspond to the shadow lines that form the shadow junction. These shadow lines can be very fragmented due to problems already mentioned. In some cases nearby lines make it difficult to establish accurate correspondences, thus, these lines are less reliable.

Medium Junctions

The junctions formed between strong lines and medium lines can also be found along the direction of the strong lines. The shape and attitude of these junctions, however, can vary widely depending on the surface they fall onto. Although less reliable, medium junctions and strong junctions falling on flat level ground match the building geometry accurately and are reliable features to measure shadow width.

Weak Junctions

Junctions and breaks in the shadow boundaries between the strong and weak junctions may denote changes in the height of the shadow casting side of a building. We have made use of these breaks in [Huer83] but consider it less useful here, as many of these breaks are caused by objects in the shadows, like trees and bushes, and by shadows cast by nearby structures. Although we also detect these, they make no contribution to the shadow evaluation process (see below) and are mostly ignored in the current version of the system.

Weak Lines

Lines laying between medium lines and not colinear with these usually correspond to shadow lines cast

crease the description of the scene. There is an exception when rectangle A is fully contained by rectangle B but the extra evidence of support for rectangle B is weak then rectangle A will be selected instead of rectangle B..

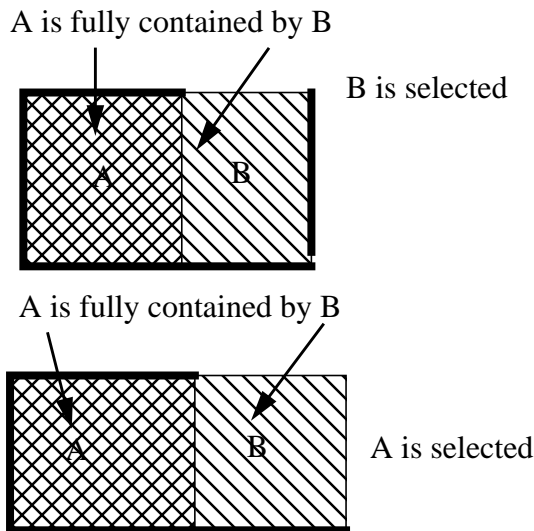


Figure 14 Rectangle A is fully contained by rectangle B

For the reason of efficiency, we actually implemented some of the global selection rules together in a function.

5 Verification and Generation of 3-D Models

The purpose of a verification mechanism is to validate the selected hypotheses to correspond to the objects of interest. The validation step should also segment the objects and generate a description of the shape of the structures in the form of a 3-D model. Our systems uses shadows for verification and estimation of height extracted from monocular views, and assumes that the sun angles are given.

5.1 Other Methods: Stereo and Probabilities

Recent stereo systems have shown improved performance by using more structure than individual edges [Coch90, Lim87, MedNev85, OhtaKan83]. Mohan and Nevatia [Moha-Neva89c, Moha89] use rectangles for stereo matching, object segmentation and shape description. The object segmentation automatically provides a shape description of the roof in terms of the component rectangles. The segmented roof area and their heights are used to generate 3-D models of the buildings.

For multiple building detection tasks, however, edge and segment based stereo matching algorithms have displayed poor performances. The following factors indicate why stereo systems based on simple image features may not perform well in

this domain:

- **Organized nature of the scene.** There are numerous parallel lines since the buildings, roads, parking lots, etc. are all parallel. This leads to the same type of problems as with monocular analysis.
- **Absence of texture.** The buildings sides mark regions with high disparity differences and there are insufficient markings on the roofs to support match-disparities at roof level while matches giving low disparities get favored due to the preponderance of features on the ground.

Mohan's system has the drawback of using stereo to select among the existing rectangles but of not using it to check for missed rectangles. Also there is a loss of accuracy in the determination of the disparities as a result of the robustness in the detection of the matched primitives. The rectangles are grouped features, and are thus primarily structural representations with low positional accuracy. The component lines of the rectangle only represent the structure among the underlying edges, not their exact positions. For obtaining accurate disparity, matching of more precisely located features, namely the edges, is required, using a system like the one described in [Coch90].

Using another approach, Lowe [Lowe85] suggests computing prior probabilities of accidental and non-accidental instances of certain relations (collinearity, proximity of end-points, and parallelism) and using this to constrain the search in model matching. The prior probabilities are computed assuming a random distribution of straight lines in the scene. The groupings are used primarily to reduce both the number of features for matching and the possible transformations (camera viewpoint and object orientation) suggested by the match. We continue to believe there are shortcomings with the approach of computing prior-probabilities for assigning significance to groupings. First, this seems hard to do in general for real scenes. Second, in our domain of aerial images, some scenes, such as urban areas, consist of an organized layout of objects, and the assumption that the objects are randomly oriented breaks down. In fact, perceptual organization has been used to exploit this very fact of organized layout of objects in a scene [QuaMoh88, Rey87]. We believe that the approach used in our work, on basing the significance of a grouping on its comparison to its alternates and its relationships with related groupings in the hierarchy (by part-of relationships), is a more reasonable solution.

Lowe proposes that segmentation is one important task of perceptual organization but proposes no mechanism for it. He states that "A major reason why perceptual organization has not been a focus of computer vision research is probably because these groupings often do not lead immediately to a single physical interpretation." In our work

possible combination.

There are lots of ways to find the approximated optimal weight assignment. The following is a simple way of finding an approximate solution. First of all, we manually classify many rectangles formed by the previous stage into good rectangles and bad rectangles. For each sampled weight assignment, we evaluate the values of all rectangles. Find the comparison value for this weight assignment which is the greatest value smaller than all the values of good rectangles. For all the sampled weight assignments, find the assignment which has the smallest number of bad rectangles with evaluated values greater than the comparison value associated with the assignment.

4.2 Global Selection Rules

Those good rectangles surviving from local selection could compete with each other. For example, some rectangles could share the same edges support or corners support and some rectangles might overlap with each other. The goal of global selection rules is to select a minimum set of rectangles which could best describe the rectangular composition of the scene.

Each global selection rule will be applied to the set of rectangles separately. It acts like a filter. Every time a global selection rule is applied, some rectangles will be filtered out. The final set of rectangles is the best selection according to our rules. The order of applying these global selection rules is also important.

There are four global selection rules described as follows in the order of applying.

Selection Rule for Duplicated Rectangles

It is possible for our system to create duplicated rectangles and they all get the same support, if their support is strong enough, these duplicated rectangles will all survive local selection. Actually one of them will be sufficient to describe the corresponding scene, so we will randomly selection one rectangle and remove all the other duplicated rectangles.

Selection Rule for Mutually Contained Rectangles

For any two different rectangles A and B, we define rectangle A contains rectangle B, in other words, rectangle B is contained by rectangle A, if more than 95% of rectangle B is inside rectangle A. If rectangle A contains rectangle B and rectangle B contains rectangle A, we say rectangle A and rectangle B are mutually contained by each other. This rule says that we will select one rectangle with highest evaluation value, that is, the best rectangle, from a set of mutually contained rectangles. Actually, this rule is just an extension of the previous rule.

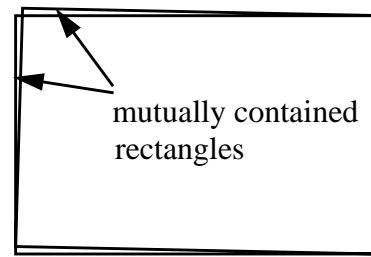


Figure 12 Mutually contained rectangles

Selection Rule for Overlapping Rectangles

If a rectangle does not overlap with any other rectangles, it is not necessary for us to worry about that rectangle, because there is no competition between that rectangle and other rectangles. When two rectangles are overlapped with each other but no one is contained by the other, there is no rule for this situation, because they are almost equally important. When two rectangles are overlapped with each other and one is contained by the other, the innermost contained rectangle, that is, no other rectangles are contained by this rectangle, will be selected with higher priority. For those rectangles containing some other rectangles, they will be given different priorities which depend on the rectangles they contain

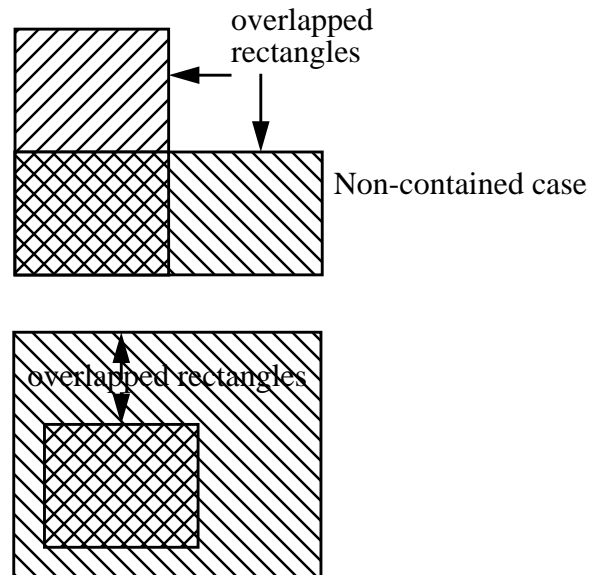


Figure 13 Overlapping rectangles

Selection Rule for Fully Contained Rectangles

For any two different rectangles A and B, rectangle A is fully contained by rectangle B if all the supporting evidence of rectangle A is included by supporting evidence of rectangle B. We tend to remove those rectangles which are fully contained by some other rectangles, because their existence can not in-

used in our system:

Lines Crossing Sides of a Rectangle

If a rectangle is a part of the contour of a building, the probability of a line crossing any side of the rectangle is very small. The case only happened when a marking line on top of a building is accidentally connected to a colinear line on the ground. Especially when the place where the line crosses the side of a rectangle has no edge support for the rectangle, it almost suggests that it could not be a good rectangle. Of course there is an exception, but the case is rare.

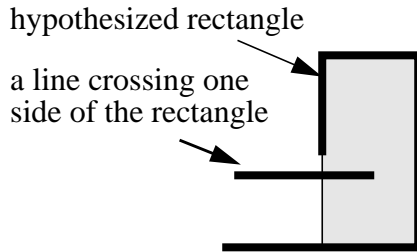


Figure 8 Lines crossing sides of rectangle

Sides of a Rectangle Going Through L-Junctions or T-Junctions

A side of a rectangle is usually confined by two consecutive specific kinds of L-junctions or T-junctions. In the case of L-junction, it must be formed by the side of the rectangle and its adjacent side of the rectangle. In the case of T-junction, the side of the rectangle must be the stem of the T-junction. Thus, it is unreasonable for any side of a rectangle to go through any of the specific kinds of L-junctions or T-junctions..

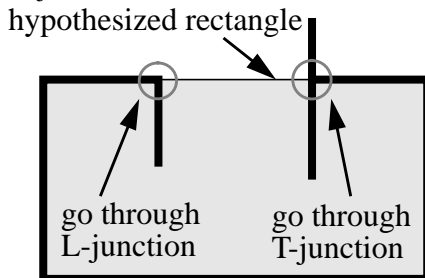


Figure 9 Sides of a rectangle going through L-junctions

Overlapping Gaps on Opposite Sides of a Rectangle

It is possible that some edges will be missing, but the chance for missing edges support on opposite sides of a rectangle is small, that is, overlapping gaps on opposite sides of a rectangle is a negative evidence of support. Overlapping gaps on opposite sides of a rectangle usually imply that it's better to separate the rectangle at the place of the overlap-

ping gaps..

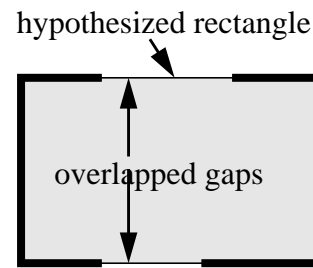


Figure 10 Overlapping gap on opposite sides of a rectangle

Displacement between an Edge Support and its Supporting Side of a Rectangle

The edges support of a rectangle should be as close to the hypothesized side of the rectangle as possible. If there is a large displacement between an edge support and its supporting side of a rectangle, the edge support could be a wrong edge support. So, we have to count it as a negative evidence of support. Although this kind of evidence is weak, we can still use it to distinguish some good rectangles from bad rectangles..

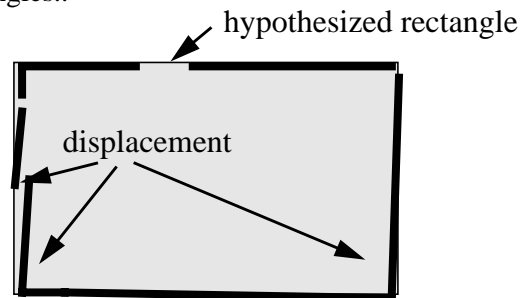


Figure 11 Displacement between an edge support and its supporting side

Actually, some of these evaluation rules could be applied earlier in the stage of forming rectangles, however, we have chosen to allow the formation of rectangles in a liberal manner, and evaluate them at this stage.

4.1.3 Weights of Evaluation Rules

Evaluation rules for positive evidences of support will be assigned positive weights and evaluation rules for negative evidences of support will be assigned negative weights. Positive weights range from 0 to 1 and negative weights range from 0 to -1. All the evaluation rules will return values between 0 and 1. Since these evaluation rules are heuristic, there is no absolute way of assigning weights to these rules. So we have to decide the weight assignment by testing all the possible combination of weights and comparing their results. Of course the possible combination of weights are infinite, so we have to take and test only some samples of these

Basically whether or not a rectangle is good depends on evidences of support of the rectangle. There exist *positive evidences* and *negative evidences* of support for a rectangle. Negative evidences are as important as positive evidences, because they will help us to remove those rectangles which are almost impossible to be part of buildings. We could formulate some evaluation rules to evaluate some important evidences of a rectangle.

The positive evidences of support we used include existence of edges, corners, parallels, and shadow. The negative evidences of support we used include existence of lines crossing any side of a rectangle, existence of L-junctions or T-junctions in any side of a rectangle, existence of overlapping gap on opposite sides of a rectangle, and displacement between four sides of a rectangle and its corresponding edges support,

There exist some difficulties for making evaluation rules. Since it is hard to make a clear and formal definition of goodness of a rectangle, it is also difficult for us to find out evaluation rules which could well define a good rectangle. So, the evaluation rules we used are all heuristic and with high probability a rectangle is good if it is evaluated to be good by these rules. Also, these rules are made to be as general as possible, so that the system can deal with the case where noise exists and thus making the system robust.

4.1.1 Positive Evidence

The following positive evidences of support are used in our system:

Existence of Edges Support

It is obvious that a good rectangle must have some edges support, otherwise we can not even recognize it as a rectangle. Not all edges of a rectangular part of a building could be detected. If the contrast between the building and its surrounding environment is low or the rectangle is connecting to other part of the building, weak edges will not be detected or the edges don't even exist..

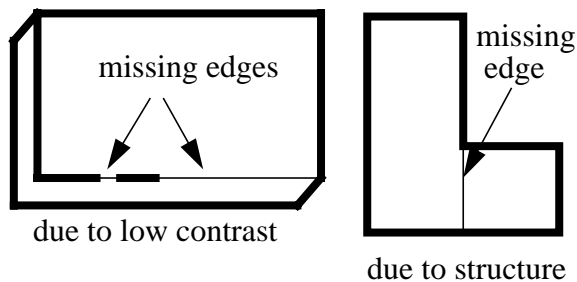


Figure 5 Missing edges

Existence of Corners Support

In most cases, a rectangular part of a building will include some corners. So, if there exist corners sup-

port in a rectangle, especially when there exists three or four corners, it highly suggests that this is a good rectangle. While edges around corners sometimes are not detected, we have to extend some edges and hypothesize some reasonable corners. The closer the edges forming the hypothesized corner are the better the corner is. Also, whether or not the angle of a corner is near 90 degree is another measurement of goodness of the corner..

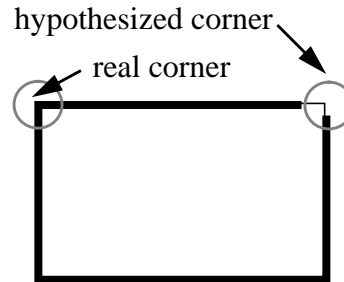


Figure 6 Hypothesises reasonable corners

Existence of Parallels Support

Two pairs of parallel sides is an important feature of a rectangle. Thus the existence of parallel edges support could be used as an evaluation rule. For the same reason as described in edges support, it is possible that some edges will be missing and then parallel edges support will be effected also by the missing edges..

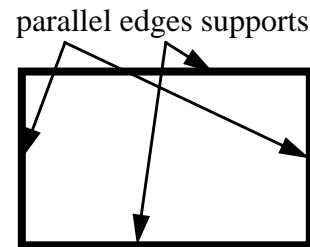


Figure 7 Parallel edges support

Existence of Shadow Support

Given the direction of illumination we look for strong evidence of a corresponding shadow (See section 5.2 below). In particular the correspondence between rectangle corners and shadow corners constitutes strong and reliable evidence. We also look for evidence of shadow boundaries cast by the vertical edges of buildings. If these can be detected, it is very possible that the rectangle is a part of the contour of a building. Since the chance of a rectangle being matched to a false strong shadow evidences is very small, this is a very good evaluation rule.

4.1.2 Negative Evidence

The following negative evidences of support are

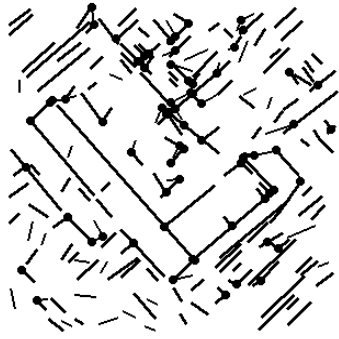


Figure 3 Linear structures and junctions

Parallels and U-structures:

Structures in urban scenes like buildings, roads and parking lots are often organized in regular grid-like patterns. These structures are all composed of parallel sides. As a consequence, for each significant line-structure detected in the scene, there is not one but many lines parallel to it. For each line, we find lines that are parallel and satisfy a number of reasonable constraints. Note that the formation of a *parallel structure* also aids in the formation of new *lines*, as they suggest extension and contraction of the parallels to achieve full overlap.

When the two lines in a parallel structure have their ends aligned, they strongly suggest the presence of a line with which the parallel structure would form a U-structure. Even if the third line does not exist in the set of *lines*, we hypothesize it and generate the U-structure.

Rectangles:

Rectangle structures are generated from the U-structures. The complete set of rectangles in our example is shown in figure [Ref: hood3-r-rec], regardless of their size. In practical applications this number can be reduced by restricting the formation of rectangles on the basis of size, as a function of image resolution, for example.

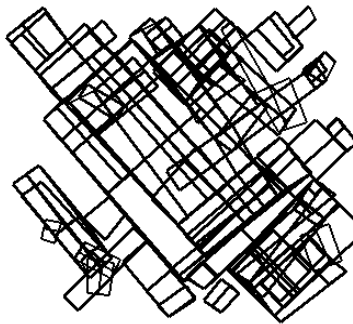


Figure 4 Rectangle hypotheses generated

4 Selection Process

After the formation of all reasonable rectangles, a selection process is used to choose rectangles which have strong evidences of support and have minimum conflict among them. The previous work used Constraint Satisfaction Networks which given the constraints among collated features, minimizing the cost of the network will select the best consistent rectangles for the given constraints. We noticed that the result of Constraint Satisfaction Networks is rather unpredictable and the maintenance and extension of the Constraint Satisfaction Networks is difficult if we want to add new constraints into the networks. So we decided to explicitly represent these constraints as rules then the system could be easily extended and improved by adding new rules.

In our system there are two kinds of rules: *local selection rules* and *global selection rules*. Local selection rules will determine whether or not a rectangle is good based on the local supporting evidences of the rectangle. Only good rectangles will be retained for global selection. These local selection rules are also called *evaluation rules*, since they are used to evaluate the goodness of a rectangle. The “goodness” of a rectangle is dependent on the probability of the rectangle being a part of the contour of a building. It is possible that some of those good rectangles retained after the local selection are mutually contained or duplicated or overlapped with some other good rectangles. Global selection rules will select the best consistent rectangles from good rectangles.

The way we apply local selection rules and global selection rules are different. Local selection rules (evaluation rules) will work together to evaluate the goodness of a rectangle, while global selection rules will work separately. Each global selection rule is like a *filter*. The set of retained rectangles will pass through all filters and the set of rectangles coming out from the last filter will be the set of rectangles selected by the selection process.

4.1 Local Selection Rules (Evaluation Rules)

Some rectangles have been formed on weak evidence; local selection rules are used to remove these rectangles. Given a rectangle, we will use all the evaluation rules to compute a goodness value of the rectangle. If the goodness value of the rectangle is greater than a given threshold, the rectangle is selected, otherwise the rectangle will be removed.

Every evaluation rule will be given a weight according to the importance of the rule. The goodness of a rectangle will then be measured by the sum of the weighted values of evaluation rules on the rectangle. The problem of measuring the goodness of a rectangle now becomes how to find good evaluation rules, how to formulate these rules, and how to assign appropriate weights to these rules.

any domain where the set of specific shapes consists of basic shapes. Apart from changes specific to the selected basic shapes, no changes to the methodology itself should be required to deal with other specific shapes.

In this work we have considered the case of object shapes (building roofs) composed of rectangles. Our system allows us to design a feature hierarchy which encodes the structural relationships specific to this set: Lines, parallels, U-contours, and rectangles are identified as the pertinent structures these shapes can be decomposed into. We now describe these with the aid of an example, referring the reader to the full details, given in [Moha89].

3.1 A Hierarchy of Features

Figure [Ref: hood3-r-img] shows a building from a scene of Ft. Hood in Texas. The building is easy for humans to see and describe, even without stereo, but it is difficult for computer vision systems. Figure [Ref: hood3-r-seg] shows the line segments detected in the image using LINEAR, our Linear Feature Extraction Software [Neva-Babu80]. We are still able to see the roof structures of the buildings readily and easily, but the complexity of the task now becomes more apparent. The building boundary is fragmented, there are many gaps and missing segments. There are also many extraneous boundaries caused by other structures in the scene. While local techniques, such as “contour-following” have proved useful for simpler instances of such tasks, they are likely to fail for the scene of the complexity shown here.

This task is difficult for several reasons. The contrast between the roof of a building and surrounding structures such as curbs, parking lots, and walkways can be low. The contrast between the roofs of various wings, typically made of the same material, may be even lower. Low contrast alone is likely to cause low-level segmentation to be fragmented. In addition, small structures on the roof and objects, such as cars and trees, adjacent to the building will cause further fragmentation and give rise to “extraneous” boundaries. Roofs may also have markings on them caused by dirt or variations in material. Shadows and other surface markings on the roof cause similar problems.

There are other characteristics of these images which may specifically cause problems for contour following type systems. Roofs have raised borders which sometimes cast shadows on the roof. This results in multiple close parallel edges along the roof boundaries and often these edges are broken and disjoint. At roof corners and at junctions of two roofs, multiple lines meet leading to a number of corners making it difficult to choose a corner for tracking. A roof cast a shadow along its side and often there are objects on the ground such as grass, trees, trucks, pavement, etc., which lead to changes

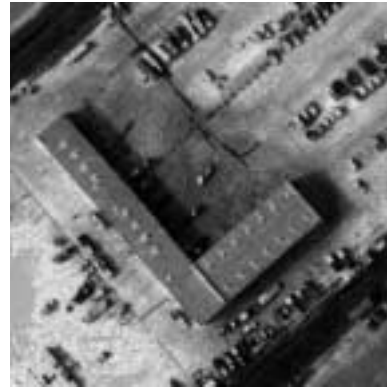


Figure 1A building from Ft. Hood, Texas

in the contrast along the roof sides. Thus while tracking one can face reversal in edge direction. Often some structures both on the roof and on the ground are so near the roof that the border edges get merged with the edges of these objects, leading contour trackers off the roof onto the ground or inside the roof area. At junctions it is difficult to decide which path to take. Searching all paths at junctions leads to a combinatorial explosion of paths. It may be difficult to decide on the correct contours since contours may not close because of missing edge information, or more than one closed contour may be generated. Contours may merge roofs or roofs and parts of the ground. Figure [Ref: hood3-r-seg] illustrates some of these problems. also in figure [Ref: hood3-r-lin] .

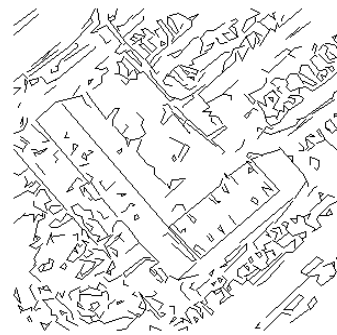


Figure 2 Line segments extracted from image

3.2 Lines and Junctions:

To deal with some of these problems, we use Mohan’s technique [Moha89] and consider a group of close parallel lines to represent the presence of a *linear structure* at a higher granularity level than the edges (see the common boundary between the building wings in Figure 4. The resulting *linears* have a length and an orientation derived from the contributing elements. Figure [Ref: hood3-r-lin] shows the lines obtained from grouping the segments in Figure 5. We then use these lines to detect L-junctions and T-junctions, shown .

leading to explanations that are primarily geometrical, for example in terms of transformational invariances [Palmer83], or probability [Lowe85,Witten83], or in terms of mechanisms such as orientation selection [Zucker83]. For groupings at the level of objects we have preferred a more functional explanation based on the *significance* (identification of structures in a image that have high likelihood of corresponding to object structures, i.e. focus of attention), *representation* (usefulness to other visual processes) of grouped features, and *selection* among multiple groupings (greater saliency and support).

Significance Issues

Our interest here is in building structures, most of which exhibit a great deal of geometric regularity. The principle of *non-accidentalness* [Lowe85,Witten83] states that “regular geometric relationships are so unlikely to arise by accident that when detected, they almost certainly reflect some underlying causal relationship.” The probability of two lines, which are not parallel in 3-D, projecting in the image as parallel lines in 2-D (due to an accidental alignment), can be considered so small that we can with high confidence claim that parallel lines detected in 2-D are also parallel lines in 3-D.

We have designed our systems to favor and be sensitive to the shapes and structure of the objects they were designed to detect. We detect viewpoint-invariant structural relationships that are common in the objects of interest and use the non-accidentalness principle to reason that the detected groups were caused by the structure of the objects in the scene. We choose groups that identify structural arrangements of visual elements that have a high probability of corresponding to object of interest.

Representation Issues

Grouped elements must encode the geometric and structural criteria that led to its formation. They basically represent higher levels of abstraction but must allow recursive recovery down to the primitive image elements. The chosen representations for groups should clearly allow the system to perform segmentation, description, focus of attention and integration of evidence at all levels:

- *Segmentation*: Relationships among substructures may also exist among structures and superstructures.
- *Description*: Most object shapes are described in terms of component shapes. The hierarchical decompositions should be easily identified.
- *Focus of Attention*: By encoding appropriate and relevant grouping information, the groups should provide guidance and contextual cues.
- *Integration*: As much as possible groups represent relationships that are invariant to viewpoint,

and thus can be used to integrate information from multiple views, site models, multiple sensors and multiple media. In correspondence processes such as stereo, motion, and model matching, improved performance has been and can be obtained by using more abstract features [MedNev85,Moh89,Lim87,GazMed89]. This also results in a significant reduction in the computational expense of matching. Perceptual groups are more complex than edges, thus there is much less ambiguity in matching higher level features.

Selection Issues

The inherent complexity of aerial scenes in terms of detail and in terms of the possible combinations leads to a large number of possible groupings. In practical terms a computer vision system must include a set of parameters that keep the data bases within a reasonable size. On the other hand there is the issue of fragmented and incomplete low level information due to image content, quality, resolution, etc. This requires that missing features be hypothesized from partial information, which must compete with less relevant groups. In section 4 below we propose a novel selection scheme based on the evaluation of the competition among hypotheses as a function of their structural properties and the underlying support.

3 Hypotheses Generation

Photointerpretation tasks require the detection of a particular set of objects in an aerial scene: buildings [FuaHan87,HerKan86,Huer-Neva88,Mathwa85,Moha-Neva89c,Venk91], roadways and storage tanks [Hue83,LanBro78], runways and taxiways [Huer-etal89,Huer-etal90], airplanes [Broo83,Stei-Medi90], and docks and ships [Neva-etal90]. In these scenes, many of the objects have restricted shapes, the viewpoint is often restricted, and many of the object shapes can be characterized as compositions of small sets of basic shapes.

Given this set of basic shapes, we can deduce the pertinent structural relationships for the domain by decomposing the basic shapes into the component structural relationships. As an image is a 2-D projection of a 3-D scene, we need to select those structural relationships (to characterize the shapes) that project invariantly over various viewpoints. If for a domain, the set of viewpoints is restricted, as is the case for most aerial images, we need to consider transformation invariance only for the restricted set of viewpoints.

This set of shapes is not specific in the sense of having a particular model for each object (e.g. the model of a Boeing-747 [Broo81b]), rather the object shapes are arbitrary compositions of known, specific, basic shapes. The methodology that we have used (see [Moha89] for details) can, by simple extensions or modifications, be applied to

able. We present results for various buildings from real and modelboard scenes provided to us by DARPA as part of the RADIUS effort.

2 Our Approach

There is no doubt that people are able to group sets of simple features, such as dots and lines, very quickly. The human process of perceptual grouping is not fully understood but the grouping criteria are believed to include proximity, collinearity and symmetry of the features and formation of closed figures. This process is usually referred to as *perceptual grouping*, and we argue that it is of crucial importance in the process of aerial image analysis. This is due to the large amount of detail in the images and the complexity of the structures in them. Perceptual grouping allows us to separate the meaningful features from the others; in this task we are further aided by the knowledge of the kinds of structures we are interested in. For example, for buildings we expect to find regular geometric structures.

Following [Moha89,Moha-Neva89c] we continue to think of the problem of perceptual grouping as consisting of three related subproblems. The first is to determine how to represent the visual features. The second is to determine what kinds of grouping operations to apply to obtain meaningful groups. The third is that there are many possible groupings and we must be able to choose among them.

Points and curves are useful representations. Points denote position but can have other attributes such as size that denotes extent of the feature. We now consider in addition, “oriented” points, when these represent the intersection of lines (i.e. junctions and their attitude). Straight curves (i.e. line segments) denote visual boundaries as well as directionality. They can also denote symmetries.

For selection among various possible groupings, we can use multiple criteria. For example, if a curve forms symmetric pairs with two other curves, we can choose the one that gives a more closed or compact figure. Essentially, we can allow the groupings to compete and cooperate depending on whether they are mutually exclusive or supportive. Mohan and Nevatia [Moha89,Moha-Neva89c] have used such a scheme to successfully detect groups of line segments that might correspond to roofs of buildings in aerial images, and to segment scenes of complex man-made objects without specific knowledge of the objects in the scene.

2.1 Perceptual Grouping

In the remainder of this section we summarize our approach to perceptual grouping. For a more detailed discussion see [Moha89].

We believe that the grouping operations

that yield organized perception consist mainly of symbolic evaluation of the properties of the features that are candidates for grouping. By features we mean representations of visual elements; by grouped features or feature groupings we mean geometrically and structurally significant groupings along a hierarchy of recursively formed groups, from points to complex objects. By object detection we mean two things: the description of the shape of the objects and the description of their structure. The shape descriptions should be at various scales using features that are invariant to changes in the viewpoint.

The construction of meaningful feature groupings in our systems is determined by the following issues:

2.1.1 Similarity Issues

It appears that humans prefer to group elements that have similar characteristics such as shape, intensity and color [Palmer83,Tries82]. We do not explicitly use color or intensity but, as mentioned above, we use two primitives, points and curves, to represent position, shape and attitude, and thus the similarity is among these simple features.

2.1.2 Structure and Scale Issues

The relationships among the points and lines convey structure and are thus a criteria for grouping. Strong relationships induced by the domain, such as those between structures and their shadows, also constitute as strong criteria for grouping. Given the importance of structural information in visual processing [WitTen83], this has been the most studied component of perceptual organization in computer vision and psychology [Lowe85,LowBin83,Moha-Neva89b,Moha-Neva89c,Palmer83,Rey87,-Stev81,Zucker83]. Some grouping processes use only structural relationships, and others are hierarchical. We can separate the elements into distinct groupings even though they do not correspond to any objects we recognize.

Structural groupings can be further subdivided on the basis of scale. At small scales the structural relationships form locally, at the level of subparts or parts of objects such as in regular textures (see [Zucker83]). Although the structural grouping process outlined in our previous work can detect such groupings as well, in the work described here we concentrate on groupings at the scale of the objects in a scene, such as buildings or part of buildings.

2.2 What to Group?

The Gestalt psychologists believed in the principle of *Prägnanz* (goodness or simplicity of form) as a fundamental criterion to group elements. Recent work has tried to develop computational criteria

ry in the image understanding domain has focused on grouping of dots and lines [Blos-Ahuj89,Fis-Bol86,Kell-etal77,Lowe85,LowBin83,Stev81,Wit-Ten83,Zucker83]. Perceptual organization ideas, however, are difficult to formalize and several authors including many in our group at USC, are working on the computational aspects of perceptual groupings. Informal derivations are hard to implement in computer vision systems due to the difficulty in detecting grouping relationships, due to our lack of understanding of suitable representations, and the processes that can make use of the established relationships in higher levels of perception [Moha89].

One of the pioneering efforts on grouping features in real scenes was done in our group [Moha89,Moha-Neva89b,Moha-Neva89c]. This system has been applied to building detection as well as object level segmentation from monocular images of complex indoor scenes. They do not claim that this is a computational theory for perceptual grouping and we are not aware of such theory in the literature. However they claimed that the use of perceptual organization principles is perhaps almost a must for aerial image analysis, and in particular, in the detection and description of cultural features such as buildings (see [Moha89]). We continue to believe that there are two main reasons for this. First, perceptual organization makes explicit the geometric relationships among *perceived features* (abstract internal representations of actual features in the visual field). Second, it provides *focus of attention*, that is, a collection of techniques designed to draw attention to significant structures in the image. Note that these two major visual abilities are applied recursively at all levels of perception. We should also point out that at the lower levels of vision we prefer to think of groupings for non-purposive perception (geometric structure without functional attributes) while at the higher levels we like to think of groupings for purposive perception (geometric structure with functional attributes).

Fua and Hanson [FuaHan87] segment the scene into regions, find edges lying on region boundaries and then see if there is evidence of geometric structure among these edges to classify the region as a building or similar object. In the VISIONS system [HanRis78], region segmentation is the primary technique used and the regions are classified by their shape and spectral properties. SPAM [McKe-etal85] is a map based system which uses region segmentation of aerial imagery. Venkateswar [Venk91] uses line segments extracted from the image to construct a data base of lines. A set of heuristic rules operates over the data base and hierarchically composes the line segments into buildings. This method relies on fairly complete and accurate line segments to detect junctions and form hypotheses in a manner similar to [Huer-Neva88]. Shadow correspondences are also used to verify buildings as in [Huer-Neva88].

For the systems mentioned above, the generic feature extraction techniques, namely region segmentation or contour following, are not suited for extracting particular shapes or organizations. If the features being detected have simple geometric properties, it is more straightforward to use specific detection algorithms. The Hough transform [Bal86,BalBro82] is a general mechanism for detecting groupings, but is practical only if the exact shapes (rather than generic descriptions) are known. The MOSAIC system [HerKan86] uses oriented junctions to complete fragmented lines. This system also uses height information obtained from stereo and sophisticated geometrical reasoning to hypothesize likely wire frame models of the buildings. The complexity of this system, and its limited performance, are due to the use of simple features (lines and junctions) to perform the detection, stereo matching and reasoning. Recently, application of perceptual grouping to locate features indicating structure has been explored by Reynolds and Beveridge [Rey87]. This systems also employs specific routines to detect various geometric organizations indicative of structure. However, this system has limited use as the groupings are sensitive to the layout of the scene rather than the object shapes, and consequently can not be used to either detect or describe any individual structures (like buildings) in the scene.

Most of these systems work on simple scenes, for example rural scenes, where the building roof can be simply segmented (and even identified) from the background on spectral properties. The detected buildings have simple shapes and no design details that may generate multiple edges at building boundaries. Only a few systems compute and use depth information. None of the systems generate a description of the buildings at the level of shape descriptions of the different wings.

In this paper we discuss the our more recent work on 2-D and 3-D analysis of aerial images of real scenes. The 2-D work deals mostly with shape issues and is applied to monocular images, and the 3-D analysis deals with objects descriptions in 3-D using shadows to obtain 3-D information. Our relevant past work, which applies mostly to the detection and description of cultural features from monocular and stereo images is given in [Huer83,Huer-Neva88,Moha-Neva89c,Moha89]. We first give a summary of our hypotheses generation process, mostly based on previous work by Mohan and Nevatia. We have added the use of shadow cues to aid in this process. Next we describe a new and more simple and robust mechanism than the one described in [Moha-Neva89c], to select promising hypotheses, which is also aided by strong shadow clues when these are available. The next section describes a building verification mechanism based on shadows (using ideas from our earlier work described in [Huer83,Huer-Neva88]) that also gives 3-D information when stereo views are not avail-

Detection of Buildings from Monocular Views of Aerial Scenes using Perceptual Grouping and Shadows

A. Huertas, C. Lin and R. Nevatia*

Institute for Robotics and Intelligent Systems
University of Southern California
Los Angeles, California 90089-0273

Abstract

Mapping, cartography, photointerpretation and guidance are some applications that can directly and readily benefit from automated aerial scene analysis. We have successfully used perceptual organization ideas and shadow cues to analyze monocular and stereo images of aerial scenes and describe cultural features of interest, such as buildings. Perceptual organization refers to the ability of a visual system to quickly capture representations of structure and similarity among otherwise random elements, features, and patterns in the visual field. We present a system that combines these ideas with the use of shadow clues to aid the hypotheses generation process, to verify buildings, and to generate 3-D descriptions of buildings from monocular views. This paper is based on, and extends the work of different people in our research group with more details of some issues found in the referenced papers.

1 Introduction

The goal of this work is to detect and describe three-dimensional structures such as buildings from monocular views of aerial scenes. This task requires robust segmentation techniques and methods to infer the 3-D structure. Traditional methods rely on edges or regions extracted from the image. Edge-based techniques attempt to collect linked edge curves into the desired object boundaries and succeed only for relatively simple scenes. Region based techniques construct closed curves that often do not correspond to the objects of interest. Model-based techniques (for a survey, see [Binf82]) can deal with fragmentation but require a-priori knowledge of the objects in the scene. Actually, they require very specific shape models. For example, it is

not sufficient to say that the building boundary is a rectangular parallelepiped; you must also supply the relative dimensions of the sides.

Perceptual grouping has been the basic approach for much of our work on detecting buildings and other structures in aerial images. While a wide variety of techniques have been applied towards this task, a systematic use of perceptual grouping has been lacking. Another observation is that while non-natural objects have rich geometric structure, little use of this structural information was made in the older systems. Cultural features such as buildings represent structures that are not random but have specific geometric properties. We rely primarily on these properties to form hypotheses. On sunny days and under favorable imaging conditions 3-D structures cast shadows and exhibit other features that allow inference of the 3-D structure from 2-D images. We have previously used techniques of *perceptual grouping*, *shadow analysis* and *shape from contour* somewhat independently for utilizing these observations. In this paper we describe the integration of these techniques and we discuss some of the relevant issues of perceptual grouping and shadow analysis to aid the hypotheses generation process. Shadow analysis is discussed below in connection with the process of generation and selection of promising hypotheses, and with the verification of buildings and generation of 3-D descriptions.

Perceptual grouping at all levels of a hierarchical process of perception yield organized perception. In general, perceptual organization refers to the ability of a visual system to quickly capture visual representations of structure and similarity among otherwise random elements, features, and patterns. These representations are the result of grouping operations that give the system or individual a sense of the objects in the visual field. Originally, perceptual grouping was studied by Gestalt psychologists in the 1920s and 30s. Unfortunately, while they provided many useful insights into the problem and many compelling demonstrations, they did not provide a computational theory.

Much of the work on a computational theo-

* This research was supported by the Defense Advanced Research Projects Agency under contract F49620-90-C-0078, monitored by the Air Force Office of Scientific Research. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.