Technical Report CS80003-R

CONTEXTUAL BOUNDARY FORMATION BY ONE-DIMENSIONAL EDGE DETECTION AND SCAN LINE MATCHING

by

R.W. Ehrich

F.H. Schroeder

Department of Computer Science
Virginia Polytechnic Institute and
State University
Blacksburg, Virginia 24061

June, 1980

The work reported in this paper was supported by the National Science Foundation under Grant ENG-7712105.

for more intelligent global edge constructors to make sense out of local events. The approach in this paper is one-dimensional, and there are three important reasons for this. First, the most important interpretations of an edge are a consequence of its behavior in a direction normal to its spatial orientation; discussion of two-dimensional behavior complicates what is already an enormously intricate structure. Second, the representation of an edge by the relational trees of its crossectional slices is both simple and useful. Third, many existing edge detectors are not, in fact, as two-dimensional as they first appear.

The most common technique for producing an edge image from a gray level image is to apply to the image a particular edge detector and then to do some sort of locally adaptive thresholding on the detector output. Milgram [2] gives an excellent discussion of the pitfalls of thresholding techniques and has proposed a boundary growing technique that produces region boundaries by examining the derivative image at multiple thresholds. In the technique described here there is no need to preselect thresholds since the various edge interpretations produced by the edge detector are self-determining. Second, it is argued that only first order difference approximations to the first derivative are viable for accurate, high resolution edge detection. Indeed, differences of 4 x 4 averages would not be suitable for the class of images with which we have been working.

Many edge detectors in common use, though two-dimensional in concept and design, have implementations that are nearly one-dimensional. An edge detector is said to be directionally separable if its output is a function $f(e_1,\ldots,e_n)$ of the responses $e_1,\ldots,e_n$ of a set of directionally selective detectors that test edge hypotheses in directions $\theta_1,\ldots,\theta_n$. An edge detector is one-dimensional if its output is computed from a one-dimensional intensity profile across an image, and it is called quasi one-dimensional if its output is computed from a linear combination of parallel intensity profiles. To give a few examples, the Hueckel operator [3] is an example of a non-separable edge detector, whereas the Roberts [4], Sobel [5], Kirsch [6], and Rosenfeld difference of averages [7] operators are examples of directionally separable edge detectors. For these four, the component detectors are one-dimensional in the case of the Roberts Cross, quasi one-dimensional in the case of the Rosenfeld and Sobel operators, and two-dimensional for the Kirsch.

Because many of the commonly used edge detectors are directionally separable and constructed from components that are at least quasi one-dimensional, it is the purpose of this paper to reexamine in more detail the semantics of one-dimensional edge profiles. There are a number of consequences of this view of the edge detection problem that are quite different from established views. On the other

hand, it will be possible to see quite clearly why such a simple detector as the Roberts Cross performs as well as it does and why non-maximum suppression is so successful in thinning edges. Also, it should become clear why for any moderately complicated scene, no particular local edge detector with given parameters can do an adequate job for all images.

One of the most important assumptions in this paper is that the images under consideration are noise free and that the extreme complexity of a picture function is due entirely to tonal, reflectivity, illumination, and curvature variations in the three-dimensional scene being analyzed. It is, for example, quite clear that sensor noise is not the source of the great difficulty of general computer vision or robotics. In the same spirit, if there are dust specks on the light table used for digitizing an industrial scene, it is assumed that these are to be removed after edge detection instead of requiring the edge detector to remove them directly by treating them as noise.

2 - Edge Regions

If one carefully considers a natural or industrial scene it is clear that edges are, by definition, regions of intensity change rather than locations at which image intensities change abruptly. In some applications it is

necessary to know only a representative point for an edge element that can be used to form a smooth region boundary. In others, such as determining the curvature of an object, it is necessary to know the width of the edge region or possibly even the exact shape of the entire edge region. In any case, many edges, such as the one whose profile is shown

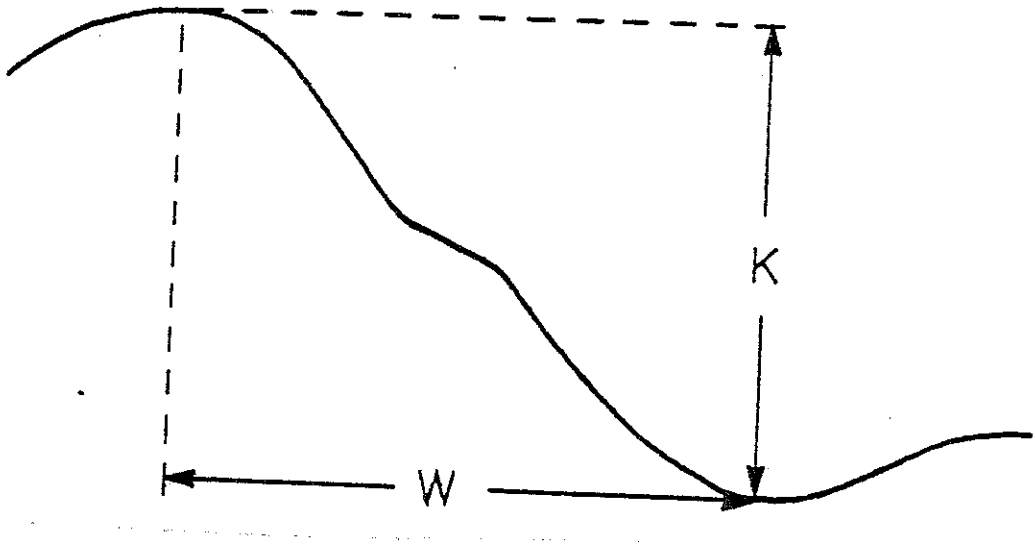


Figure 1 - Ambiguous edge.

in Figure 1, are ambiguous. Given such a profile, one can always find an image in which the entire slope region must be interpreted globally as one large high contrast edge. In another image the very same profile may have to be interpreted on the basis of its context as two adjacent low contrast edges, separated by a narrow plateau. Consequently, there is no way that an edge detector that produces the single edge interpretation can perform properly

on the second image, and so forth. In fact, neither the thresholded Kirsch nor the thresholded Sobel operators are even capable of resolving the double edge in the center of this shadow plateau. However, this is a relatively unimportant point compared with the problem of the commitment such a detector has to one interpretation or another. The only alternative is to construct an edge detector that is smart enough to report both interpretations so that the proper one may be chosen when more information is available. Before leaving Figure 1, notice that the edge contrast has little to do with local slope but is a global property of the entire edge region.

Unless one is analyzing very restricted scenes, it is important that an edge detector report only what is happening in an image rather than an interpretation of the data. This is an instance of Marr's Principle of Least Commitment [8] which states roughly that whenever there is insufficient evidence for making a decision, the decision is deferred. Since the goal of edge detection is usually boundary formation or segmentation, the interpretation of edge profiles should be deferred until the time of the boundary formation process itself.

Least commitment is the key to understanding the deficiencies of fixed size local edge operators. Most such operators are matched filters which measure the

crosscorrelation between an image and a filter mask. In such an operator the edge model is embedded in the filter mask, and the detector output is a measure of the degree to which the data can be interpreted to be identical with the shape of the filter mask. Marr also recognized the problems associated with the use of a fixed size edge operator; instead he used a set of filter masks with varying sizes to determine an interpretation for the intensity variation. Still, Marr would select one of a small number of interpretations for each edge, and to some extent he appears to violate his own principle. His use of multiple filter masks is similar to the work of Rosenfeld and Thurston [7], except that they retained all filter outputs at each image point.

The information required to be reported by an edge detector is determined in large part by the application. If one wishes to deduce the shape of an object in a scene, a functional approximation of a slope region might be appropriate. Other applications require much less information, and the edge features considered most important in this paper are the contrast and the locations of the maximum slope point, full edge region boundaries, and high derivative region boundaries.

In order to motivate the presentation of the edge detection algorithm proposed in this paper, let us consider

some image data. The house scene in Figure 2 is a moderately complex image that contains both gray level edges and texture region boundaries. This image will be used for demonstrations throughout the remainder of the paper. Notice, in particular, the small circled region above the left window. A 25 point segment of a vertical profile through this region oriented from the top toward the bottom is shown in Figure 3a. The proper interpretation of this profile would be two parallel edges, separated by the light shadow region above the window. The shadow boundary is not the important edge, however, although that is obvious only because we can deduce it from other facts about the image.

Figure 2 - House scene.

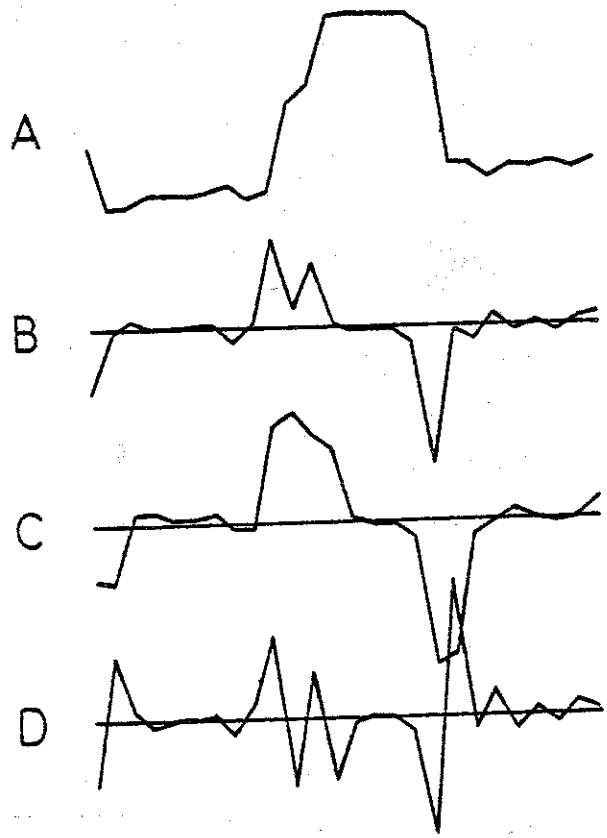


Figure 3 - Vertical profile from part of Figure 2 (a), First difference (b), Central difference (c), Laplacian (d).

Clearly the main function of an edge detector is detection of the presence of local slope, and Figures 3b and 3c show the first forward difference and central difference approximations to the first derivative, respectively. The plateau shows dramatically on Figure 3b and is totally missed on Figure 3c because it is only one picture element wide. Consequently it is necessary to use first forward differences if we are to resolve plateaus that are this narrow. Curiously, the edge detector with which most others have been compared over the years, the Roberts Cross, is made from two detectors that difference the intensities of

diagonally adjacent picture elements. Typically one would threshold the output of the Roberts Cross at some intensity whose value is just above the valley between the two adjacent peaks on Figure 3b, and the two edge elements would be resolved. The problem, of course, is that there is no way of knowing in advance what threshold to use for this particular edge.

Looking again at Figure 3b it is obvious that the overall edge regions should be bounded by the intersection points of the derivative peaks and a line parallel to the horizontal axis at a height equal to the detector threshold. In order to bound the high slope region one might use the extrema of the second derivative shown in Figure 3d. Notice that there are two positive-negative peak pairs - one pair for each of the two adjacent edges. It was thought at first that the centers of the high slope regions would be excellent places to mark in order to produce thin and visually smooth edge boundaries. In all our experiments with various definitions of the high slope region, we were never able to make that idea work. The best place to mark is invariably the maximum slope point. This explains why the Roberts Cross produces such visually pleasing results, and it also justifies non-maximum suppression procedures for thinning edges. With that motivation, the next section defines precisely the edge detector that is being proposed.

An edge region is defined as a maximal set of consecutive points in an intensity profile whose slopes exceed a threshold T. Let D_1 be the derivative operator that differences the intensities of adjacent picture elements. If an intensity profile is given by

$$X = x_1, x_2, \ldots, x_{n-1}, x_n$$

then
$$D_1 X = x_2 - x_1, x_2 - x_1, x_3 - x_2, \ldots, x_n - x_{n-1}$$
$$= \dot{x}_1, \dot{x}_2, \ldots, \dot{x}_n$$

are the first forward differences of X. Since x_0 is unavailable for computing $\dot{x}_1$, $\dot{x}_1$ is simply set equal to $\dot{x}_2$. Then

$x_j, x_{j+1}, \ldots, x_{k-1}, x_k$ is a wide edge region
iff $\dot{x}_{j+1}, \dot{x}_{j+2}, \ldots, \dot{x}_{k-2}, \dot{x}_{k-1} \geq T$ and
$\dot{x}_j < T$ and $\dot{x}_k < T$.

The contrast of an edge is the signed difference between the extrema of X over the wide edge region. Let

$$x_\ell = \max \{x_j, \ldots, x_k\}, \quad j \leq \ell \leq k$$
and let $$x_m = \min \{x_j, \ldots, x_k\}, \quad j \leq m \leq k.$$
Then the contrast, K, is given by

$$K = x_\ell - x_m \quad \text{if } m < \ell$$
$$= x_m - x_\ell \quad \text{if } m > \ell$$
$$= 0 \quad \text{otherwise.}$$

The contrast is defined in this way rather than by $x_{k-1} - x_{j+1}$ so that the definition still works properly when another difference operator is substituted for D_1. The high derivative edge region is determined from the second derivative of the intensity profile X, and it is referred to as the narrow edge region. Let D_2 be the Laplacian operator defined by

$$D_2(x_i) = x_{i-1} + 2x_i + x_{i+1} .$$

Then

$$D_2X = x_1 - 2x_2 + x_3,\ x_1 - 2x_2 + x_3, \ldots,\ x_{n-2} - 2x_{n-1} + x_n$$
$$= \ddot{x}_1, \ddot{x}_2, \ldots, \ddot{x}_{n-1}, \ddot{x}_n .$$

Once again since x_0 and x_{n+1} are unavailable, $\ddot{x}_1$ is set equal to $\ddot{x}_2$ and $\ddot{x}_n$ is set equal to $\ddot{x}_{n-1}$. The extrema of D_2X are the points where the maximal changes are occurring in the slope of X, and as discussed previously, these points seem natural for bounding the narrow edge region. Within the wide edge region, D_2X frequently has more than two extrema, and there needs to be a consistent way to select the proper ones. The situation is shown in Figure 3 if T=1 so that the wide edge region consists of the combination of the two subpeaks. To do this, the maximum slope point is determined, and then D_2X is searched for extrema on either side of the maximum slope point. Since it may happen that D_1X has several maximum slope points within the wide edge region, the maximum slope point, s, is defined as follows.

Let $\dot{x}_i = \max \{\dot{x}_j, \ldots, \dot{x}_k\}$, $j \le i \le k$, if $K>0$

and $\dot{x}_i = \min \{\dot{x}_j, \ldots, \dot{x}_k\}$, $j \le i \le k$, if $K<0$

Then $s = (\ell+m)/2$ where

$$\ell = \{p \mid \dot{x}_p \in \{\dot{x}_j, \ldots, \dot{x}_k\},\ \dot{x}_p = \dot{x}_i,\ \text{and p is minimum}\}$$

and $$m = \{p \mid \dot{x}_p \in \{\dot{x}_j, \ldots, \dot{x}_k\},\ \dot{x}_p = \dot{x}_i,\ \text{and p is maximum}\}.$$

Thus s is determined by finding the extreme value of $D_1 X$ within the wide edge region and searching from both ends of the wide edge region for the first points that achieve that value. Next the narrow edge region is located by searching in either direction from the maximum slope point, s, for the first values of the second derivative that achieve the proper extreme values. First,

let $\ddot{x}_\ell = \max\{\ddot{x}_j, \ldots, \ddot{x}_s\}$

and $\ddot{x}_m = \min\{\ddot{x}_s, \ldots, \ddot{x}_k\}$ if $K>0$

or $\ddot{x}_\ell = \min\{\ddot{x}_j, \ldots, \ddot{x}_s\}$

and $\ddot{x}_m = \max\{\ddot{x}_s, \ldots, \ddot{x}_k\}$ if $K<0$.

Then the narrow edge region is given by

$x_r, \ldots, x_t$ where

$$r = \{p \mid \ddot{x}_p \in \{\ddot{x}_j, \ldots, \ddot{x}_s\},\ \ddot{x}_p = \ddot{x}_\ell,\ \text{and p is maximum}\}$$

and $$t = \{p \mid \ddot{x}_p \in \{\ddot{x}_s, \ldots, \ddot{x}_k\},\ \ddot{x}_p = \ddot{x}_\ell,\ \text{and p is minimum}\}.$$

At this point it still appears that there is a threshold, T, that must be selected before the algorithm above is run. This is a clear disadvantage since it fixes the detector's interpretation of a low slope region and causes exactly the problems we have sought to avoid. In the next section it is shown how T can be eliminated by allowing

the edge detector to generate alternative hypotheses about the edges.

4 - Edge Hypotheses and Representations

Consider again the ambiguous edge of Figure 3a which has been redrawn together with the first derivative in Figure 4. Let the polarity of an edge describe whether it is a white-black or a black-white transition when traversed from left to right. As long as the edge profile is monotonic with polarity P it is necessary to form an edge hypothesis for the region over which the profile is monotonic. If there are two adjacent monotonic regions of polarity P separated by a monotonic region with polarity -P, then one must hypothesize three edge regions and require a more intelligent edge analyzer to determine whether the central edge region was caused by an undesired disturbance or by an anticipated and significant part of the scene.

The derivative in Figure 4b shows that the monotonic edge region in Figure 4a has substantial structure. With threshold T as shown, only one edge region would be reported by the edge detector, and with T set at T′, two separate edge regions would be reported. The alternative of fixing the detector threshold at some high value so that multiple hypotheses are always produced is not viable because then high contrast, low slope edges would become undetectable. Therefore it is necessary to vary the threshold T to produce

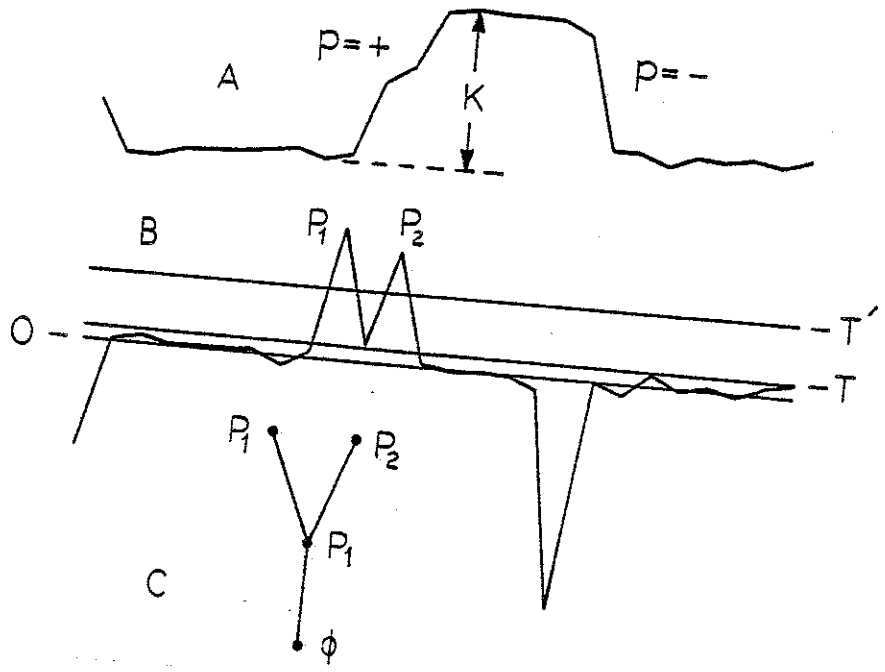


Figure 4 - Profile (a) and Derivative (b) from Figure 3 and Relational tree of derivative (c).

all possible hypotheses, and the tree data structure that is used to store these edge hypotheses is called a relational tree.

The relational tree of an edge can be easily motivated with the help of the example in Figure 4. Suppose T is set to some large value and then lowered in discrete steps to T=0. First peak P_1 will exceed the threshold and then P_2 will as well. As soon as T is lowered below the level of the valley V_1, peaks P_1 and P_2 merge to form a superpeak that is labeled by P_1 since it is the higher or dominant peak of the two that have merged. The merger of small peaks

to form larger and larger superpeaks is represented naturally by a tree such as the one shown in Figure 4c, and this tree is called a relational tree because it represents relations among relations. Each valley induces a node in the relational tree, and each node is labeled with the dominant descendant peak. In general, relational trees are computed by a multi-stack algorithm, and the full development is given in [9]. The principal difference is that in this application a separate relational tree is computed from the first derivative for each monotonic portion of the profiles of a scene.

Each node of a relational tree consists of a pointer to a list of descriptive attributes for the peak it represents. In this case the attribute list would contain the following information (See Figure 1).

1 - Coordinates of the peak, giving maximum slope point and value of maximum slope.
2 - Horizontal coordinates of the wide edge region, W, defined with T at the valley that induced the peak.
3 - Contrast, K, for the peak, measured over the wide edge region.
4 - Horizontal coordinates of the narrow edge region.

The following properties of the relational trees of edges show why they are so useful in representing alternative hypotheses.

Property 1 The attribute list for each edge hypothesis contains the descriptive information about edge attributes discussed earlier.

Property 2 As a relational tree is traversed upward from the root, one encounters finer and finer interpretations of an edge region.

Property 3 If one prunes a relational tree by removing all descendants of various nodes, the frontier of the pruned tree always represents adjacent, non-overlapping edge hypotheses. The description is complete in that all parts of the wide edge region of the root node are covered by the frontier nodes.

Property 4 Since most edge regions are very narrow, the relational trees of edges are usually very simple.

Property 5 The edge hypotheses represented by the nodes of the relational tree are determined by the depths of the valleys of the derivative, hence by the slopes connecting the high slope regions of the edge. For example, in Figure 5, there is no hypothesis P_2 P_3 because the slope between them is so low. Even if P_3 were higher than P_1, the structure of the relational tree would be the same. This property ensures that the hypotheses represented in the relational tree are also the best ones for structural

reasons. On the other hand, there is enough information in the relational tree to form a compound hypothesis like $P_2 P_3$, although the contrast will be slightly underestimated if the contrast is computed from information stored elsewhere in

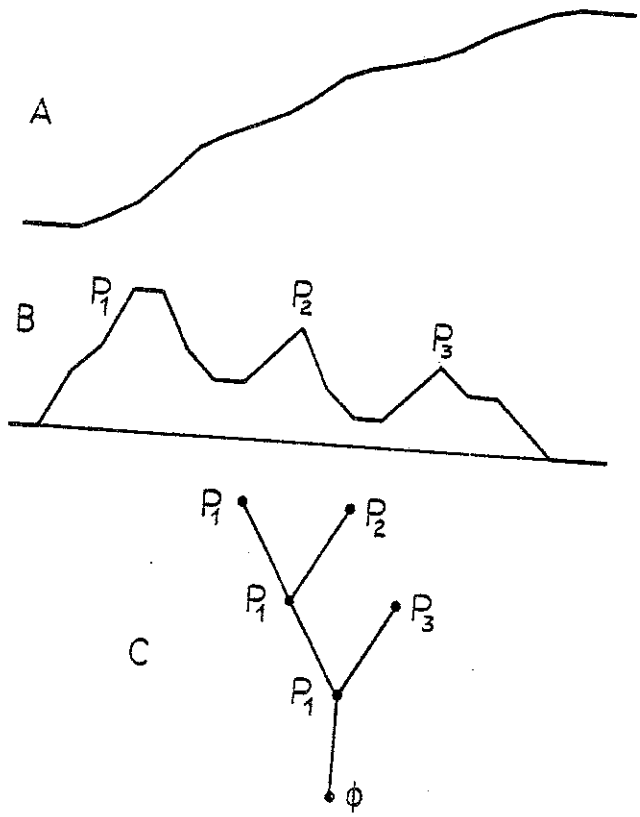


Figure 5 - Part of an intensity profile (a), First derivative (b), and Relational tree(c).

the tree.

In the preceding development, relational trees were used to represent only alternative hypotheses about edges that were monotonic slope regions. Figure 6 shows a situation commonly found in low slope shading edges. In Figure 6a there are three monotonic regions, and three relational

trees would be generated since the slope of the center region has opposite polarity. If the wide edge region for the center segment was sufficiently narrow and the slope sufficiently low, a relational tree for the combined regions could easily be generated as illustrated. While it seems more consistent to require the subsequent edge constructor to form compound hypotheses from non-monotonic seqments, the example shows how easy it is to have the edge detector do the work directly. Notice, however, that no explicit record of the center slope region is stored in the tree.

5 - Experimental Results

The following results were computed from the house scene in Figure 2. In viewing these results it is necessary to keep in mind both the sensitivity of the edge detector and the vast amount of detailed information it produces. The edge constructor, to be described next, produces linked edges and a cleaned image that is as good as any result we have have seen for this image. We have superimposed the results for both vertical and horizontal processing and have plotted only the high slope point of each edge. Wherever edges appear thickened, it is a consequence of detection of adjacent edges of alternating polarity.

Earlier in the paper it was mentioned that the high slope point was the best location for an edge. Several

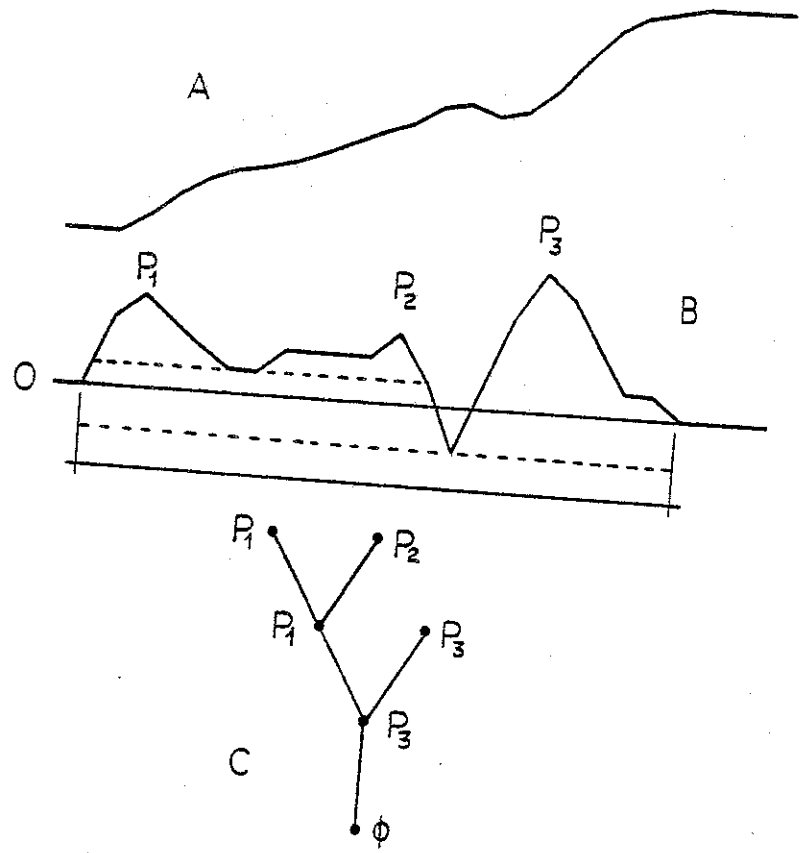
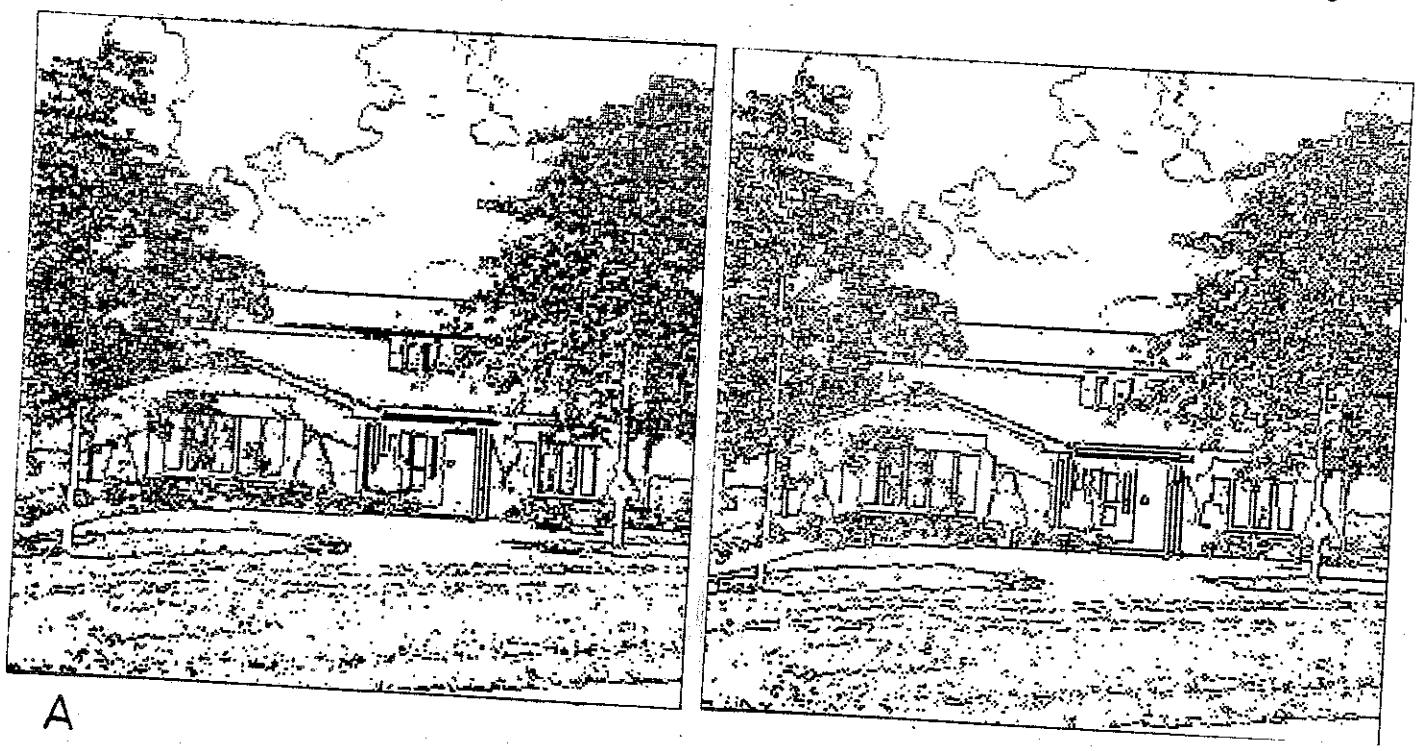


Figure 6 - Part of an intensity profile (a), First derivative (b), and Relational tree for merged edge regions.

unsuccessful attempts were made to define a narrow edge region that would be centered on a good location point. In one experiment the wide edge region was searched from both ends for extrema of the second derivative, and the edge location point was taken to be the center of this region. Curiously, the high slope point was invariably within this region, but rarely at its center. The significant point is that we were unable to improve on the use of the high slope point as an edge locator. This is consistent with Marr's

later experimental results [10] in which edges were located at the zero crossings of the second derivative.

First, let us compare the results of substituting a central difference operator for the forward difference operator, D_1. Figure 7 shows all edge points with contrast K>30 and T=1. While at first glance the central difference detector appears to produce a clearer image, this is because many significant edge points have been missed, such as the multiple edges on the roof line and on several of the windows. In Figure 8 are the detector responses with K>30



A B

Figure 7 - K>30, T=1, Forward differences (a), and Central differences (b).

using T=20. Again, many edges have been missed by the central difference detector, although edge location seems to

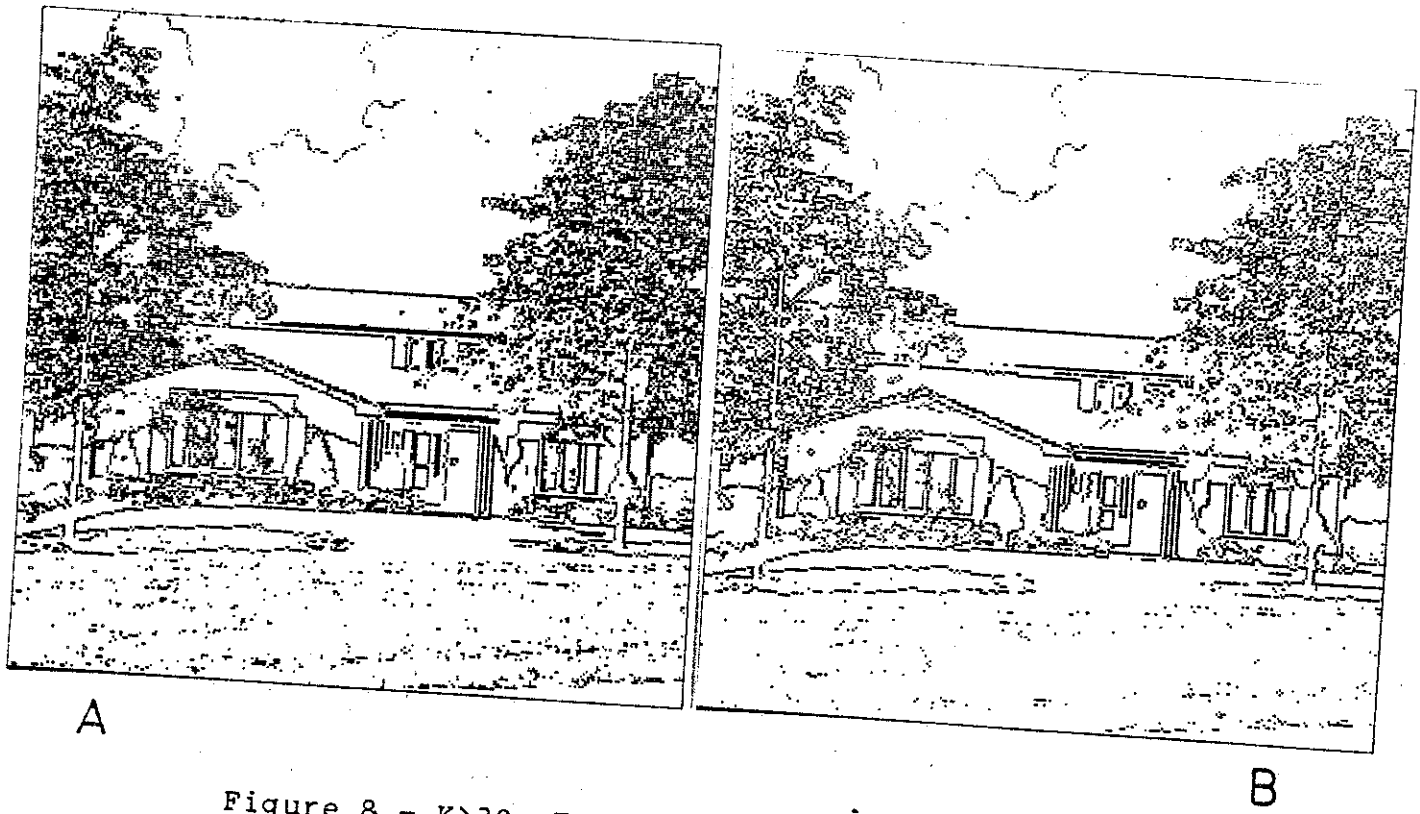


Figure 8 - K>30, T=20, Forward differences (a), and Central differences (b).

be slightly better. In the area above the left hand windows the shadow boundary cannot be completely detected by the central difference detector. A very careful comparison of Figures 7 and 8 reveal that although far fewer edges were detected with T=20 than with T=1, in many cases edges appear in Figure 8 where they were absent in Figure 7. The reason can be seen in Figure 4 -- for a high threshold both P_1 and P_2 are produced, while for a low threshold, only P_1 is produced. Finally, notice that the texture boundaries between sky and clouds are not well defined. It is firmly believed that special techniques should be used to detect

such boundaries and that general edge detectors should not be used here.

6 - Global Edge Linking

In this section the problem of constructing global edges from local hypotheses is discussed. The need for algorithms for producing line drawings from gray level images was recognized in the early years of computer vision research, and a number were constructed. Most, like the Binford-Horn algorithms [1], were designed to function in the blocks world, and they had built into them a substantial amount of knowledge of this restricted domain. Most algorithms are multi-step algorithms consisting of an edge detection phase followed by one or more tracking or linking phases. Because scene-dependent information could be used to correct the output of the edge detector whenever the interpretations it produced were erroneous, good results could be obtained without too much concern about the behavior of the edge detector. However, the analysis of natural scenes stimulated the development of numerous new edge detectors because the more difficult problem domain required the best possible edge information that could be obtained.

The algorithm investigated here is a scan line matching procedure that links edges across an image in parallel. The

matching procedure optimizes the linking in ambiguous situations while forming as many contiguous links as possible. The procedure depends upon an edge detector, such as the one described earlier, that produces edge assertions whose interpretations are left open until they are fixed by the linking procedure. Thus the linker does not correct erroneous edge interpretations but rather forms the interpretations on the basis of contextual information. Most important, edges are considered variable width regions rather than just points of high slope or high edge confidence; consequently, the scan line matching procedure may be considered to be a region grower.

Line construction algorithms generally fall into four categories - those that track individual edges, those that track multiple edges in parallel through adjacent scan lines, those that generate all parts of all edges in parallel, and those that generate edges as a by-product of a region growing process. Typical of the first approach are tracking algorithms such as those discussed in Rosenfeld [11], heuristic search methods such as Martelli's [12], and adaptations of dynamic programming algorithms such as that described by Montanari [13]. The Binford-Horn linefinder [1] is typical of the one-dimensional scan line approach; their procedure is run once in both the vertical and horizontal directions, and then the results are merged. Examples of parallel procedures include that of O'Gorman and

Clowes [14] who use the Hough transform to locate collinear points and that of Hanson and Riseman [15] who use relaxation to do simultaneous edge linking and enhancement.

Pavlidis has used a combination of the one-dimensional approach and region growing to generate boundaries by linking together portions of scan lines that have similar slopes [16,17]. His procedure is based upon the assumption that edges can be determined by approximating a scan line with straight line segments; edges would correspond to segments with particularly high slope. While this procedure yields good results on simple high contrast images using a very modest scan line matching algorithm, the procedure does not work well in complicated natural scenes. The difficulty is characteristic of the endemic problem of most edge construction algorithms: too little attention has been given to the extraction of information from the raw image data. As a consequence, the linking algorithms are frequently given incorrect data and have insufficient knowledge to recover from previous errors. Stated more directly, the data extraction phase that produces linker input frequently violates the Principle of Least Commitment.

In this particular problem there are interpretations of edge-like structures in an image that are completely impossible for an edge detector to make without more global contextual information. An edge detector must not produce

invariant interpretations of image data. Instead it must generate the most viable hypotheses about the nature of the image data and store it in a convenient representation. Then the edge constructor must evaluate the alternative interpretations based upon global context. In the case of the straight line approximation approach, edge interpretations are fixed at the time the approximation is computed, and there is no possibility of changing the interpretations without simply ignoring them.

It is inevitable that a system of the type being discussed here is going to have increased complexity, both in the edge detector which must form multiple hypotheses and in the linker which must evaluate numerous alternative hypotheses in context with another. One of our reasons for adopting a one-dimensional approach is to allow us to deal directly with the very complicated semantics of edges. Another reason is that meaningful interpretations of edges are determined principally by analyzing behavior along their normals, rather than in their principal directions. In fact, many popular edge detectors are essentially one-dimensional detectors whose outputs are combined so as to make them appear two-dimensional.

7 - Scan Line Matching

In this section a scan line matching algorithm is introduced that approaches the problems associated with boundary formation in a manner very different from earlier approaches such as those of Pavlidis. In its most general form, the matching algorithm would be a parallel tree matching algorithm that matches the relational trees of the edges in adjacent scan lines. The matching algorithm described here is more modest in the sense that while alternative edge hypotheses are evaluated, they are formed by the linker itself rather than taken from the relational trees. The edge detector that produces the simple edge hypotheses for the linker is a simplified version of the detector described above. All it does is detect those edge regions where the slope is greater than some minimum slope, S_o, and the wide edge regions for these simple hypotheses are simply the coordinates between which the slope exceeds S_o. Linking is achieved through a search procedure that determines the optimal linking arrangement based upon the strengths of the matches between the various edge hypotheses that are generated from the simple hypotheses put forth by the edge detector.

As discussed in the previous section, it is important for the linking procedure to have the ability to decide the interpretations of the edge regions in an intensity profile on the basis of context. This can be done by generating additional edge hypotheses called compound hypotheses from

the simple hypotheses formed by the edge detector whenever an edge region is sufficiently ambiguous. An example of where it is important to form a compound hypothesis is shown in Figure 9. Here, region A has two possible

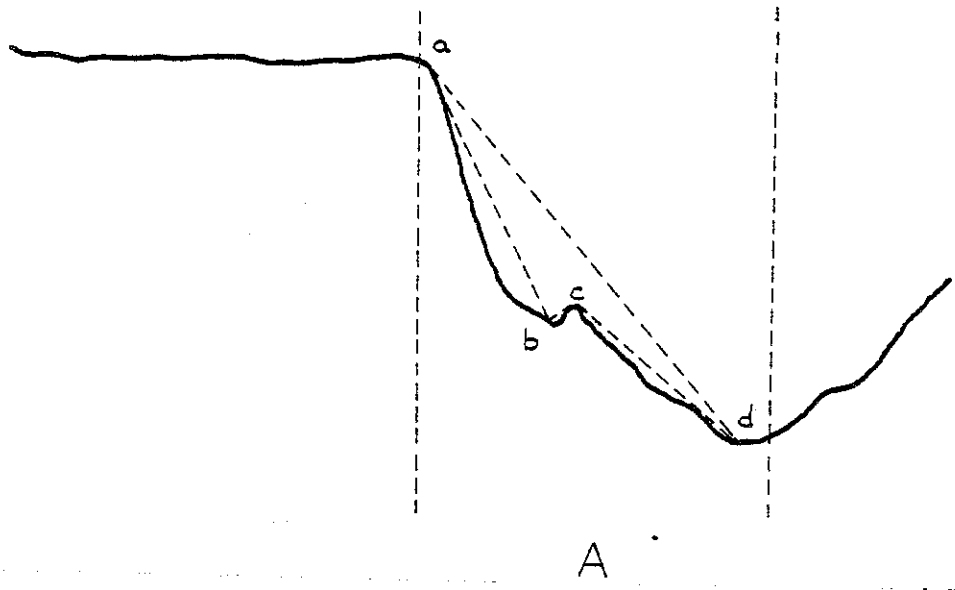


Figure 9 - Simple and compound hypotheses.

interpretations. In the first, the minor peak in the center of region A is considered to be a true peak, and the region is composed of the three simple hypotheses ab, bc, and cd. The second hypothesis is that the minor peak in the center of region A is caused by a very local event, and the region is correctly represented by the compound hypothesis ad. All of these edge hypotheses must be considered, and the three simple hypotheses and the compound hypothesis mutually constrain each other.

The linking procedure is based upon several observations about adjacent profiles.

a) Smooth edge boundaries in the image that are within 45 degrees of vertical have corresponding edge hypotheses in adjacent intensity profiles that are spatially aligned in the horizontal direction.

b) Homogeneous edge boundaries in an image have properties such as contrast, slope and spatial extent that vary slowly along their lengths.

c) An edge hypothesis in one intensity profile may have no match in an adjacent intensity profile for three possible reasons: other events have altered the continuity of the edge boundary, the end of an edge has been located, or the edge turns and becomes horizontal. The linking procedure searches through the competing edge hypotheses and generates an optimal linking arrangement between adjacent intensity profiles by considering the profiles in overlapping pairs.

After the edge detector produces the simple edge hypotheses, the linking algorithm has the following tasks:

1) Generate compound edge hypotheses where needed.
2) For each pair of adjacent intensity profiles, generate an edge pair table that gives the cost associated with each feasible edge link.
3) For each pair of adjacent intensity profiles, search for lowest cost consistent linking arrangement.

4) Postprocess the linking arrangement to increase the number of contiguous edge segments.

The most significant aspects of the edge formation process are:

a) Linking is a one-dimensional process.

b) The procedure is highly parallel, and the context across the entire image is used to link edge regions in the direction normal to the context.

c) Linking is achieved through a search that involves what is called the "strongest first" paradigm, and the procedure closely adheres to the principle of least commitment.

d) The output consists of linked directed edge segments rather than independent edge elements.

e) No a priori knowledge of the image is needed except that edges are smooth and continuous.

f) Links are not allowed to cross one another: edge intersections are resolved after linking using more global information.

g) Once an edge region has been determined it may be involved in only one link to each of its adjacent intensity profiles.

7.1 - Generation of Edge Hypotheses

The generation of edge hypotheses is a very important part of processing edge regions. The edge hypotheses

include the simple hypotheses that were produced by the edge detector and the compound edge hypotheses that are formed by the linker. The generation of compound edge hypotheses is based upon the spatial relationships between adjacent edge hypotheses. If the relative positioning of adjacent simple hypotheses implies that they could have originated from the same edge region, a compound edge hypothesis is formed that represents the combination of those two simple hypotheses. This merging procedure is based upon the differences of the end point values for the edge hypotheses as shown in Figures 10a and 10b. Let the two sets of end points $(X_{a1},Y_{a1}),(X_{a2},Y_{a2})$ and $(X_{b1},Y_{b1}),(X_{b2},Y_{b2})$ represent the edge hypotheses in both cases.

Then,

$$d_1 = |Y_{b1} - Y_{a2}|$$

$$d_2 = |X_{b1} - X_{a2}|$$

$$d_3 = |Y_{a2} - Y_{b2}|$$

The two edge hypotheses are joined together if and only if:

$$d_1 < D_1$$

$$d_2 < D_2$$

$$d_3 > D_3$$

for some threshold values, D_1, D_2, and D_3. If the two hypotheses are joined, the resulting edge hypothesis is defined by the end points $\{(X_{a1},Y_{a1}),(X_{b2},Y_{b2})\}$. In Figure 11, two adjacent intensity profiles from an image are shown together with the compound hypotheses generated using $D_1=25$, $D_2=3$, $D_3=2$.

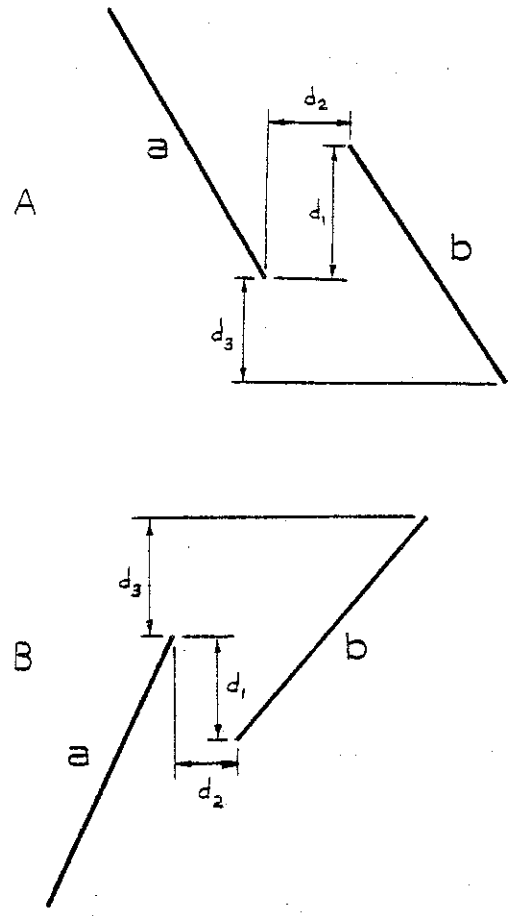


Figure 10 - Formation of compound edge hypotheses.

It can be seen that the parsing of intensity profiles into hypotheses that correspond to the most significant interpretations of the edge regions can be achieved by a relatively simple procedure. Unlike earlier methods, the edge construction algorithm presented here is not based upon approximation methods and is very sensitive to the structure of the intensity profile.

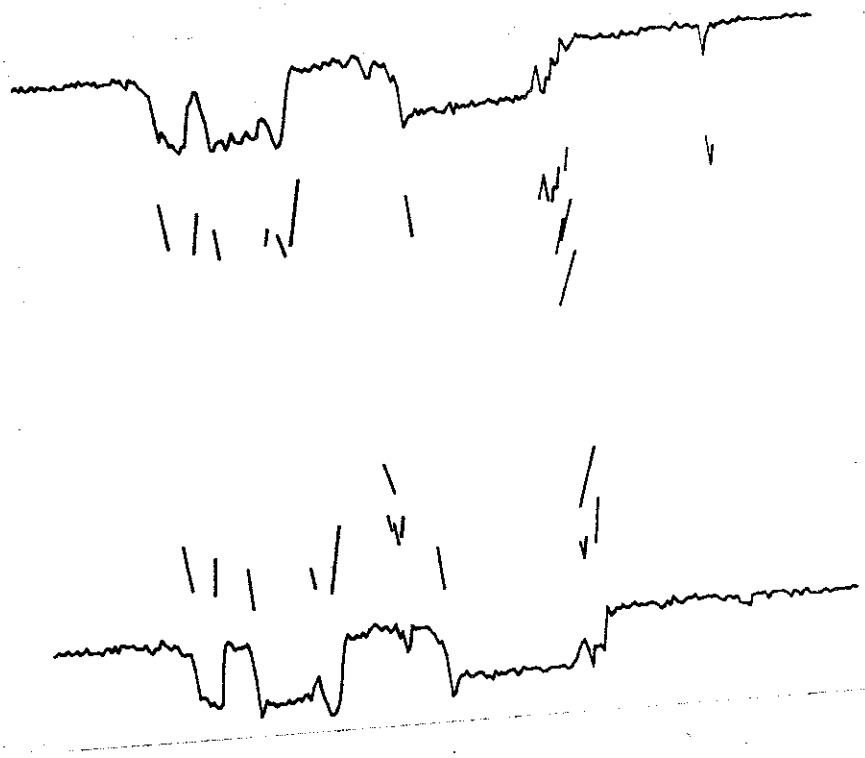


Figure 11 - Two adjacent profiles from an image and all simple and compound edge hypotheses.

7.2 - Generation of the Edge Pair Table

As we have seen, the edges of an image that are nearly vertical have edge regions in adjacent intensity profiles that have similar slopes and spatial translation. There should then exist a corresponding pair of edge hypotheses that are similar. If there is more than one edge hypothesis for an edge region, a decision must be made as to which one should be involved in a link to an adjacent intensity profile.

A simple method for making this decision is to link the most similar pair of edge hypotheses. As seen in Figure 12, each edge region in two adjacent intensity profiles is represented by three hypotheses a,b,c and d,e,f, respectively, and there are eight possible links ad, ae, af, be, bf, cd, ce, and cf. On the basis of similarity alone

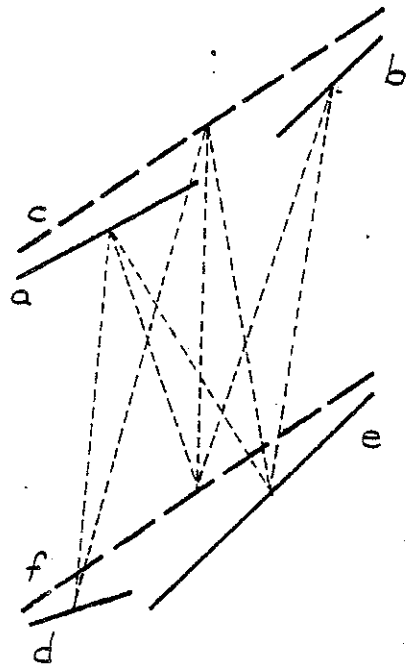


Figure 12 - Edge hypotheses in adjacent intensity profiles.

the link cf would be formed. Unfortunately, this simple linking procedure is not always correct. Figure 13 shows two adjacent intensity profiles across an edge boundary oriented at 45 degrees. It is obvious that the correct set of links is ad, be, and cf, but because of the minor differences in the intensities, the first link to be formed

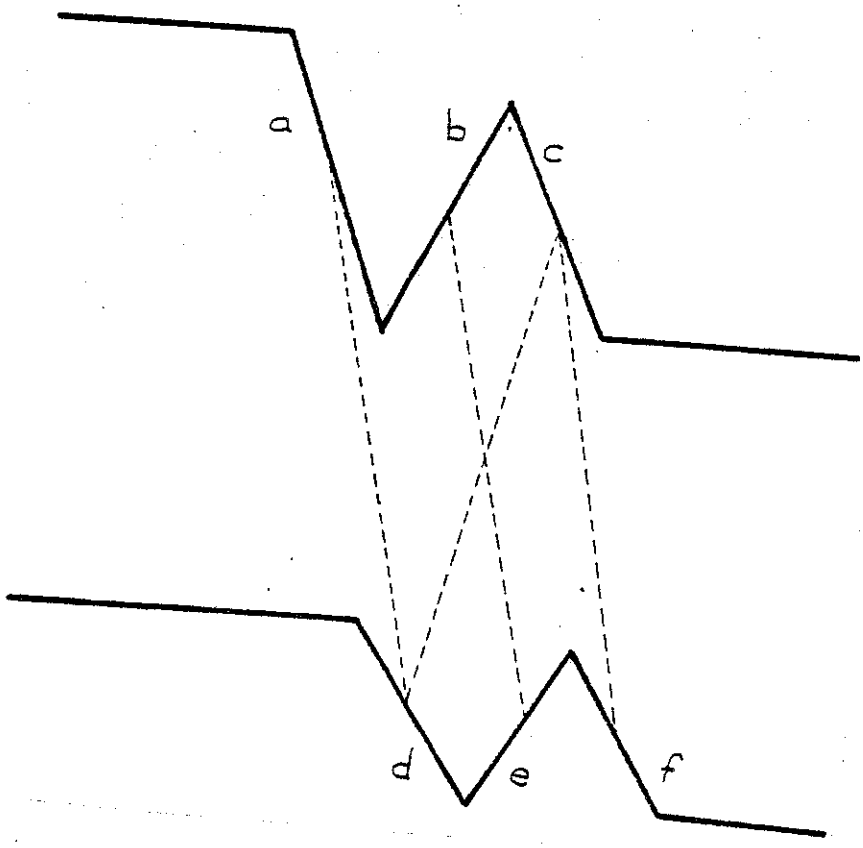


Figure 13 - Possible edge links between adjacent intensity profiles.

on the basis of similarity would be cd. This link would then constrain the others since links are not allowed to cross. Therefore, the linking process cannot be based solely upon the similarity of edge hypotheses. It must also consider the global effect that each link will have on the entire linking arrangement. Because of this, the linking procedure must involve a search of some kind that includes the "strongest first" paradigm.

The "strongest first" paradigm states that the strongest links should be made first unless they constrain links with similar strength. The strongest first paradigm is closely related to least commitment because it is undesirable to match weak hypotheses first and allow them to constrain other links about which one is more confident. The strength of a link is based upon the similarity between the edge hypotheses to be linked. Since we wish to define a cost measure that decreases in proportion to the similarity between the candidate hypotheses, similarity is defined on the basis of endpoint proximities. In cases where two or more links with similar strengths constrain one another as in Figure 13, the decision for linking must be based upon other criterea. Before linking begins, an edge pair table is formed that contains the costs for each feasible link between a pair of edge hypotheses in adjacent intensity profiles.

The cost of forming an edge link is a combination of the match between the two hypotheses and the strengths of the individual hypotheses. Roughly, the cost is the ratio

$$\text{cost} = \frac{\text{Differences between edge hypotheses}}{\text{Strength of individual edge hypotheses}}$$

The cost value is determined by the differences in the end points of the edge hypotheses as shown in Figure 14. Let the two sets of end points $\{(X_{a1}, Y_{a1}), (X_{a2}, Y_{a2})\}$ and

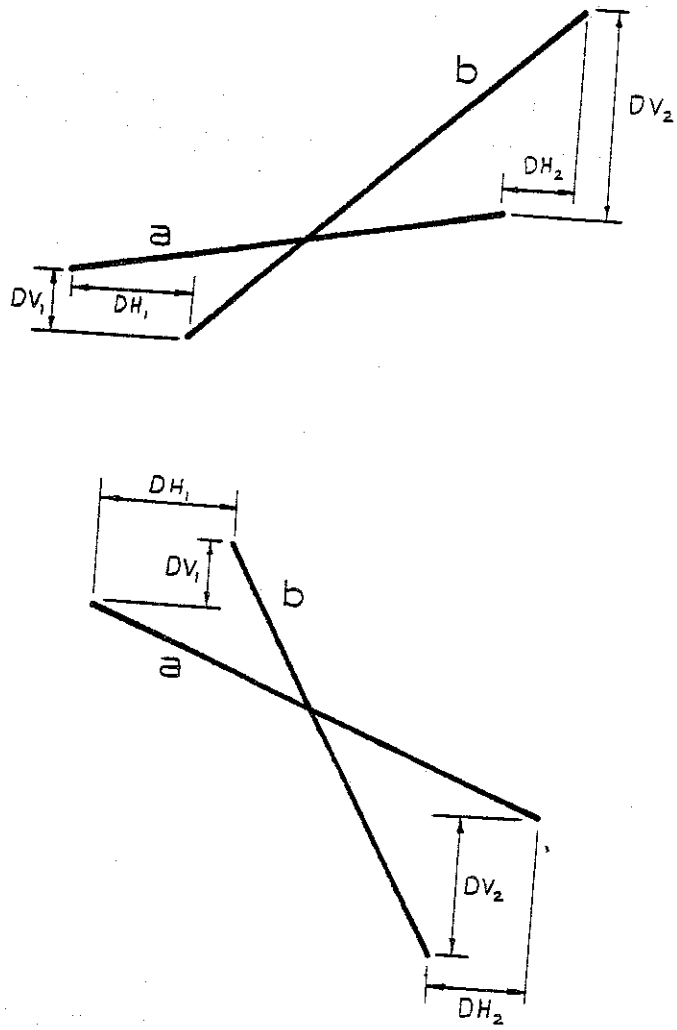


Figure 14 - Pairs of superimposed edge hypotheses from adjacent intensity profiles.

$\{(X_{b1}, Y_{b1}), (X_{b2}, Y_{b2})\}$ represent the end points of pairs of edge hypotheses from adjacent intensity profiles. Then

$DV_1 = |Y_{a1} - Y_{b1}|$

$DV_2 = |Y_{b2} - Y_{a2}|$

$DH_1 = |X_{a1} - X_{b1}|$

$DH_2 = |X_{b2} - X_{a2}|$

S_1 = contrast of edge hypothesis a

S_2 = contrast of edge hypothesis b

The edge contrasts are determined by the edge detector and stored in the attribute lists. Given these difference values, the cost of the link is given by

$$\text{cost} = \frac{DV_1 + DV_2 + DH_1^3 + DH_2^3}{S_1 + S_2}$$

The two values DH_1 and DH_2 have the largest influence in the cost function since they represent the differences in horizontal placement which is very important in the matching procedure. The cost function was made inversely proportional to the contrasts of the edge hypotheses since a link between strong edges of given similarity should have lower cost than a link between weaker edges with the same similarity.

Given two sets of edge hypotheses, not every pairing is entered into the cost table. Since edge region pairs in the intensity profiles are spatially aligned in the horizontal direction and have similar slopes, it is unreasonable to enter pairings that are drastically different. Therefore, the following contraints must be met before a pair is entered. First, edge hypotheses must overlap each other horizontally or have a common X coordinate in their end points. Second, their slopes must have the same sign. Third, given that the first two contraints have been met, the cost of the link must be below a preset value. This value is the same for all images and was determined by

examining the costs of linking edge hypotheses with varying similarities and contrasts. The value chosen was 1.5.

7.3 - The Linking Procedure

The first step in the development of the linking procedure was the determination of the search cost function. The cost function must allow as many links to be made as possible so long as the total cost for the linking arrangement remains reasonable. If this was not done, the optimal linking arrangement would be the trivial one where no links are made. The search is to find that linking arrangement that maximizes the total number of links subject to the condition that no link cost exceeds a preset threshold t.

Let P_i and P_{i+1} be two adjacent intensity profiles with respective sets of edge hypotheses H_i and H_{i+1} where $H_i = \{e_1, e_2, \ldots, e_n\}$ and $H_{i+1} = \{f_1, f_2, \ldots, f_m\}$. Then $H_i \, XH_{i+1}$ is a set that contains all possible links between P_i and P_{i+1}.

Let $c(e_j f_k)$, $e_j f_k \in (H_i \, XH_{i+1})$ be the cost of linking e_j to f_k.

Then the minimal cost arrangement is given by

$$\min_h C = \sum_{j,k \in K_h} c\ (e_j f_k) + \frac{t}{2}\ (N - 2|K_h|)$$

where K_h = set of links in a given linking arrangement h

$N = |H_i| + |H_{i+1}|$

t = maximum cost link to be considered

Paths are followed in the search tree until the cost of the lowest cost unused link exceeds a threshold t. The longest path in the search tree then represents the linking arrangement to be used. If the longest path is not unique, the one with the lowest cost is used.

Since the cost depends upon the entire search path and, in fact, decreases as a goal node is approached in the search tree, it appears that a full depth first search is required to determine the optimal path. This does not cause as many difficulties as it appears. Consider Figure 15 which shows two separate edge regions A and B in adjacent intensity profiles. It is clear that the links in one region do not interfere with the links of the other. Therefore, the search should consider these regions separately. In general, the search is then a series of local searches which drastically reduces the size of the search tree.

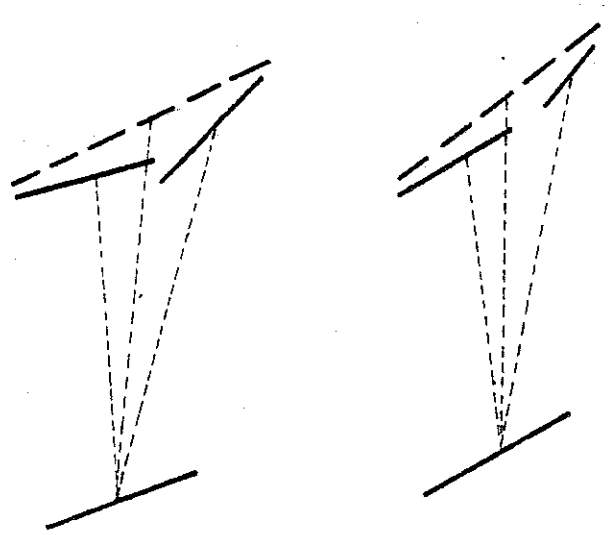


Figure 15 - Two separate edge regions in adjacent intensity profiles.

The size of the search tree can be reduced again by using the "strongest first" paradigm. Let $l_1, l_2, \ldots, l_n$ be a set of links over which a local search is to take place. If link l_i is the lowest cost link in this set and its cost is much less than those of the links it constrains, then link l_i may be included in every linking arrangement. If l_i interferes with a link that has a similar cost then both possibilities must be kept open.

7.4 - Edge Postprocessing

Next, we have implemented a postprocessing phase that generates additional links in cases where there are no

ambiguities. If no compound edge hypotheses were involved in the linking procedure, then the edge linking process for a pair of adjacent intensity profiles would be complete. If compound edge hypotheses are involved, then postprocessing can be done to increase the number of links.

As seen in Figures 16a and 16b, the same edge regions are present in three adjacent profiles although the edge boundary is not continuous because linking is done between pairs of profiles. Cases such as this can be located and corrected very easily at this stage of the processing. Figures 17a and 17b show the results after forcing a link through the middle edge regions.

This type of processing is not always possible. As seen in Figures 18a, 18b, 18c, and 18d there exist ambiguous cases. At this level of processing no further linking decisions can be made. Further linking must be left to a process that can examine more global information or to a process that has a priori knowledge of the image.

8 - Results

In this section, the results of the edge linking procedure are presented and are compared to the unlinked output of the Roberts Cross and Sobel operators.

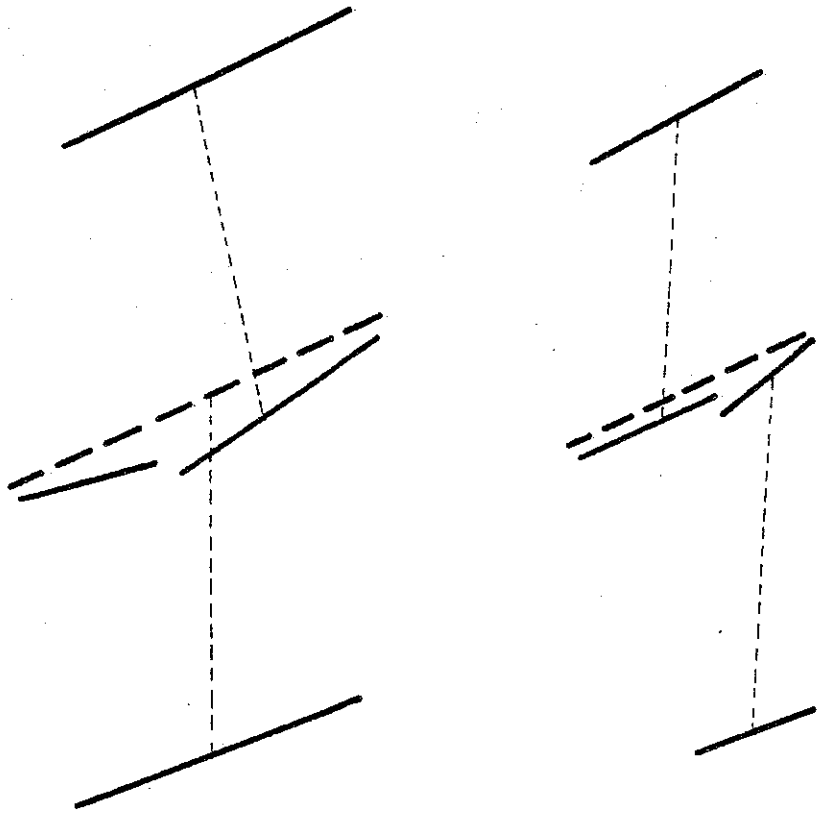


Figure 16 - Two edge regions in three adjacent intensity profiles.

Figure 2 shows the natural scene containing a house, trees, grass, and sky from which the first set of results were produced. Edges are very difficult to extract from an image of this type because of the large amount of texture present. Figure 19 shows the thresholded results of the Roberts Cross and Sobel operators on Figure 2. Each operator produces a slightly different edge output, and each

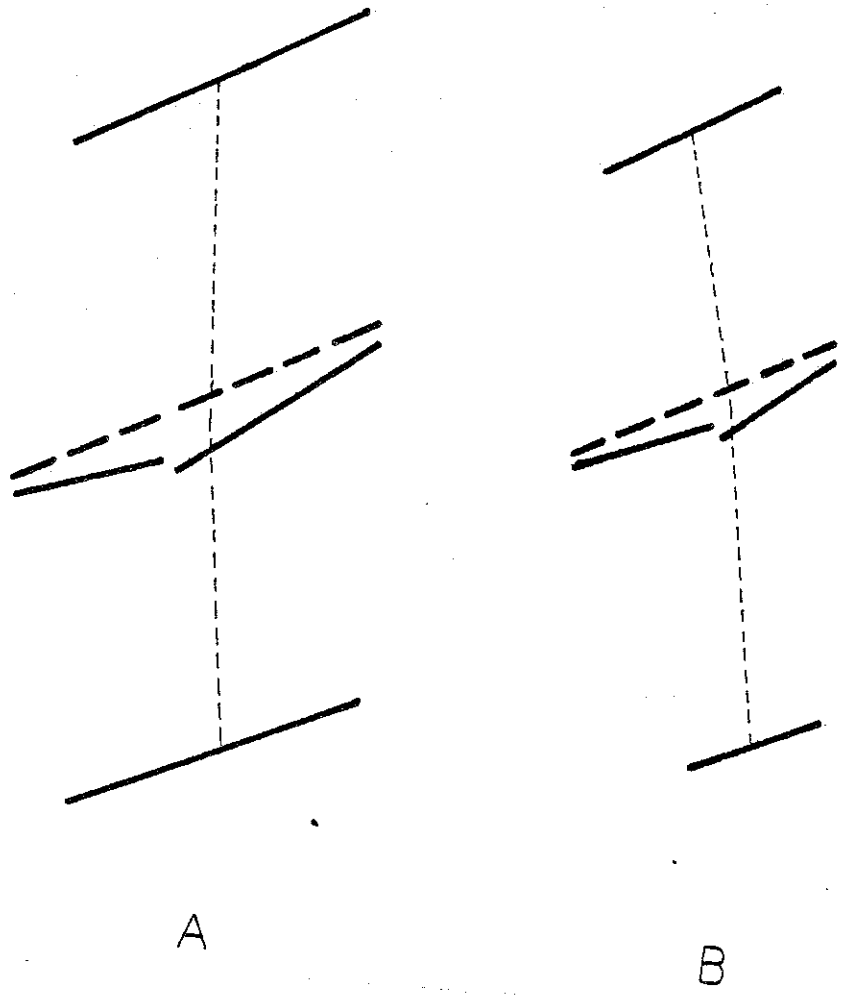


Figure 17 - Postprocessed links in three adjacent intensity profiles.

edge image has both good and bad aspects. The Sobel operator appears to produce better results for areas A and B of Figure 19 while the Roberts Cross operator appears to be superior for areas C and D. However, neither does very well on the obscured corner of the left set of windows in area E. The differences between the operators are due, of course, to the fact that the Sobel operator is a larger operator. Therefore it will do a better job on texture boundaries but a much poorer job on fine detail.

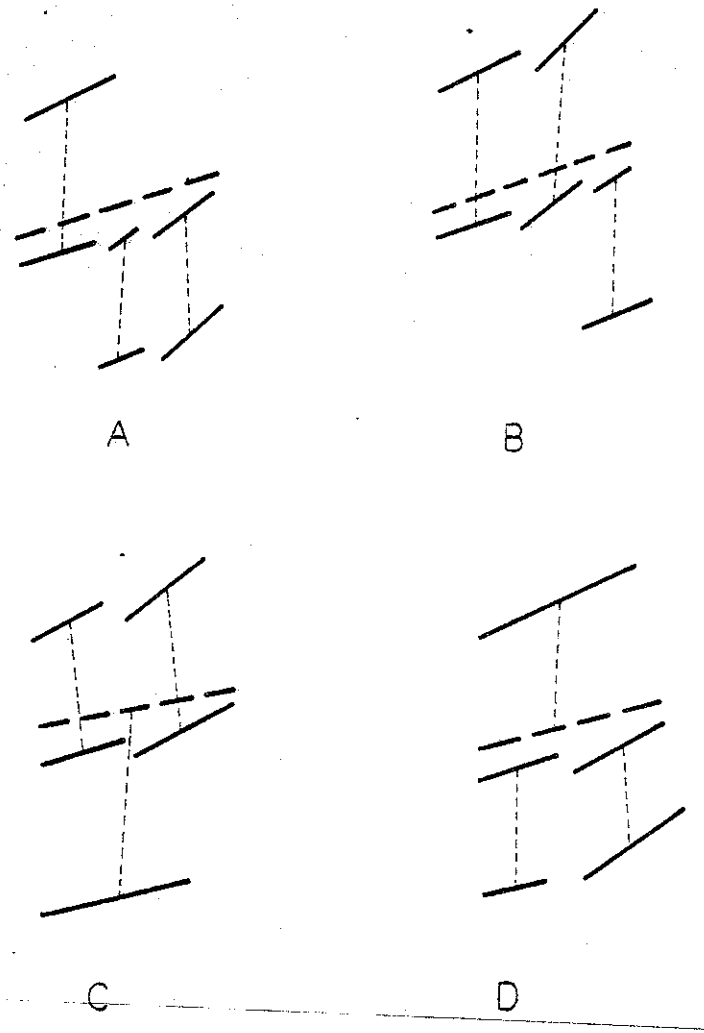


Figure 18 - Ambiguous links in three adjacent intensity profiles.

Figure 20 is an overlay of vertical and horizontal linked edges produced by the linker from Figure 2. In all there are over 14,000 edge elements and over 3,700 lines. However, the line drawing can be simplified dramatically by discarding the short edges. Figure 21 shows the results of eliminating lines containing fewer than 3 or 10 edge elements. Notice that in Figure 21b a significant portion of the house remains even though only 293 lines are required. Notice also the larger branches of the trees that are difficult to observe when shorter lines are included.

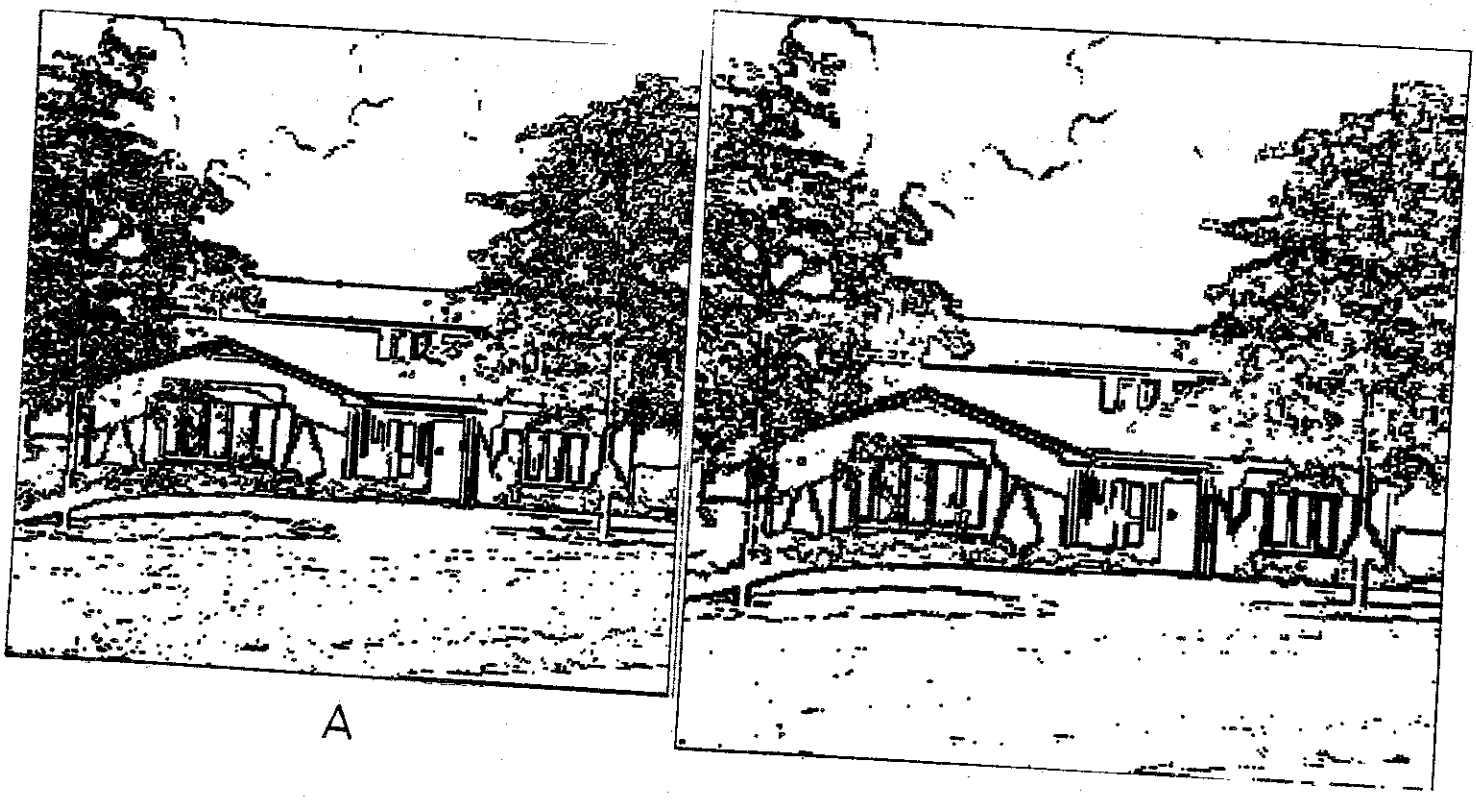


Figure 19 - Roberts Cross (a) and Sobel operator (b) output from Figure 2, thresholded at 45.

Figure 22 shows enlargements of four areas of special interest in the house scene of Figure 2. All lines of length greater than 2 have been shown. In Figure 22a, the procedure generates very accurate and complete edges even in the area where the window is obscured by the tree. It also locates the boundaries caused by shadows at the top of the window and near the peak of the roof. In comparison, the Roberts Cross generated ambiguous patches of strong edge points between the windows, and the Sobel operator generated unaligned points for the shadow near the peak of the roof.

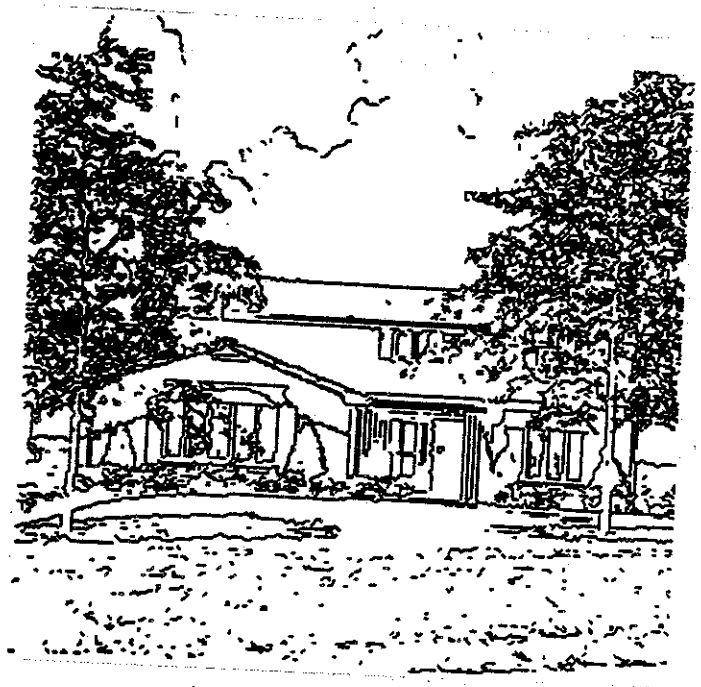


Figure 20 - Overlayed linked edges from linking procedure applied in vertical and horizontal directions.

As shown in Figure 22b, the edge linking procedure locates most of the detail near the door of the house scene. The Sobel operator appears to generate a more complete representation for the first window to left of the door, but it fails to separate the individual pillars to its right and the windows to the far left of the door. Figure 22c shows a window that has been obscured by tree branches. Both the Sobel and Roberts Cross operators generate ambiguous connections between the edges of the windows and the tree while the edge linking procedure tends to segment the window from the background. Finally, Figure 22d shows that the

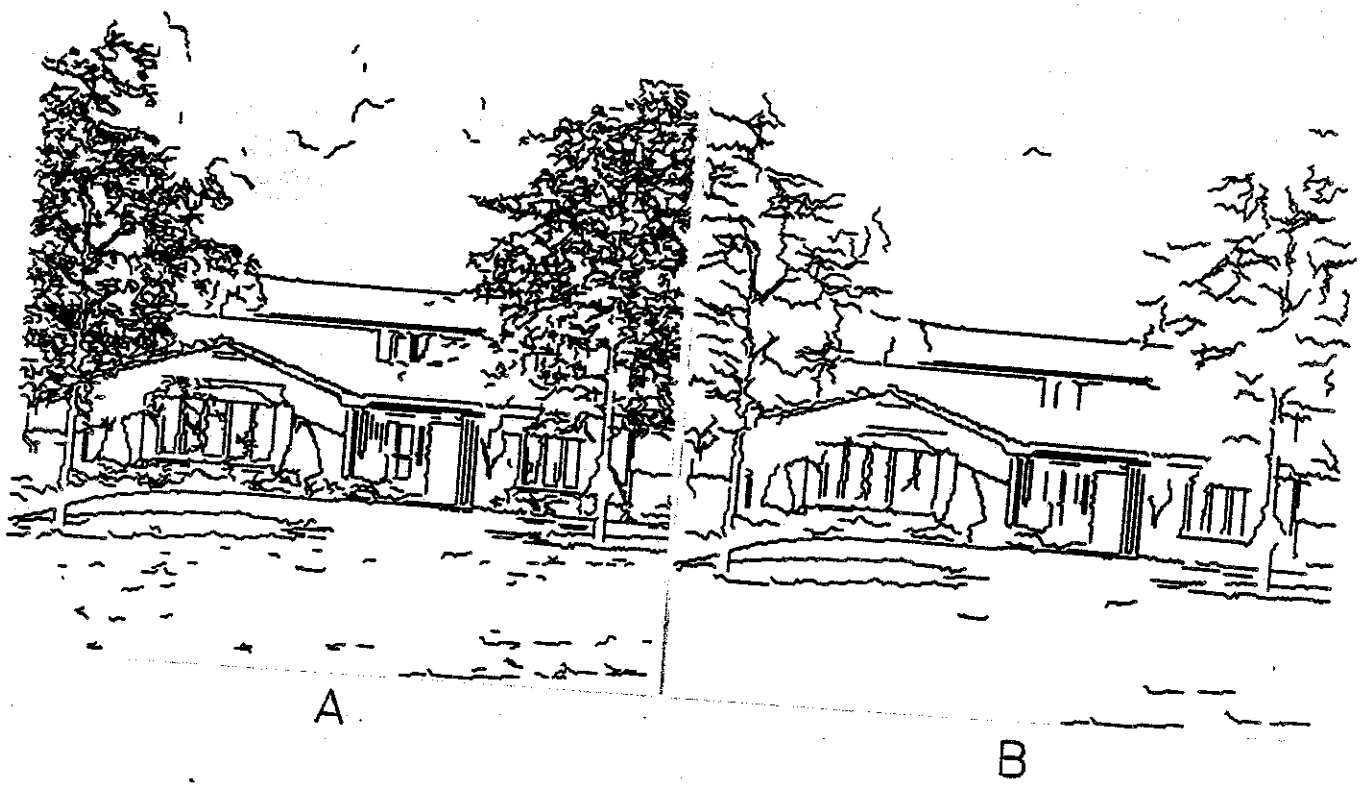


Figure 21 - Same as Figure 20 except lines shorter than 3 edgels removed (a) and lines shorter than 10 edgels removed (b).

detail of the upstairs window is much better in the line drawing generated by the edge linking procedure, particularly for the middle window. The Roberts Cross operator tends to fragment the portion obscured by the tree into ambiguous patches while the edge linker generates a smooth boundary.

Continuing with some other examples, Figure 23 shows a high contrast industrial scene with a noisy background. Figure 24a shows the Roberts Cross thresholded at 35, and Figure 24b shows the lines produced by the edge linker

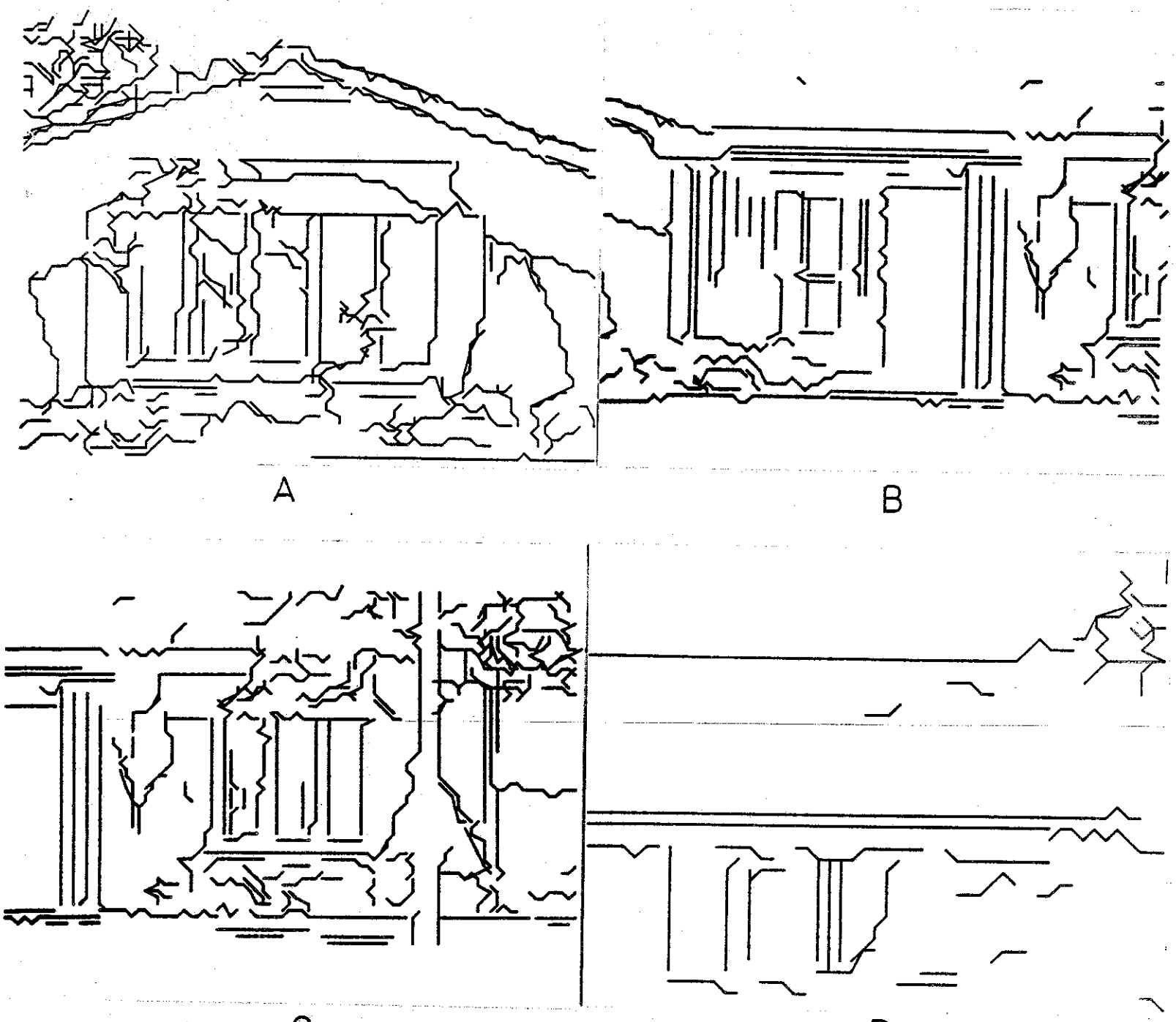


Figure 22 - Enlargements of left window (a), door (b), right window (c), upper window (d) showing edges of length 3 or more.

having length at least 10 edgels. Notice the smoothness of the edges. The double edges are caused by shadows, and these edges can be distinguished since they have opposite polarities.

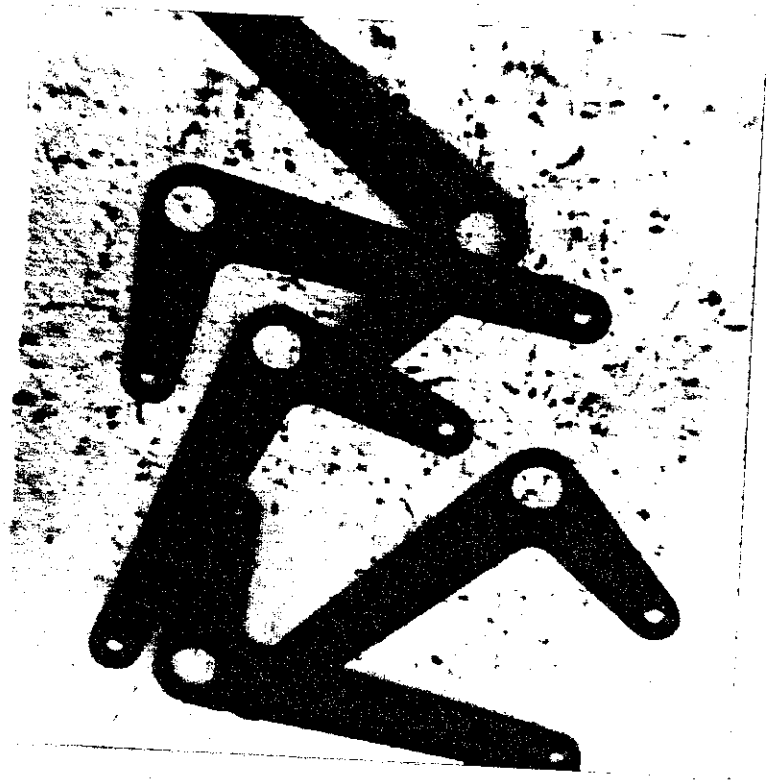


Figure 23 - High contrast industrial scene.

Finally, Figure 25 shows another very complicated industrial scene with occluded objects and complicated surface texture. Figure 26a shows the Roberts Cross with threshold 50, and Figure 26b shows the lines produced by the linker with length at least 5 edgels. Notice how clearly regions A, B, and C are defined. The parallel edges of the capacitor leads cannot be confused, since they have opposite polarities.

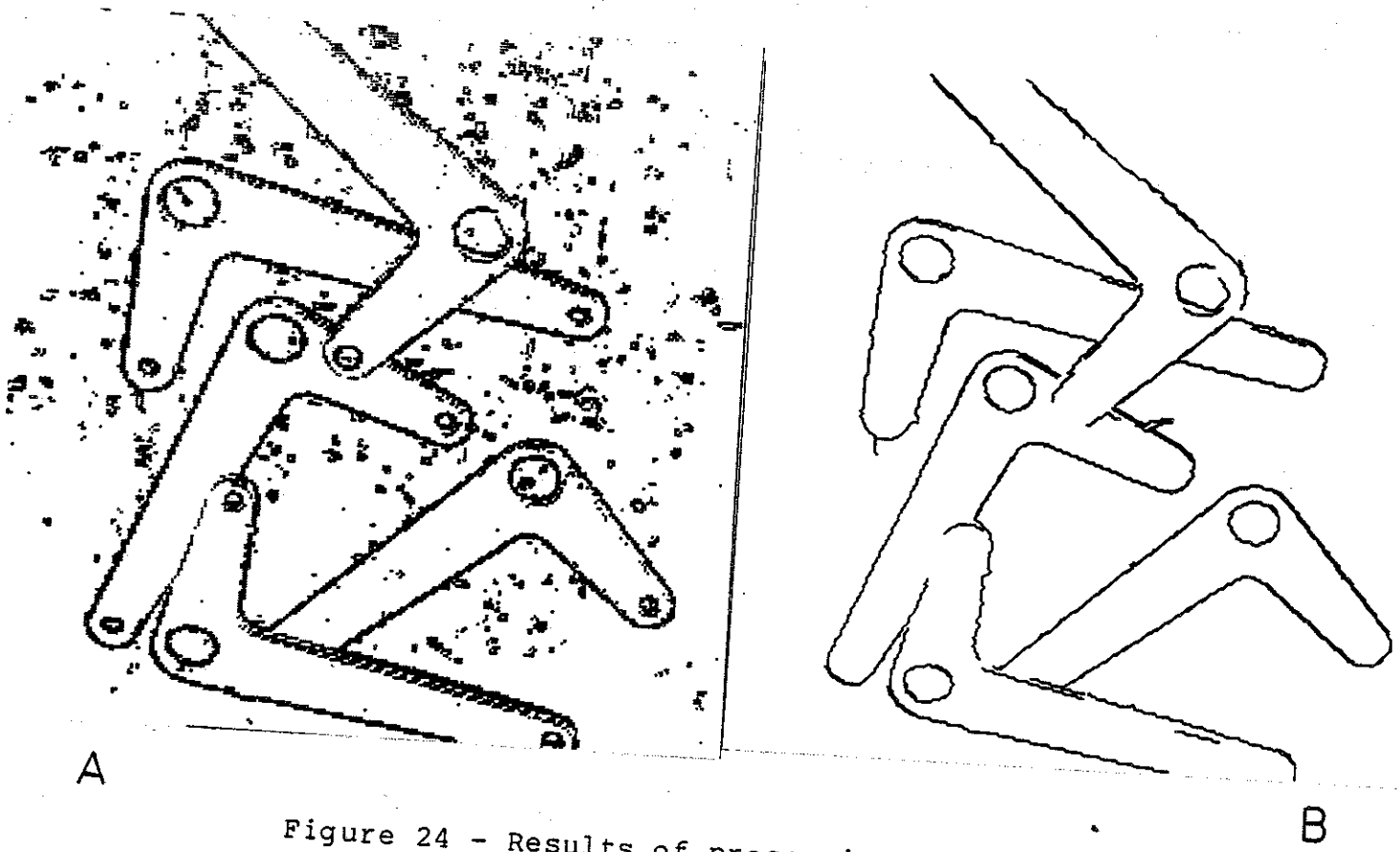


Figure 24 - Results of processing Figure 23. Roberts Cross thresholded at 35(a) and lines produced by linker with length at least 10 edgels (b).

In general, the edge linking procedure generates very good results for most images. In regions of low contrast the edges tend to become fragmented, but this is a problem encountered by all other edge detectors as well. The linking procedure does extremely well in regions where ambiguous edge points are generated by local operators.

In all the experimental results, it must be emphasized again that even though the linker output may appear dense, the edge elements are linked together, and the edge

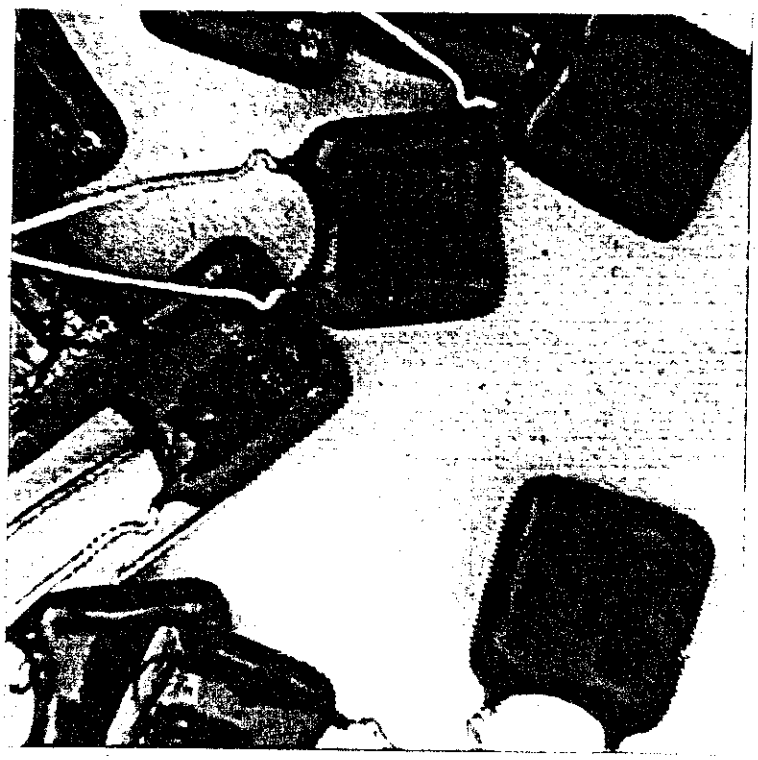


Figure 25 - Complex industrial scene.

polarities are known. Hence, vastly more information is available than appears in these images.

9 - Conclusions

It is felt that the edge detector described in this paper is more suitable for many applications in image analysis than many currently in use because it is more sensitive, more accurate, and because it does not make irrevocably incorrect interpretations of edges.

Figure 26 - Results of processing Figure 25. Roberts Cross thresholded at 50 (a) and lines produced by linker with length at least 5 edgels (b).

By treating edges as regions and by observing least commitment, good line drawings can be produced from gray level images. There are a number of cases where the procedure does not generate contiguous lines; however, these are ususally natural texture boundaries that do not meet the initial requirement of smooth and continuous edges. Texture boundaries such as these must be dealt with by algorithms designed explicitly for this purpose.

Further work in this area could involve not only matching pairs of adjacent profiles but also larger sets of adjacent profiles. This would provide much more information for the linking procedure, and it would eliminate many of the ambiguous cases encountered by the postprocessing procedure.

The edge formation procedure presented in this paper does not produce adequate results for textural boundaries since it was not designed for this purpose. Also, the problem of matching the full relational trees for edges has not been solved. This does not imply that the concepts are inappropriate for textural boundary formation; rather, the current implementation is inadequate. We believe that generalized region matching algorithms will be very powerful tools for region growing as well as for boundary formation.

The next step in the processing of these raw line drawings is a phase in which regions are formed, subjective contours are completed, and unwanted objects are removed. For this purpose we are using a special data base system for manipulating line drawings. In this system information such as polarity, length, and contrast of a line is retained so that subsequent processing steps will have the necessary information. The processing of the raw line drawings is a sophisticated process that will be the subject of a later paper.

References

1. Horn, B.K.P., "The Binford-Horn linefinder," Artificial Intelligence Memo 285, MIT AI Laboratory, Cambridge, March 1973.

2. Milgram, D.L., "Region extraction using convergent evidence," Computer Graphics and Image Processing 11, January 1979, pp. 1-12.

3. Hueckel, M.H., "An operator which locates edges in digital pictures," J. ACM 18, January 1971, pp. 113-125.

4. Roberts, L.G., "Machine perception of three-dimensional solids," in Optical and Electrooptical Information Processing, J.T. Tippett et.al., eds., MIT Press, Cambridge, 1965, pp. 159-197.

5. Duda, R.O. and P.E. Hart, Pattern Classification and Scene Analysis, Wiley, New York, 1973, pp. 271-272.

6. Kirsch, R., "Computer determination of the constituent structure of biological images," Computers and Biomedical Research 4, June 1971, pp. 315-328.

7. Rosenfeld, A. and M. Thurston, "Edge and curve detection for visual scene analysis," IEEE Trans. Comp. 20, May 1971, pp. 562-569.

8. Marr, D., "Early processing of visual information," Phil. Trans. Royal Soc. London 275 on Biological Sciences, 19 October 1976, pp. 483-524.

9. Ehrich, R.W. and J.P. Foith, "Representation of random waveforms by relational trees," IEEE Trans. Comp. 25, July 1976, pp.725-736.

10. Marr, D., "Visual information processing: the structure and creation of visual representations," Proc. 6th International Joint Conference on Artificial Intelligence, Tokyo, Japan, August 1979, pp.1108-1126.

11. Rosenfeld, A. and A.C. Kak, Digital Picture Processing, Academic Press, New York, 1976.

12. Martelli, A., "Edge detection using heuristic search methods," Computer Graphics and Image Processing 1, August 1972, pp. 169-182.

13. Montanari, V., "On the optimum detections of curves in noisy pictures," Comm. ACM 14, 1971, pp. 335-345.

14. O'Gorman, F. and M. Clowes, "Finding picture edges through collinearity of feature points," Proc. 3rd JCAI, August 1973, pp. 543-555.

15. Hanson, A.R. and E.M. Riseman, "Segmentation of natural scenes," in Computer Vision Systems, A.R. Hanson and E.M. Riseman, eds., Academic press, 1978.

16. Pavlidis, T., "A Minimum storage boundary tracing algorithm and its application to automatic inspection," IEEE Transactions on Systems, man, and Cybernetics 8, January 1978, pp. 66-69.

17. Feng, H.Y. and T. Pavlidis, "The generation of polygonal outlines of objects from gray level pictures," IEEE Transaction on Circuits and Systems 22, May 1975, pp. 427-439.

List of Captions

Figure 1 - Ambiguous edge.

Figure 2 - House scene.

Figure 3 - Vertical profile from part of Figure 2 (a), First difference (b), Central difference (c), Laplacian (d).

Figure 4 - Profile (a) and Derivative (b) from Figure 3 and Relational tree of derivative (c).

Figure 5 - Part of an intensity profile (a), First derivative (b), and Relational tree (c).

Figure 6 - Part of an intensity profile (a), First derivative (b), and Relational tree for merged edge regions (c).

Figure 7 - K>30, T=1, Forward differences (a), and Central differences (b).

Figure 8 - K>30, T=20, Forward differences (a), and Central differences (b).

Figure 9 - Simple and compound hypotheses.

Figure 10 - Formation of compound edge hypotheses.

Figure 11 - Two adjacent profiles from an image and all simple and compound edge hypotheses.

Figure 12 - Edge hypotheses in adjacent profiles.

Figure 13 - Possible edge links between adjacent intensity profiles.

Figure 14 - Pairs of superimposed edge hypotheses from adjacent intensity profiles.

Figure 15 - Two separate edge regions in adjacent intensity profiles.

Figure 16 - Two edge regions in three adjacent intensity profiles.

Figure 17 - Postprocessed links in three adjacent intensity profiles.

Figure 18 - Ambiguous links in three adjacent intensity profiles.

Figure 19 - Roberts Cross (a) and Sobel operator (b) output from Figure 2, thresholded at 45.

Figure 20 - Overlayed linked edges from linking procedure applied in vertical and horizontal directions.

Figure 21 - Same as Figure 20 except lines shorter than 3 edgels removed (a) and lines shorter than 10 edgels removed (b).

Figure 22 - Enlargements of left window (a), door (b), right window (c), upper window (d) showing edges of length 3 or more.

Figure 23 - High contrast industrial scene.

Figure 24 - Results of processing Figure 23. Roberts Cross thresholded at 35 (a) and lines produced by linker with length at least 10 edgels (b).

Figure 25 - Complex industrial scene.

Figure 26 - Results of processing Figure 25. Roberts Cross thresholded at 50 (a) and lines produced by linker with length at least 5 edgels (b).