Pattern Recognition Letters 9 (1989) 127 136
North-Holland February 1989

Matching topographic structures in stereo vision

Ting-Chuen PONG
Department of Computer Science, University of Minnesota, Minneapolis, MN 55455, USA

Robert M. HARALICK, Linda G. SHAPIRO
Department of Electrical Engineering, University of Washington, Seattle, WA 98195, USA

Received 9 March 1988
Revised 9 November 1988

Abstract: Stereo vision is important in determining three-dimensional positions of the visible surfaces in images. Most existing stereo matching methods involve matching low-level image features such as edges. A stereo matching technique based on matching high-level image features is described in this paper. The high-level image features extracted are topographic structures, arc segments and region segments. Matching results in stereo images are also discussed.

Key words: Stereo matching, feature extraction, edges, topographic structures, computer vision.

1. Introduction

A major industrial application of computer vision systems is in automatic inspection. One approach to such an inspection task is a coordinated system that employs robot arms with measurement devices for inspection along with a stereo vision system that provides information to a planner that guides the robot arms [20]. The functions of the vision system are to determine the position of the objects in the scene in relation to the camera(s) and to determine the portion of object which the camera is viewing. The first problem can be solved by finding correspondences between features extraced from the stereo pair of images. The second problem can be approached by matching structures extracted from the images to parts in a three-dimension description of the object to be inspected. One such object model can be found in [20] which employs a hierarchical relational model to describe a complex F-15 bulkhead.

This work was supported in part by NSF Grant ECS-8505662.

A stereo approach to automatic inspection must solve the following problems. First, how are structures in the images to be extracted? In particular, we are interested in extracting structures which are usuful for finding correspondences between the images of the stereo pair and for matching structures in the three-dimensional object model. To extract structures from an image usually involves some kind of segmentation procedure. Unfortunately, none of the existing segmentation techniques seems to produce satisfactory results. In this paper, we will examine a scheme for extracting line-like and surface-like structures which are used frequently in modeling three-dimensional objects [20]. Preliminary experiments show that this scheme produces improved results over the existing techniques.

The second problem is to find correspondences between structures in the stereo pair. Most existing stereo matching techniques use local features as primitives for matching. Although continuity constraints are imposed in some systems to resolve ambiguity, it seems that the use of local features is far from adequate. In solving this problem, we will describe a stereo matching technique based on matching higher-level primitives such as arc segments, region segments and topographic structures.

2. Feature extraction

The goal of early visual processing is to extract a rich symbolic representation of the gray scale intensity changes in an image. Marr [14] names this representation the primal sketch. It is important for the primal sketch to capture all gray scale intensity variations in an image because any intensity changes can be related to changes in scene characteristics such as surface orientation, surface discontinuity, surface reflectance and illumination. Marr's primal sketch relies solely on a special form of intensity changes which are detected as zero-crossings of the Laplacian of a Gaussian filtered image [16]. Unfortunately, the use of this representation alone cannot provide a rich enough description for intensity changes, since it fails to account for smooth intensity changes or shadings which frequently appear in images of curved surfaces.

Haralick [11] proposed a rich and robust representation for all types of two-dimensional intensity variations. This representation is called the topographic primal sketch. We will present here a feature extraction scheme which extracts topographic structures such as edges, ridges, valleys and hillsides. Edges, which usually correspond to sharp changes in gray scale intensities, are used to describe the basic structure of an image. The rest of the topographic structures are used to describe the intensity variations within regions which are extracted as connected sets of non-edge structures.

The proposed feature extraction scheme can basically be outlined as follows:
1. detect raw edge elements by a local edge operator;
2. extract arc segments by linking fragmented edges and cleaning noisy edges;
3. assemble non-edge pixels into regions by a connected components algorithm; and
4. compute topographic structures with regions.

2.1. *Extracting edge elements*

Edge detection was pioneered by Roberts [19]. A comprehensive survey of edge detection can be found in [7]. A set of edge detection criteria that capture the desirable properties of an edge operator is formulated in [5]. A comparison of the various edge operators is given by Haralick [12]. Based on the results of [12], Haralick's second directional derivative edge operator was selected for our initial edge detection process. A precise mathematical definition of the operator can be found in [12]. The results of the second directional derivative edge operator applied to the stereo images of Figure 1 are illustrated in Figure 2.

Although the second directional derivative zero-crossing edge operator is less sensitive to noise than the classical gradient edge operators, it still produces undesired and fragmented edges for a complex and noisy image such as the one shown in Figure 1. Moreover, industrial objects often consist of shiny metallic parts. Specular reflection creates further problems in edge detection. In order to obtain more meaningful edge structures, a procedure is used to extract arc segments by linking fragmented edge segments and cleaning noisy edges.

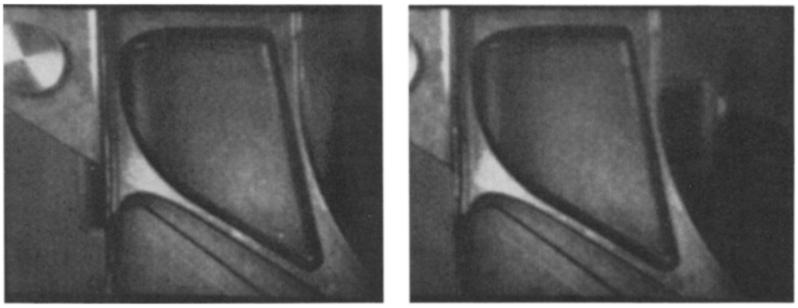

Figure 1. Stereo images of a F-15 bulkhead.

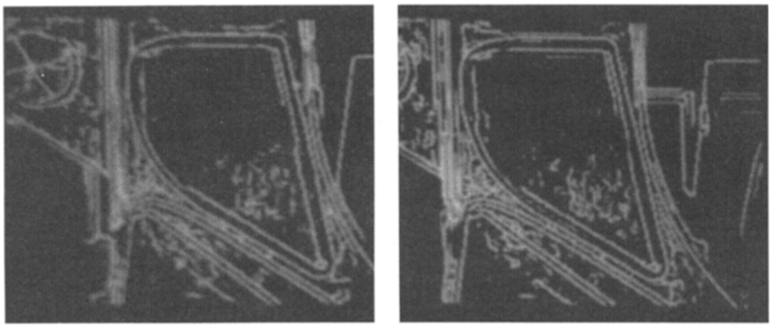

Figure 2. Edges detected from the images of Figure 1.

2.2. *Extracting arc segments*

Edge cleaning and edge linking are two difficult problems which may lead to conflicting consequences. Edge cleaning tends to create wider gaps between already broken edges. Linking noisy edges often results in more undesired edges. An attempt is made here to give an even-handed treatment of both. The proposed technique is based on the observation that true edges are usually detected as sequences of orderly oriented pixels, while noisy edges are usually oriented in some random manner. The idea behind this scheme is to first link edge pixels into arc segments based on their detected orientations; isolated or short segments will then be deleted and a final gap filling scheme is used to create closed boundaries.

The edge linking process attempts to link together edge pixels which are detected as edges with similar orientations. This edge linking process is in some sense similar to Burns et al.'s [4] method for extracting straight lines. Our approach is more powerful than theirs in the context that our method extracts arc segments other than straight lines. Instead of estimating edge orientation by a 2×2 mask as in [4], we compute edge orientation directly from the facet fitting coefficients. Specifically, edge orientation is defined to be the direction perpendicular to the direction which extremizes the first directional derivative. A region growing scheme is employed for the edge linking process. This region growing scheme differs from the traditional region growing scheme in that the similarity measure is derived directly from edge orientation instead of from gray level intensity and that the resulting regions are long-thin edge regions.

An image is scanned left-to-right and top-to-bottom. Each edge pixel's orientation is compared to the mean orientation of all its neighboring edge segments. An edge segment is selected such that its mean orientation is closest to the edge pixel's orientation. The edge pixel is merged into the edge segment provided that their orientations are not too different. If no merging is possible, the edge pixel starts a new segment.

After a single scan through the image, a set of initial segments is created. In order to produce more meaningful and longer arc segments, a second merging processing is performed. In this step, mean orientations of adjacent edge segments are compared. Adjacent edge segments are merged if their mean orientations are not too different. A cleaned edge image can then be obtained by eliminating arc segments of sizes less than a certain predetermined threshold. Results show that our technique not only removes most of the undesired edges, but also groups edges into meaningful arc segments which are useful for higher level matching. One undesirable result of removing all short segments is that it disconnects edges at locations where there are sharp changes in orientation, for example, at corners. This problem can be fixed by joining nearby end points of neighboring arc segments.

2.3. *Extracting region segments*

Once a final set of arc segments is determined, the regions extracted are the largest connected areas of pixels which are entirely surrounded by arc segments. This process is accomplished by a connected components algorithm which assigns a unique label to each maximally connected group of non-edge pixels. An efficient memory-limited connected components algorithm can be found in [13]. This algorithm requires only one top-down scan and one bottom-up scan of the entire image. The non-edge pixels are labeled as regions, and the arc segments which exist between regions remain unaffected. This region extracting scheme can sometimes create small regions which may not contain enough information for meaningful matching, therefore, regions of sizes less than a certain threshold are usually eliminated. The region segments of the images of Figure 1 are shown in Figure 3.

2.4. *Extracting topographic structures*

Multiple images of the same scene can be obtained under different illumination conditions and with a variety of camera gain settings. Such photometric variations can create problems if raw intensity values are used for stereo matching. We are interested in describing intensity variations within region segments in a way which is insensitive to photometric variations. The major categories {peak, pit, ridge, valley, saddle, hillside and flat} of the topographic primal sketch [11] have been proven to be invariant under monotonic gray scale transformations. Although its subcategories {edge, slope, convex hill, concave hill and saddle hill} may change under gray scale transformation, to the extent we deal only with slight alterations in imaging condition (which is usually the case for stereo vision), it is reasonable to assume that such changes are tolerable.

The topographic primal sketch can be used to represent the underlying intensity surface of a digital image. A digital image may be interpreted as a sampling and quantizing of a real-valued function f. While the image is a discrete matrix of values, the underlying surface is continuous. The gradient magnitude and the directional derivatives computed from the underlying surface are used in determining the topographic labeling of the surface. A complete mathematical treatment of the topographic primal sketch is given in [11].

3. Stereo matching

Stereo matching has been approached mainly in two directions. The first approach matches area features, while the second approach matches lineal features such as edges. The area based approach attempts to find a match for each neighborhood in the first image by searching for neighborhoods of maximum similarity in the second image. A typical similarity measure is the mean-square difference of the gray levels within the two neighborhoods. Since the search process is computationally expensive and not all pixels can be unambiguously matched, matches should be restricted to neighborhoods with high information content. One such technique for finding candidate points for matching is Moravec's interest operator [18]. Moravec's operator computes an initial interest measure for each pixel by finding the minimum of the four directional variances (horizontal, vertical and two diagonals) over the neighborhood centered around the pixel. A pixel is of interest if its interest measure attains a

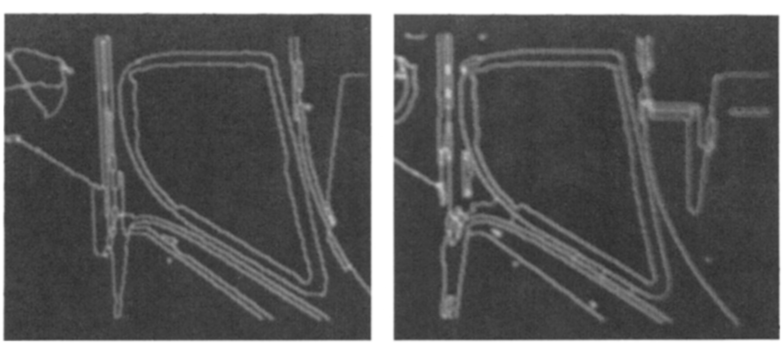

Figure 3. Regions extracted from the images of Figure 1.

local maximum. Moravec's interest operator tends to select corners.

Barnard and Thompson [3] use a relaxation labeling technique to find matches for area features. They use Moravec's interest operator to select matchable points. The searching space is restricted by assuming a maximum disparity. An initial probability is given to each match based on the similarity of the neighborhoods around the pixel. A relaxation labeling technique is then used to improve the initial probabilities by using the local continuity property of disparity. A common problem with area-based approaches is that they are sensitive to changes in contrast and brightness between the pair of images.

The edge-based approach attempts to search for edge correspondence between edges detected from the stereo pair. Since an edge often corresponds to a discontinuity in the depth of an object, edge-based methods are less sensitive to photometric characteristics of the object in the scene. Furthermore, edge-based matchings are in general more accurate and less expensive. This is because there are fewer edges than pixels and edges can usually be detected at reasonably precise locations.

A computational theory of human stereo vision was formulated by Marr and Poggio [15]. They suggest matching features which are located as zero-crossings of the Laplacian of a Gaussian filtered image. This scheme works well on random-dot stereograms, but fails for real images. Grimson [10] implemented a modified version of Marr's algorithm. Grimson's approach seems to generate reasonable results for some real images. A coarse to fine approach is used in these methods. This is done by convolving the image with filters of various resolution. This approach tends to fail for complex scenes because edges detected from low resolution filters do not usually correspond to any true edges and incorrect matches are often introduced. Thus, correct matches as well as incorrect matches propagate from the low levels to the higher levels.

More recently, Baker and Binford [2] and Arnold [1] developed two similar methods for stereo matching based on a dynamic programming technique. The Viterbi dynamic programming algorithm is used in the matching process. They attempt to find an optimum match for each pair of epipolar lines in the image by assuming that there is no order reversal of edges. The optimum match is obtained by maximizing a measure derived from local properties of edges and by requiring that an edge angle and interval constraint be satisfied. One limitation of these methods is that a dynamic programming implementation is possible only if the edge ordering assumption is not violated.

Until recently, most feature based stereo matching systems match low level edge primitives. Medioni and Nevatia [17] proposed a technique for matching higher level primitives, namely, line segments. Matching line segments from the two images are searched within a parallel sided window which has two of its sides parallel to the orientation of a line segment and the other two sides horizontal. By using such a window, all epipolar lines that intersect the line segment to be matched are considered and a maximum disparity is allowed for each match. This is an iterative technique. At each iteration, each possible match is evaluated by assuming a minimal differential disparity. A preferred match for each segment is selected to be the one which is most consistent with the disparities suggested by its neighbors.

While Mediono and Nevatia solve the correspondence problem by matching line segments, Goshtasby [8] suggests using region segments. Goshtasby first uses an image segmentation procedure to determine optimally similar regions, by an appropriate measure, from multiple images. Region correspondences are then achieved by a probabilistic relaxation labeling process.

Cheng and Huang [6] present a method for registering images. A relational description is used in their method to represent structures extracted from images. They extract straight line segments as primitives and use relations such as parallel, antiparallel, collinear and adjacent to describe the interrelationships between line segments. The image correspondence problem is then solved by matching relational structures. As another step toward using higher level primitives, the proposed matching scheme matches arc segments, region segments, and topographic structures, and uses spatial relations among these structures to resolve ambiguous matches.

3.1. Stereo imaging geometry

The stereo imaging geometry is illustrated in Figure 4. We assume that the image planes of the stereo cameras are coplanar and are placed at a distance f, the focal length of the lens, in front of the two camera foci. The cameras are arranged so that the camera baseline, the line joining the two camera foci, is parallel to the row scan direction of the image (the x-axis in Figure 4). Given any point in the scene, the epipolar plane is defined as the plane determined by the point and the camera foci. It is clear that the projections of a point in 3-space onto the two image planes must lie somewhere along the intersecting lines between the epipolar plane and the image planes. The intersecting lines on the left and right image planes are called epipolar lines. Therefore, the corresponding problem is reduced to finding matches along corresponding epipolar lines. If the input images are taken under a different arrangement, a preprocessing step is required to account for the differences. One such technique can be found in [9].

3.2. Notation

The segmentation process partitions the images I^l and I^r into arc and region segments. As a notation, we will use superscripts l and r to denote, respectively, descriptors for the left and right image (descriptors without superscript can be applied to either image). The matching process to be described is performed in both directions, that is, from the left to the right image, and vice versa. In the following discussion, we will concentrate on finding matches from the left image to the right.

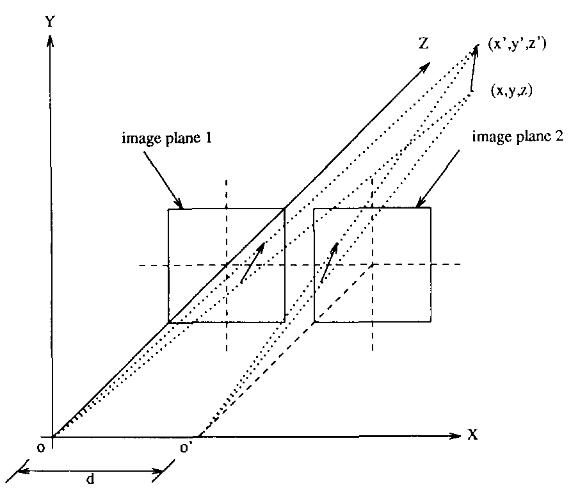

Figure 4. Stereo imaging geometry.

Let

N_A = number of arc segments,

$A = \{a_i \mid i = 1, \ldots, N_A\}$ = set of arc segments,

N_R = number of region segments,

$R = \{r_i \mid i = 1, \ldots, N_R\}$ = set of region segments.

Associated with each arc segment, the mapping $o{:}A \to [0,360]$ gives the mean orientation for each arc segment in A. Similarly, the mapping g: $R \to [I_{\min}, I_{\max}]$ gives the mean gray level for each region in R.

Define a boolean function $p(x,y)$ for $x \in A^l$ and $y \in A^r$ by

$$p(x,y) = \begin{cases} \mathit{true} & \text{if } d(x,y) < t_o, \\ \mathit{false} & \text{otherwise} \end{cases}$$

where t_o is a threshold for edge orientation and

$$d(x,y) = \begin{cases} |o(x) - o(y)| & \text{if } |o(x) - o(y)| \geq 180, \\ 360 - |o(x) - o(y)| & \text{otherwise.} \end{cases}$$

Thus $p(x,y)$ is *true* if and only if arcs x and y are oriented similarily enough. Define a boolean function $q(x,y)$ for $x \in R^l$ and $y \in R^r$ by

$$q(x,y) = \begin{cases} \mathit{true} & \text{if } |g(x) - g(y)| < t_g, \\ \mathit{false} & \text{otherwise} \end{cases}$$

where t_g is a threshold for mean gray level. Thus $q(x,y)$ is *true* if the mean gray scales of regions x and y are similar. The selection of the above thresholds is quite loose. Although tight thresholds can reduce the search space, it increases the risk of ruling out potential matches. In the examples presented in this paper, t_o is set to 30 and t_g is set to 50.

3.3. Matching along epipolar lines

Since arc and region segments can be very large, it will be expensive in terms of both memory and computation to correlate segments. Instead of treating segments as whole units, we will divide the matching problem into subproblems by first matching along epipolar lines. Region or arc segments correspondences, which are discussed in the next section, will then be determined globally by combining matching results on the epipolar lines.

Let L^l and L^r be a pair of corresponding epipolar lines from the topographic images. We have

$$L^l(i) = t^l \quad \text{and} \quad L^r(i) = t^r$$

for $i = 1, \ldots, nppl$ where *nppl* is the number of pixels per line, and t^l and $t^r \in$ {ridge, valley, convex hillside, concave hillside, saddle hillside, slope, flat}. Since the images are partitioned into arc and region segments, an image line is divided into conntected *arc-sections and region-sections*. For an epipolar line, let

n_a = number of arc-sections,
$S_{arc} = \{b_i \mid i = 1, \ldots, n_a\}$ = set of arc-sections,
n_r = number of region-sections,
$S_{reg} = \{c_i \mid i = 1, \ldots, n_r\}$ = set of region-sections.

Notice that each region-section corresponds to part of a region segment and is bounded by a pair of arc-segments except at the boundaries of the image. We thus have for $y \in S_{reg}$, the mapping REG: $S_{reg} \to R$, which gives the region segment to which y belongs. Correspondingly, for $x \in S_{arc}$, the mapping ARC: $S_{arc} \to A$ gives the arc segment to which x belongs.

Let BEGIN: $S_{reg} \to [1, nppl]$ and END: $S_{reg} \to [1, nppl]$ be functions which return respectively the beginning and ending locations of a region-section. We can then define a function, LARC: $S_{reg} \to \{A^r, 0\}$, which returns the arc segment to the immediate left of x. If BEGIN$(x) = 1$, LARC$(x) = 0$. Similarly, the function RARC: $S_{reg} \to \{A^r, 0\}$ returns the arc segment to the immediate right of x. If END$(x) = nppl$, RARC$(x) = 0$.

We can now define the set of all feasible matches for $x \in S^l_{reg}$ as

$$\begin{aligned} M^l(x) = \{y \mid\ & y \in S^r_{reg}, \\ & q(\text{REG}^l(x), \text{REG}^r(y)) = \textit{true}, \\ & p(\text{LARC}(x), \text{LARC}(y)) = \textit{true} \text{ or} \\ & p(\text{RARC}(x), \text{RARC}(y)) = \textit{true}, \\ & |\text{BEGIN}(x) - \text{BEGIN}(y)| < d_{max}, \\ & \text{and } |\text{END}(x) - \text{END}(y)| < d_{max}\} \end{aligned}$$

where d_{max} is a maximum allowable disparity. That is, $x \in S^l_{reg}$ and $y \in S^r_{reg}$ is a feasible match if

(1) the mean gray levels of the region segments to which the region sections belong are not too different,

(2) the mean orientations of the arc segments to the immediate left (or right) of the matching region sections are not too different, and

(3) the disparity of the end points of the matching region sections is less than a certain maximum allowable disparity.

For each possible match, its similarity measure is computed by correlating topographic structures within the corresponding region-sections. If we let $x \in S^l_{reg}$ and $y \in M^l(x)$, the lengths of x and y are given by

$$d^l = \text{END}(x) - \text{BEGIN}(x) \quad \text{and}$$
$$d^r = \text{END}(y) - \text{BEGIN}(y),$$

and the arc-sections adjacent to x and y are related by

$$p_1 = p(\text{LARC}(x), \text{LARC}(y)) \quad \text{and}$$
$$p_2 = p(\text{RARC}(x), \text{RARC}(y)).$$

It is worth noting that p_1 (or p_2) is *true* if the arc segments to the immediate left (or right) of x and y are matchable. In order to handle occlusion, we only require that one of the p's be *true*. There are three cases to be considered:

(1) if p_1 is *true* and p_2 is *false*, we compute a similarity measure which is proportional to the number of times that the topographic labels to the immediate right of LARC(x) and LARC(y) are compatible;

(2) if p_1 is *false* and p_2 is *true*, we compute a similarity measure which is proportional to the number of times that the topographic labels to the immediate left of RARC(x) and RARC(y) are compatible; and

(3) if both p_1 and p_2 are *true*, we compute a similarity measure which is proportional to the number of compatible topographic labels at both ends of the region-sections.

Formally, we compute a similarity measure $m(x, y)$ in the following way:

$$m(x, y) = \frac{1}{M'} \left[\sum_{k=0}^{m'-1} h(\text{BEGIN}(x) + k, \text{BEGIN}(y) + k) \right],$$

if $p_1 = \textit{true}$ and $p_2 = \textit{false}$,

$$m(x, y) = \frac{1}{M'} \left[\sum_{k=0}^{m'-1} h(\text{END}(x) - k, \text{END}(y) - k) \right],$$

if $p_1 = \textit{false}$ and $p_2 = \textit{true}$,

$$m(x,y) = \frac{1}{M'}\left[\sum_{k=0}^{m'/2-1} h(\mathrm{BEGIN}(x)+k, \mathrm{BEGIN}(y)+k) + \sum_{k=0}^{m'/2-1} h(\mathrm{END}(x)-k, \mathrm{END}(y)-k)\right],$$

otherwise,

where

$$M' = \max(d^l, d^r), \qquad m' = \min(d^l, d^r)$$

and

$$h(i,j) = \begin{cases} 1 & \text{if } L^l(i) = L^r(j), \\ 0 & \text{otherwise.} \end{cases}$$

Notice that $h(i,j)$ determines whether the topographic labels at the ith location of the first region-section and the jth location of the second region-section are the same. The motivation for using m as a similarity measure is that if the topographic structures around two potentially matching segments are similar, the likelihood that they really match is high.

We call y^* a *preferred match* of x if $m(x,y^*)$ is the largest among all possible matches. We also call LARC(y^*), the arc-section to the immediate left of y^*, a preferred match of LARC(x) if p_1 is *true*. Similarly, RARC(y^*) is a preferred match of RARC(x) if p_2 is *true*. In order to avoid ambiguous matches, no preferred match is assigned to x if $m(x,y^*)$ is smaller than some predetermined threshold.

3.4. Matching segments

At this point, each region (or arc) section in the epipolar line of one image is associated with a preferred matching section, if one exists, in the other image. The next attempt is to find global matches for the segments as a whole. Since each segment (region or arc) is decomposed into a set of subsections during the epipolar line matching process, we define SECTIONS(x) to be the set of all subsections of x, that is

$$\mathrm{SECTIONS}(x) = \{y \mid \mathrm{REG}(y) = x\} \quad \text{for } x \in R,$$

or

$$\mathrm{SECTIONS}(x) = \{y \mid \mathrm{ARC}(y) = x\} \quad \text{for } x \in A.$$

Since a preferred match is determined for each $y \in$ SECTIONS(x), we can define a mapping M^*: SECTIONS $\to R^r$ which gives the segment to which the preferred match of y belongs. The set of all possible matches for x is then

$$\mathrm{ALLMATCH}(x) = \{M^*(y) \mid y \in \mathrm{SECTIONS}(x)\}.$$

We now have for each segment in one image a set of potentially matching segments in the other image. In order to obtain the best of the potential matches, each possible match, $x \in R^l$ and $y \in$ ALLMATCH(x), is evaluated. Two criteria are used in determining the best matches for region segments:

(1) the number of times that y matches each of the region sections in SECTIONS(x); and

(2) the widths of region sections of which y is the perferred match.

Specifically, the similarity measure sm(x,y) is given by

$$\mathrm{sm}(x,y) = \frac{1}{|x|}\sum_{z \in \mathrm{SECTIONS}(x)} \mathrm{len}(z,y)$$

where $|x|$ denotes the size of x and

$$\mathrm{len}(z,y) = \begin{cases} \mathrm{END}(z) - \mathrm{BEGIN}(z) & \text{if } M^*(z) = y, \\ 0 & \text{otherwise.} \end{cases}$$

Notice that len(z,y) returns the length of z if y is a preferred match of z and returns zero otherwise. The best match is considered to be the one which gives the highest similarity measure. We also require that the similarity measure corresponding to the best match be greater than some predetermined value. A similar evaluation can be computed for matching arc segments, except that for an arc segment, its width or thickness does not play as important a role as for region segments.

A final global matching process is performed as an attempt to resolve ambiguity and to preserve consistent matches. To measure the degree of consistency between two sets of matches, we examine the region adjacency graphs of the matching segments. Ideally, if x matches y, then each neighbor of x should match some neighbor of y. So far, we have employed only the segment adjacency relation for the global matching process. A possible extension of the current scheme is to include relations such as parallel, antiparallel, and collinear.

In summary, the proposed stereo matching scheme is divided into three stages:

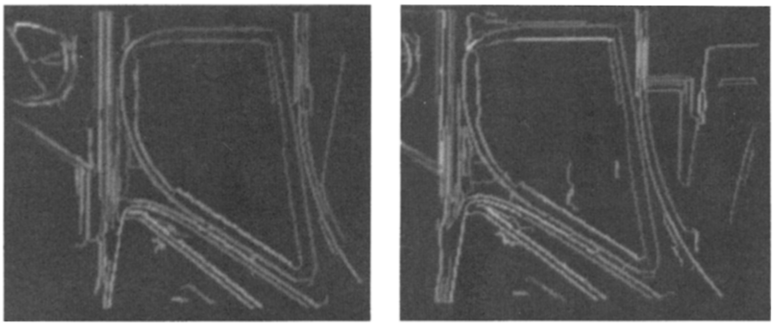

Figure 5. Matching results of the arc segments.

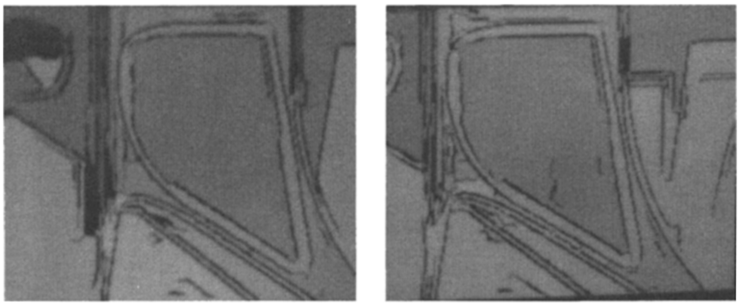

Figure 6. Matching results of the region segments.

(1) an epipolar line matching process which uses epipolar geometry to reduce the problem to a one-dimensional matching process;

(2) a segment matching process which establishes globally optimal matches for region and arc segments; and

(3) a global matching process which uses spatial relations among high-level structures to resolve ambiguous matches.

4. Results

Experiments have been performed on the stereo images (Figure 1) of the F-15 bulkhead. Figures 5 and 6 illustrate respectively the matching arc segments and matching region segments. Matching segments in the pictures are displayed with the same brightness. All matching segments have been identified correctly. It is interesting to note that horizontal arc segments were also matched correctly. Matching between horizontal arc segments was possible because segment adjacency information was incorporated in the global matching process.

Instead of using low level image features, we have presented here a stereo matching procedure based on high level structures such as arcs, regions and topographic structures. Results show that this is a promising step toward using high level features for stereo matching. For both the feature extraction and the stereo matching methods, we anticipate a great deal more work. We would need to carry out these methods for a large number of images. Only then will we be able to explore methods for generating a full disparity map accurately from the matched features.

References

[1] Arnold, R.D. (1983). Automated stereo perception. Stanford University, Ph.D. Dissertation, March 1983.

[2] Baker, H.H. and T.O. Binford (1981). Depth from edge and intensity based stereo. *Proc. 7th Intern. Joint Conf. Artificial Intelligence*, Vancouver, Canada, August 1981, 631–636.

[3] Barnard, S. and W. Thompson (1980). Disparity analysis of images. *IEEE Trans. Pattern Anal. Machine Intell.* 333–340.

[4] Burns, J.B., A.R. Hanson and E.M. Riseman (1984). Extracting straight lines. *Proc. 7th Intern. Joint Pattern Recognition*, Montreal, August 1984, 482–485.

[5] Canny, J.F. (1983). Finding edges and lines in images. Massachusetts Institute of Technology, Master Thesis, June 1983.

[6] Cheng, J.K. and T.S. Huang (1982). Image registration by matching relational structures. *Proc. 6th Intern. Joint Conf. Pattern Recognition*, 354–356.

[7] Davis, L.S. (1975). A survey of edge detection techniques. *Computer Graphics and Image Processing* 4, 248–270.

[8] Goshtasby, A. (1983). A symbolically-assisted approach to digital image registration with application to computer vision. Michigan State University, Ph.D. Dissertation.

[9] Gennery, D.B. (1977). A stereo vision system for an autonomous vehicle. *Proc. 5th Intern. Joint Conf. Artificial Intelligence*, MIT, Cambridge, MA, August 1977, 576–582.

[10] Grimson, W. (1981). *From Images to Surfaces*, MIT Press, Cambridge, MA.

[11] Haralick, R.M., L.T. Watson and T.J. Laffey (1983). The topographic primal sketch. *Intern. J. Robotics Research* 2 (1), 50–72.

[12] Haralick, R.M. (1984). Digital step edges from zero crossing of second directional derivatives. *IEEE Trans. Pattern Anal. Machine Intell.* 6 (1), 58–68.

[13] Lumia, R. (1982). A new connected components algorithm for virtual memory computers. *Proc. IEEE Conf. Pattern Recognition and Image Processing*, 560–565.

[14] Marr, D. (1976). Early processing of visual information. *Phil. Trans. Roy. Soc. London* B 275, 483–534.

[15] Marr, D. and T. Poggio (1976). A theory of human stereo disparity. *Science* 194, 283–287.

[16] Marr, D. and E.C. Hildreth (1980). Theory of edge detection. *Proc. Roy. Soc. London* B 207, 187–217.

[17] Medioni, G.G. and R. Nevatia (1985). Segment-based stereo matching. *Computer Vision, Graphics, and Image Processing* 31, 2–18.

[18] Moravec, H.P. (1977). Towards automatic visual obstacle avoidance. *Proc. 5th Intern. Joint Conf. Artificial Intelligence*, MIT, Cambridge, MA, August 1977, 584.

[19] Roberts, G. (1965). Machine perception of three-dimensional solids. In: J.T. Tippett et al., Eds., *Optical and Electroptical Information Processing*. MIT Press, Cambridge, MA, 179–197.

[20] Shapiro, L.G. and R.M. Haralick (1984). A hierarchical relational model for automated inspection tasks. *Proc. Intern. Conf. Robotics*, Atlanta, GA, March 1984, 70–77.