DEPTH FROM EDGE AND INTENSITY BASED STEREO

H. Harlyn Baker and Thomas 0. Binford
Artificial Intelligence Laboratory, Computer Science Department
Stanford University, Stanford, Calirornia, 94305

Abstract

The past *few years have seen* a *growing interest in the application" of three-dimensional image processing. With the increasing demand for 3-D spatial information for tasks of passive navigation[7,12], automatic surveillance[9], aerial cartography\l0,l3], and inspection in industrial automation, the importance of effective stereo analysis has been made quite clear. A particular challenge is* to *provide reliable and accurate depth* data *for input* to object *or terrain modelling systems (such* as [*5*]. *This paper describes an algorithm for such stereo sensing It uses an edge-based line-by-line stereo correlation scheme, and appears to be fast,* robust, and parallel implementable. *The processing consists of extracting edge descriptions for a stereo pair of images, linking these edges to their nearest neighbors* to obtain the edge connectivity structure, correlating the edge descriptions on the basis *of local edge properties, then cooperatively removmg those edge correspondences determined to be in error - those which violate the connectivity* structure of the two images. *A* further correlation *process, using a technique similar to that used for the edges,* is *applied to the image intensity values over intervals defined by the previous correlation The result of the processing is a full* image array disparity map *of the scene viewed.*

Mechanism and Constraints

Edge-based stereo uses operators to reduce an image to a depiction of its intensity boundaries, which are then correlated. *Area-based* stereo uses area windowing mechanisms to measure local statistical properties of the intensities, which can then be correlated. The system described here deals, initially, with the former, *edges,* because of the:

a) reduced combinatorics (there are fewer edges than pixels),

b) greater accuracy (edges can be positioned to sub-pixel precision, while area positioning precision is inversely proportional to window size, and considerably poorer), and

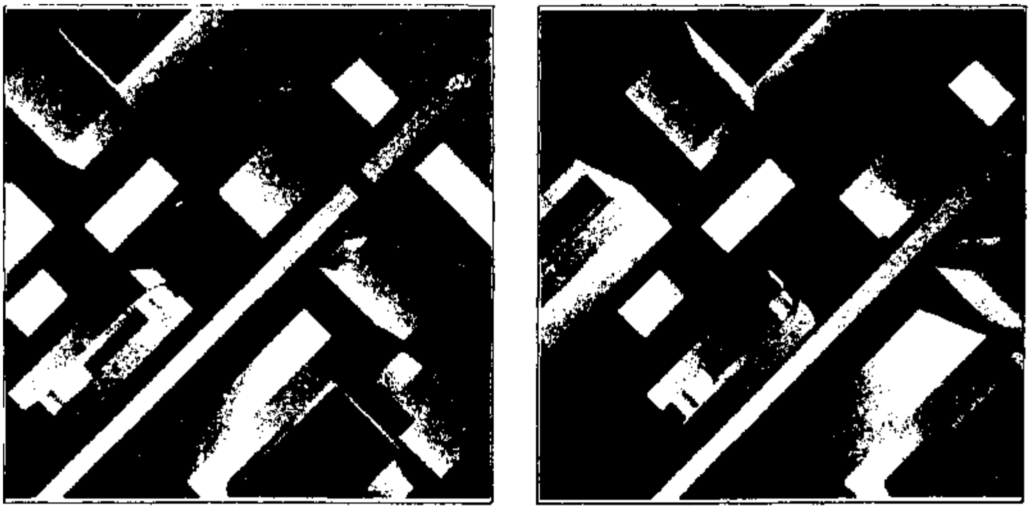
c) more realistic in variance assumptions (area-based analysis presupposes that the *photometric* properties of a scene arc invariant to viewing position, while edge-based analysis works with the assumption that it is the *geometric* properties that are invariant to viewing position).

Edges are found by a convolution operator They are located at positions in the image where a change in sign of second difference in intensity occurs. A particular operator, the one described here being 1 by 7 pixels in size, measures the directional first difference in intensity at each pixel' Second differences are computed from these, and changes in sign of these second differences are used to interpolate sero crossings (i.e. peaks in first difference). Certain local properties other than *position* are measured and associated with each edge - *contrast, image slope,* and *intensity to either side* - and *links* are kept to nearest neighbours above, below, and to the sides. It is these properties that define an edge and provide the basis for the correlation (see the discussions in [1,2]).

The correlation is & search for edge correspondence between images Fig. 2 shows the edges found in the two images of fig. 1 with the second difference operator (note, all stereo pairs in this paper are drawn for cross-eyed viewing) Although the operator works in both horizontal and vertical directions, it only allows correlation on edges whose horizontal gradient lies above the noise - one standard deviation of the first difference in intensity With no prior knowledge of the viewing situation, one could have any edge in one image matching any edge in the other.

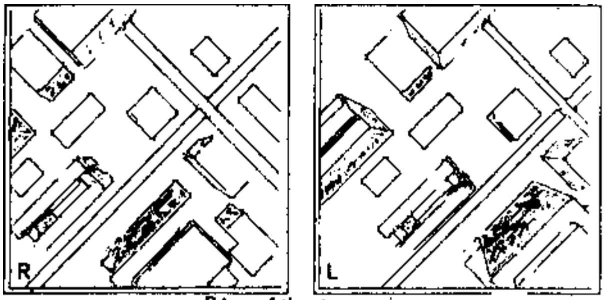
By constraining the geometry of the cameras during picture taking one can vastly limit the computation that is required in determining corresponding edges in the two images. Consider fig. 3. If two balanced, equal focal length cameras are arranged with axes parallel, then they can be conceived *of* as sharing a single common image plane. Any point in the scene will project to two points on that joint image plane (one through each of the two lens centers), the connection of which will produce a line parallel to the baseline between the cameras. Thus corresponding edges in the two images must lie along the tame line in the joint image plane This line is termed an *epipolar* line. If the baseline between the two cameras happens to be parallel to an axis of the cameras, then the correlation only need consider edges lying along corresponding lines parallel to that axis in the two images. Fig. 3 indicates this camera geometry - a geometry which produces *rectified*

The edge operator is simple, basically one dimensional, and is noteworthy only in that it it fast and fairly effective.

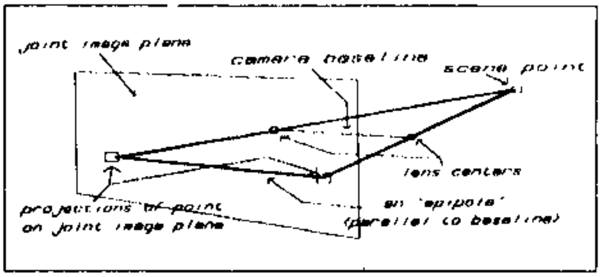


A stereo pair of images (from Control Data Corporation) **[256 × 256 × 6]**
Figure 1

images. The algorithm described assumes the stereo pair to be rectified, and if this is not the case then the appropriate transformation of one image relative to the other mutt be made before further processing is done. Note that a less restrictive solution would be to have the correlation informed of the camera geometries, and have it solve for the more general epipolar situation.



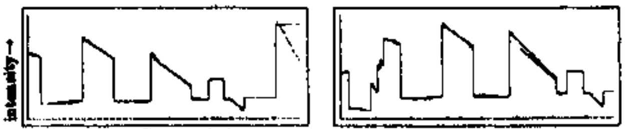
Edges of the stereo pair
Figure 2


joint image plane
camera baseline
scene point
lens centers
projections of point on joint image plane
an 'epipole' (parallel to baseline)

Two Camera Epipolar geometry
Figure 3

Fig. 4 shows a pair of corresponding lines from the stereo pair of ig. 1, while fig. 5 plots the actual intensities of these lines, seen as dots, superimposed on the edges determined by the edge operator. Since one needs to compare edges only from corresponding lines in the two images, the correlation can be applied to the edges shown in fig. 5. The process of correlating could proceed at this point by searching for the 'best' assignment of edges along such corresponding lines - one which optimises some goodness measure. However, normal combinatoric search is quite inadequate here. Typical lines have upwards of n — 30 edges each and the combinatory space, with a naive upper bound of n!, grows rapidly with n (the CDC imagery above isn't typical, but rather is synthetic, and actually quite noise-free - in contrast sec the images of fig. 10). Even with extensive heuristic pruning, runs with a combinatoric search approach often take seconds per line ... and sometimes minutes (on a DEC KL-10). A superior approach to the correlation task lies in using the Viterbi algorithm [6|, a dynamic programming technique used extensively in speech processing, and first used in vision research in some recent work at Control Data Corporation(9].

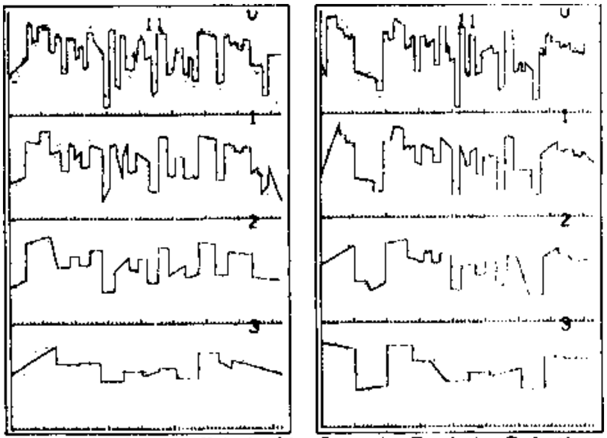
Right and Left Image Corresponding Epipolar Line Intensities
Figure 4



Edges of these lines with intensities marked
Figure 5

What distinguishes the Viterbi technique from normal search is the requirement that one be able to partition the original problem into two subproblems, each of which can be solved optimally and whose results can be processed to yield a global optimum for the original problem ('optimal' with respect to an evaluation function on the chosen parameters). In a recursive way, each of the subproblems may be divided and the solution process repeated Geometrically, the partitioning constraint here is one of *monotomcity* of edge order A left-right ordering of edges in one image cannot correspond to a right-left ordering in the other, i.e. there can be no positional reversals of edges in the image plane This constraint allows one to make n tentative assignments of an edge on one line with the edges of the corresponding line in the other image, with each tentative assignment partitioning the correspondence problem into two subproblems The two subproblems are the matching of edges lying to the left and right of the selected edge on the one line with edges lying to the left and right respectively of its tentative match on the other line The optimality criterion selects the series of such assignments judged 'best'. This constraint excludes from analysis, for the time being, features such as *WIVES* or *over hanging surfaces* which lead to positional reversals in the image. Experience suggests that this reversal also causes the human vision system trouble - we can fuse one or the other, the nearer or the further, but not both at the same time.

The correlation could use the edges as indicated in fig. 2 above, but in the interests of robustness and efficiency a different approach is taken here. A large amount of small scale detail in the images will exponentially increase the cost of the correlation Reducing the level of detail and narrowing the extent of the required search will reduce the computation time and enhance noise immunity This is achieved through the use of a *coarse to fine* analysis in which a reduced resolution correlation is first applied to bring the two images into rough correspondence. The removal of the small scale detail brings quite a reduction in the number of edges to be dealt with Successive refinements in resolution bring successively finer detail into the analysis, and each such phase can use the results of the previous lower resolution analysis to narrow its search Such an approach has had previous successful application in visual processing (e.g. (8,11,12)), and has relevant ties to the neurophysiology of vision, where it is felt a multiple spatial frequency analysis is part of the human system's processing |14] (although the filtering used here is low pass, and not bandpass) It was our intent to use the low resolution components of the images to determine local approximate *disparities,* and to use these as guides for the full resolution correlation To obtain the resolution reduction we use a linear smoothing filter to successively halve image resolutions, continuing until the image signal content reaches an acceptable level (smoothing reduces noise, so increases signal-to noise ratio) Fig. 6 shows the edges in successive resolution reductions of a sample line pair from the images of fig. 10, again with dots marking the intensities



Right and Left Image Epipolar Line Successive Resolution Reductions
Figure 6

The tame basic second difference operator is used throughout the resolution reduction analyses, but its size and noise-based thresholds are altered to keep it matched to the characteristics of the 'new' reduced resolution image. These lowest resolution edges are correlated in a manner to be described below. The *intervals* specified between nearest-correlated edge pairs and their mates in the other image define local disparities to be used by the full resolution correlation process.

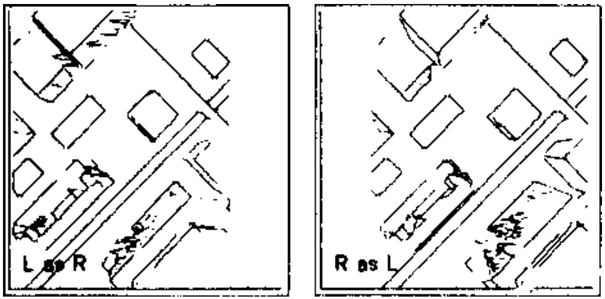
The Correspondence Process

The process of determining edge correspondences is basically the same for both the reduced resolution and the full resolution correlations; the only difference is in the set of parameters used by the optimization function. Full resolution correlation uses *edge angle, side intensities, relative disparity* (as measured by the reduced resolution phase), and *interval compression* implied by the correspondence (which uses edge position to determine the foreshortening of scene surfaces |4|). In reduced resolution correlation, *side intensities, contrast,* and *interval compression* arc used These parameter measures all enter the computation as probabilistic weightings: $0 < P < 1$.

Each edge from a line in the reference image is associated with a set of possible matching edges from its corresponding line in the other image (including the null match, which would imply that the edge is either spurious or is obscured in the other image) Slightly complicating this, each such edge in treated as a doublet, being a left side (the termination of the interval to its left) and a right side (the start of the interval to its right); left sides of edges can only match left sides of edges, right sides only right sides (quite obviously) This left-right distinction is essential in domains where surfaces may occlude one another, leaving a surface to one side of an edge hidden from one viewpoint while it is visible to the other. Each side of an edge is termed a *half-edge.* For each pairing in the set of possible matches the *static* probability of correspondence is determined - this is the product of all of the mentioned probability measures except *interval compression* (which is determined dynamically) The optimization process then uses these probabilities, composing them with the dynamic *interval compression* probabilities in determining the 'best' correspondence of half-edges along a line in one image with half-edges along the corresponding line in the other image (details in (3j). This computation is $O(n^3)$ for n edges along either image line - it would be $O(n^2)$ were it not for the use of *interval compression* probabilities.

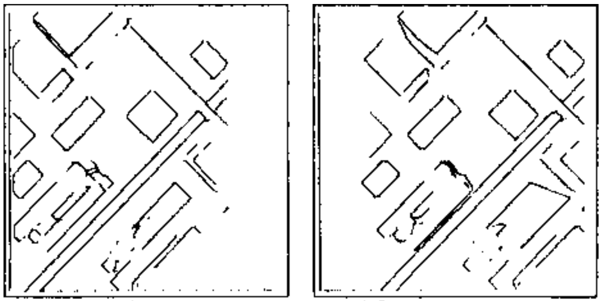
The Connectivity Constraint

Fig 7 shows the results of the whole image line-by-line correlation Wherever there is a noticeable horizontal jag across the image, there is an error in the correlation What is really being depicted here is *change in disparity* along connected edges in each image. This is achieved by plotting between the connected edges of an image, but rather than using just each edge's coordinate, use its coordinate plus associated disparity. Thus, when a connected stretch of edges in one image is matched to various parts of the other image, the drawing will jump horizontally back and forth in the other image's space, touching the various parts correlated with the connected stretch. At these horizontal jumps, the process is suggesting that there is a large change in depth. This is a suggestion of a *break in depth continuity.*

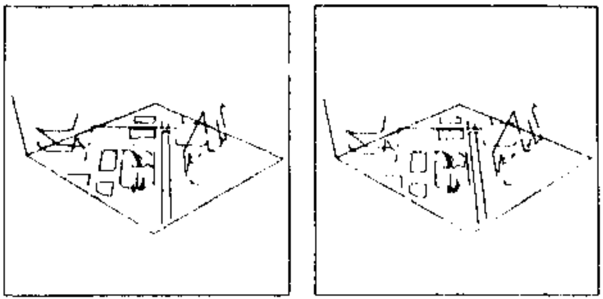


Preliminary Correlation results
Figure 7

Let us emphasise, the dynamic programming algorithm above performs a local optimisation for the correlation of individual lines in the image - it uses no information outside of those lines. A very strong global constraint is apparent and available here, that of *edge connectivity.* It may be presumed (by general position) that, in the absence of other information, a connected sequence of edges in one image should be seen as a connected sequence of edges in the other, and that the structure in the scene underlying these observations may be inferred to be a continuous surface detail or a continuous surface contour. A cooperative procedure uses this connectivity assumption to remove edge correspondences which violate surface continuity. The evidence for these miscorrelations is found through the tracking of disparities along connected edges on adjacent lines (this is what fig. 7 depicts). The results of the correlation after this process has functioned are shown in fig. 8 (with the same type of depiction as fig 7). Fig. 9 shows a perspective view of these connected edge elements in depth. Notice that this figure drawn by the program shows that it has truly captured something of the 3-D structure of the scene.



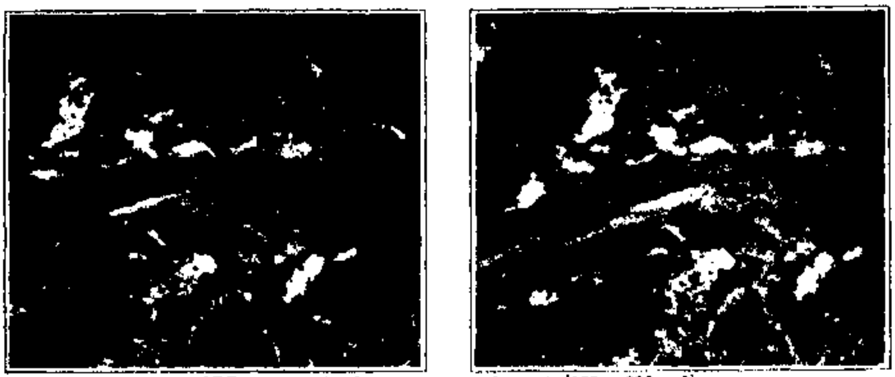
Final (post-connectivity constraint) Correlation results
3000 half-edge correlate pairs
Figure 8



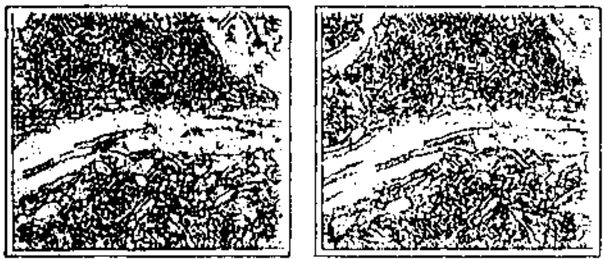
Perspective view of connected edge elements
(the z axis is disparity, not depth)
Figure 9

Figs 11 through 14 show the various stages of the edge-based stereo correlation process when applied to the image pair of fig. 10. This data (provided by the Night Vision Laboratory of the U.S. Army), an aerial view of natural terrain, is considerably noisier and significantly more detailed than the synthetic urban imagery of fig. 1. It more clearly demonstrates the importance of the interline connectivity constraint and, as shown in fig. 6, the use of resolution reduction during the correlation.

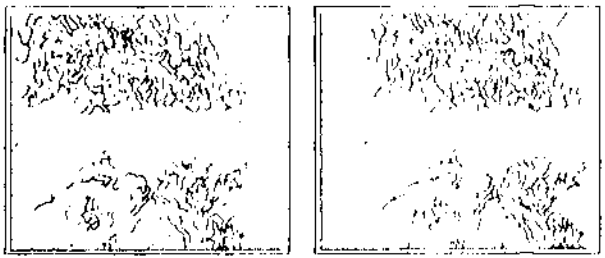
The description so far has been of an edge-based stereo correlation scheme - one which uses a Viterbi optimality condition and a cooperative continuity enforcement process in establishing reliable correspondences between the intensity boundaries, or edges, of a stereo pair of images. Yet there is much more information one could have about scenes such as those depicted in figs. 1 and 10 The edge descriptions of figs. 9 and 14 just highlight the structure of the scenes, providing rather sparse disparity measures. One would like fuller stereo detail from the correlation, and a subsequent correlation process, this time based on image intensity values, supplies this.



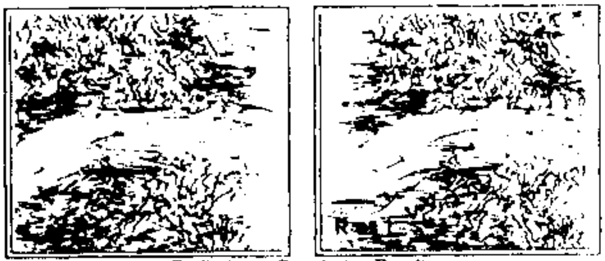
NVL stereo pair of images · natural terrain [168 × 200 × 9]
Figure 10



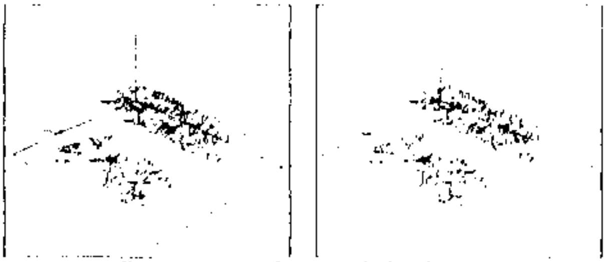
Edges of the stereo pair
Figure 11



Final (post-connectivity constraint) Correlation results
3700 half edge correlate pairs
Figure 13



Preliminary Correlation Results
Figure 12

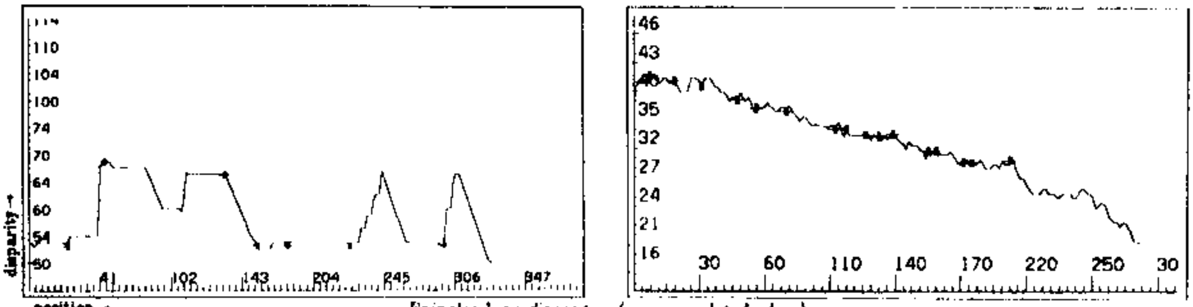


Perspective view of connected edge elements
Figure 14

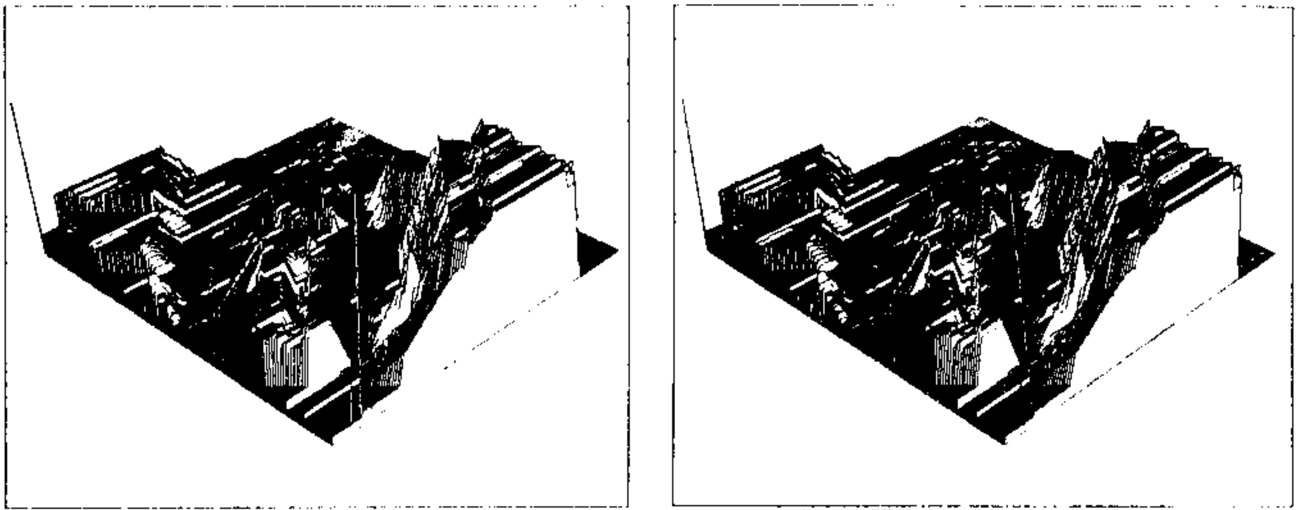
Intensity Bated Correlation

The above edge-based correlation, in indicating corresponding edges in the two images, provides strong local 'vergence' information which greatly constrains the matching problem for any remaining correlation. One can take unpaired edges from an interval along one line (epipolar line) and correlate them (again, via Viterbi) with unpaired edges from the corresponding interval of the same line in the other image (the intervals are bounded by either correlated edge pairs or the periphery of the image). This correlation serves to "fill in the gaps" of the primary edge-based correlation. A final correlation (the fourth!) takes intensity values from the intervals between paired edges along corresponding lines and does yet one more Viterbi correlation on them We are still developing the metrics used in this correlation - at the present we use 1) intensity variance and 2) deviation from linearity of the interpolated surface Fig 15 indicates the results of this procetsmg on two sample lines (CDC line to left, NVL to right) Arrowheads (-> and <--) show half edge pairings (to the *left* side and to the *right* side, respectively), and the plotted contours show the disparity values assigned by the intensity correlation (depth can easily be determined from disparity once the camera parameters are known).

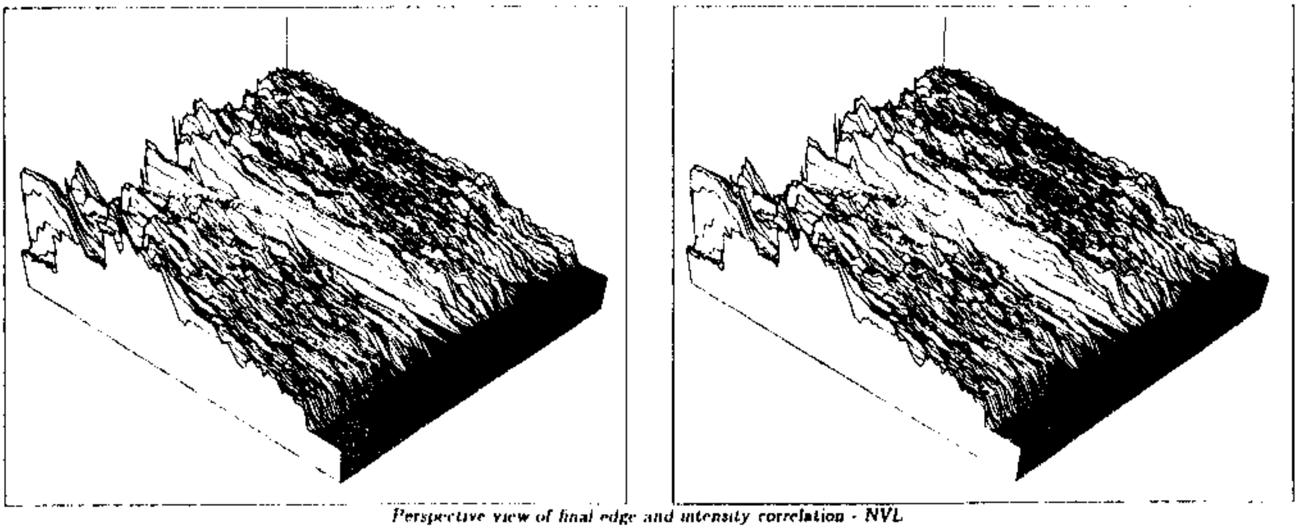
Figs. 16 and 17 show stereo perspective views of the full correlation results for the two sets of imagery (disparities are those of the left camera image), and fig 18 shows single views of the same surfaces after convolution with a 4 X 4 smoother (for easier viewing, although with less surface clarity) These figures show that the correlation algorithm produces a full image array disparity map of the viewed scenes. We have not yet measured its accuracy.



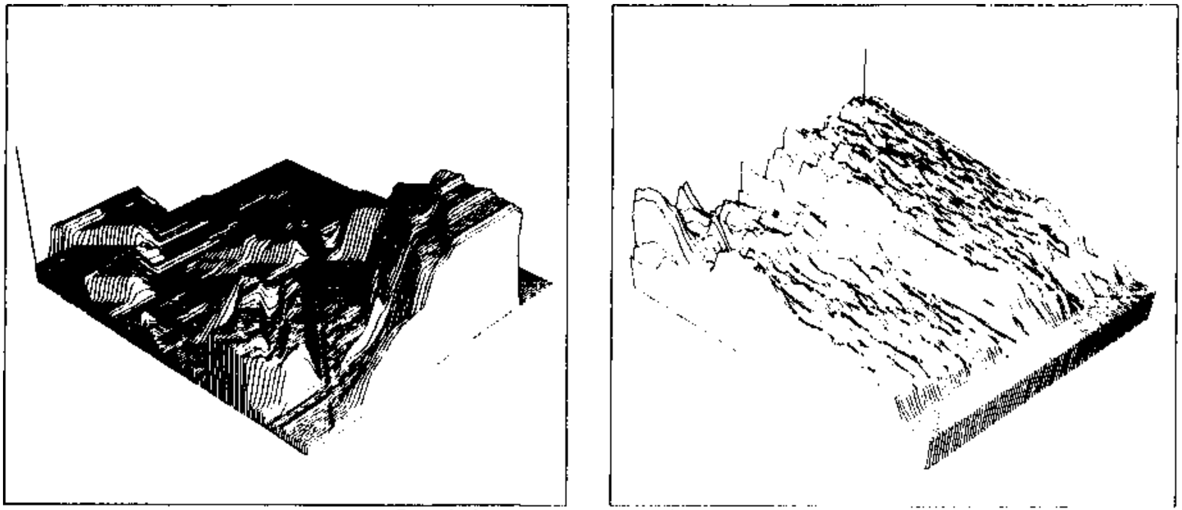
Epipolar Line disparities (+ + correlated edges)
Figure 15



Perspective view of final edge and intensity correlation - CDC
Figure 16



Perspective view of final edge and intensity correlation - NVL
after median filtering
Figure 17



Smoothed versions of final plots (figs. 16 & 17)
Figure 18

Performance

The correlation algorithm described here provides this three-dimensional sensing in a *fast, robust,* and *parallel implementable* way.

[fast] The analyses as shown in figs. 9 and 14 took roughly 25 and 35 seconds apiece from image input to the final edge correlation results as depicted. The remaining edge and intensity correlations of figs. 16 and 17 required a further 45 and 15 seconds, respectively (on a KL-10).

(robust) The use of line-by-line correlation, each processed independently (accumulating a good deal of redundant evidence), and the use of a coarse-to-fine strategy (where the more reliable lower frequency components are correlated first) give a good basis for obtaining the correct global consensus in the subsequent cooperative process.

[parallel tmplementable] Since there is no interline dependence during the various correlations, and the subsequent cooperative process has only pairwise interline interactions, there is a high potential for a parallel (n-processors for n lines) realisation.

Ahead

There is still, of course, considerable research to be done within this depth determination process. We will be:

- looking into improvement issues in the present algorithm, such as having the correlator use colour information,
- testing it on further stereo imagery, both aerial and near range, using digital terrain models for accuracy tests, where possible, and
- studying the applicability of the disparity measures derived by the algorithm to surface and object modelling (as in ACRONYM [5]).

Further afield, we are interested in developing such a correlation scheme into a continuous, multi-image correlator, capable of integrating analyses over a series of passively sensed stereo views in building a highly accurate and detailed map of scene depth.

Acknowledgements

This research has been supported by the Defense Advanced Research Projects Agency on contract MDA 903-80-C-0102. Special thanks to Sid Liebes and Peter Blicher for their comments on a draft of this paper.

References

|1] Arnold, R.D and T.O. Binford, 'Geometric Constraints in Stereo Vision," *Proc SPIE Meettng,* San Diego, California, July 1980.

[2J Baker, H Harlyn, 'Edge Based Stereo Correlation,' *Proc ARPA Image Understanding Workshop,* University of Maryland, April 1980, 168-175

|3] Baker, H. Harlyn, 'Depth from Edge and Intensity Bated Stereo,' *(forthcoming Ph D. thesis)*

(4) Blakemore, Colin, 'A New Kind of Stereoscopic Vision,' *Vision Research,* Vol 10, 1970, 1181-1199.

|5] Brooks, Rodney A., 'Symbolic Reasoning Among 3-D Models and 2-D Image*,' *Artificial Intelligence Journal,* Vol. 16, 1981

(6) Forney, G David Jr., 'The Viterb. Algorithm,' *Proc IEEE,* Vol 61, No 3, March 1973

[7] Gennery, Donald B , 'Modelling the Environment of an Exploring Vehicle by Means of Stereo Vision,' *Stanford AJ Lab Memo AIM 339,* June 1980.

(8) Crimson, W E. L.,'Computing Shape Using a Theory of Human Stereo Vision,' *Department of Mathematics, MIT* (thesis), June 1980

[9] Henderson, Robert L , Walter J Miller, C B Crosch, 'Automatic Stereo Recognition of Man Made Targets,' *Soc Photo Optical Instrumentation Engineers,* Vol 186, August 1979.

(10] Kelly, R., P McConnell and S Mildenberger, 'The Gestalt Photomapping System,' *Journal of Photogrammetric Engineering and Remote Sensing,* Vol. 43, November 1977, 1407-1417.

[II] Marr, D. and T. Poggio, 'A Theory of Human Stereo Vision,' *MITAI Memo No. 451,* November 1977.

[12] Moravec, Hans P., 'Obstacle Avoidance and Navigation in the Real World by a Seemg Robot Rover,' *Stanford AI Lab Memo AIM 340,* September 1980.

[13] Panton, Dale J., 'A Flexible Approach to Digital Stereo Mapping,' *Photogrammetnc Engineering and Remote Sensing,* Vol 44, No. 12, December 1978, 1499-1512.

[14] Wilson, Hugh R., 'Quantitative Prediction of Line Spread Function Measurements: Implications for Channel Bandwidths,' Vision *Research,* Vol 18, 493-496.