SRI International

REIMPLEMENTATION OF THE STANFORD STEREO SYSTEM

INTEGRATION EXPERIMENTS WITH THE SRI BASELINE STEREO SYSTEM

Technical Note No. 431

February 1989

By: H. Harlyn Baker

Artificial Intelligence Center
Computer and Information Sciences Division

The research reported herein was supported by DARPA Contracts MDA903-83-C-0027 and DACA76-85-C-0004, and the Advanced Technology and Applications Division of Boeing Computer Services.

SRI International®

333 Ravenswood Ave. • Menlo Park, CA 94025
(415) 326-6200 • TWX: 910-373-2046 • Telex: 334-486

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **FEB 1989**	2. REPORT TYPE	3. DATES COVERED **00-02-1989 to 00-02-1989**

4. TITLE AND SUBTITLE **Reimplementation of the Stanford Stereo System Integration Experiments with the SRI Baseline Stereo System**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **SRI International,333 Ravenswood Avenue,Menlo Park,CA,94025**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**13**	

REIMPLEMENTATION OF THE STANFORD STEREO SYSTEM

—

INTEGRATION EXPERIMENTS WITH THE SRI BASELINE STEREO SYSTEM*

H. Harlyn Baker
Senior Computer Scentist
Artificial Intelligence Center
SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025.

Abstract

We describe experiments in stereo matching using a Lisp Machine implementation of the Baker stereo system developed at Stanford University. The processing is one of edge matching in a hierarchy of long to short image contours, finishing with interedge intensity correlation to yield a dense map of scene disparities. An experiment and the results obtained in coupling this with the SRI STEREOSYS mapping system are presented.

*This research was supported by DARPA Contracts MDA 903-863-C-0027 and DACA 76-85-C-0004, and the Advanced Technology and Applications Division of Boeing Computer Services. This research was performed in 1985.

1. FEATURE BASED STEREO SYSTEM

This report describes a Symbolics 3600 reimplementation of the Baker stereo mapping system from Stanford University. The details of the system have been described earlier (see publications [1 – 4]). The original version, written in a mix of SAIL and assembly language, ran on a DEC-10. Generally, the system operates iteratively on first edges (zero crossings in difference-of-Gaussian (DOG) images), then on pixels (*i.e.*, the intensity values themselves), using a dynamic programming optimization technique. The processing operates on corresponding epipolar lines. This reimplementation effort had two purposes: first, to bring up the stereo system in a language and environment that could serve as the basis for further research, integration, and development; second, to experiment during reimplementation with certain alterations in the control structure and the matching algorithm.

To explain its operation, I shall describe the processing sequence of a stereo pair, pointing out, wherever appropriate, any differences between the new implementation and its predecessor.

2. THE PROCESSING

Image edges are detected at a particular scale (Gaussian standard deviation σ). Figure 1 shows the stereo pair used (termed the I5 data set); Figure 2 shows the edges obtained for this pair at the chosen σ value. These edges are transformed by the known camera relationships so that edges on corresponding epipolar lines have the same ordinate (as described in an earlier report [5]); and these are depicted in Figure 3. Matching begins on these edges with no initial disparity constraints, and takes first those contours having the greatest extent (beyond 2σ in the contour extent distribution of the image as a whole).

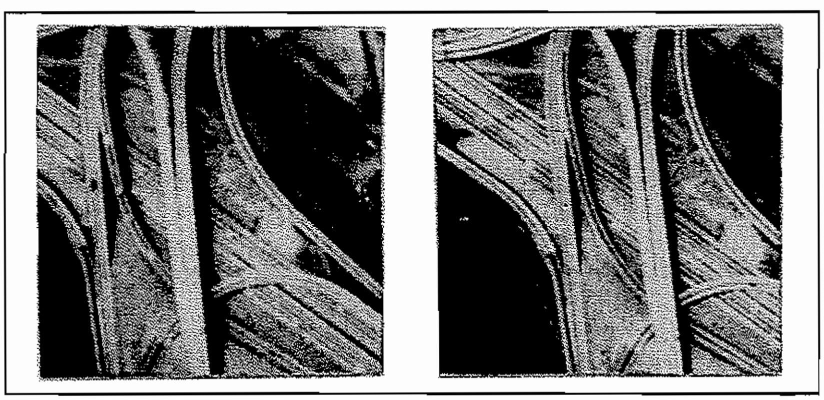

Figure 1. I5 Data Set

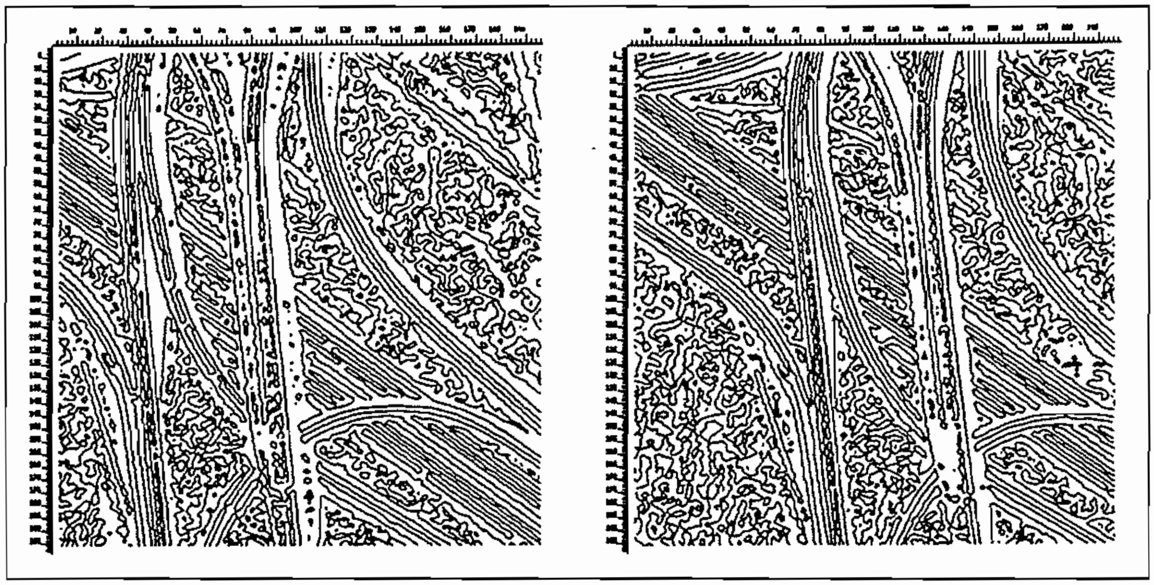

Figure 2. I5 Edges

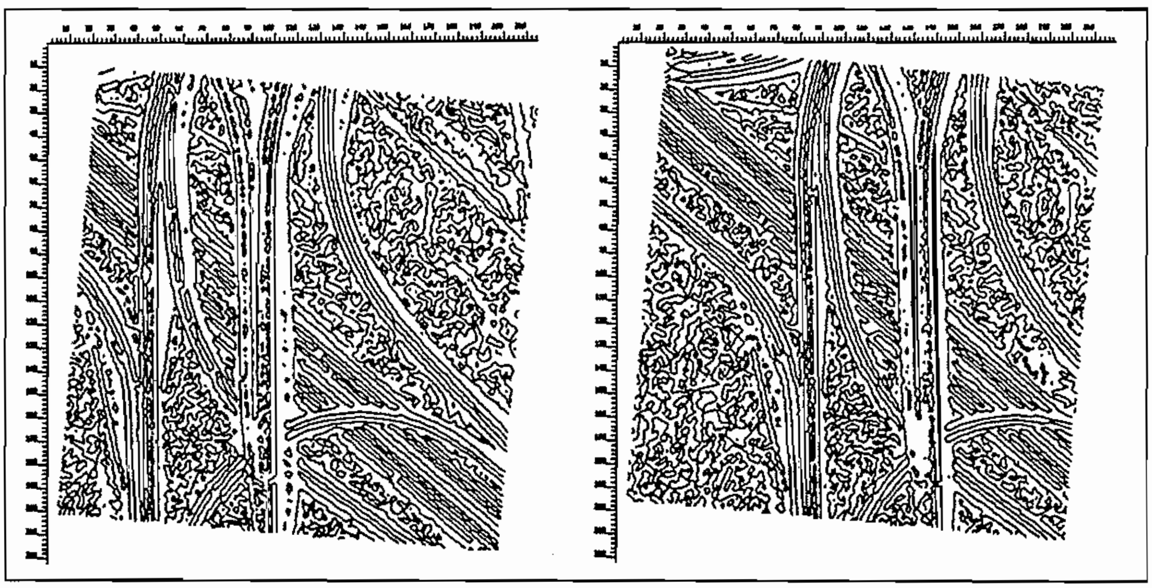

Figure 3. Edges in Epipolar Registration

Underlying both implementations was the philosophy of controlling matching by considering the strongest features first and then, in subsequent iterations, allowing increasingly weaker features to be introduced. "Strength'" was to be a measure of a feature's *significance* – the ease, distinctness, and accuracy that would characterize its matching. The earlier system progressed in a range from lower to higher spatial resolution, operating on the premise that power in the frequency domain was a good measure of each feature's significance in the spatial domain. The more recent implementation, which was not meant to replace the former, just to provide another perspective, operated on the principle that an edge that was part of a larger structure in the spatial image was more likely to be easily matched and therefore could be matched more reliably than one that was more isolated. The criterion here for "being part of a larger structure" is the edge's *extent*, namely, the distance (in epipolar lines) between the extrema along its connected contour. Figure 4 shows the results of matching edges on these initial major contours.

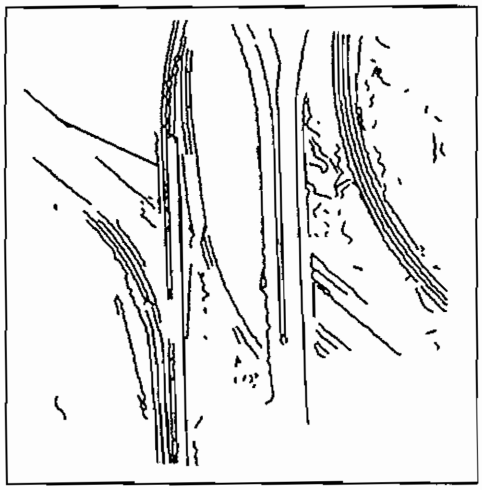

Figure 4. Initial Match Results

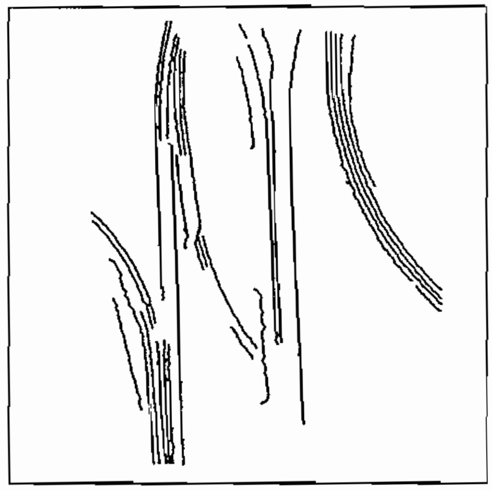

Figure 5. Matches After Preliminary Filtering

Statistics of the matched edges are accumulated, disparity outliers are discarded, and contours of lesser extent (in the lower 30%) are discarded. Figure 5 shows the edge matches left after this filtering. An important distinction to be noted is the manner in which edges are considered. Each is treated as a doublet, a left and a right side, and the matching allows these to be put into correspondence independently. Such a treatment increases the computational load, but allows for proper consideration of occlusions, where only one side of the edge relates to a physically identifiable point on a surface, *i.e.,* the one that is on the occluding side of the boundary.

The disparity constraints determined from this first pass apply to the next iteration, in which contours of smaller extent are included. Figure 6 shows the increased set of matches after this second matching stage. The iteration continues, and at each stage contours of smaller extent are introduced and outliers are discarded. Figure 7 shows the next level of this iterative processing. When certain termination criteria are met (that is, when all contours have been considered and only some minimal number of new matches is added between iterations), the edge matching terminates. Figure 8 shows the final set of matched edges.

Let me proceed through another data set, one of some familiar natural terrain (the ETL data set). Figure 9 shows the imagery, while Figure 10 shows the edges in correct epipolar registration. The next three figures show the successive matching iterations, resulting in the edge correspondences of Figure 14.

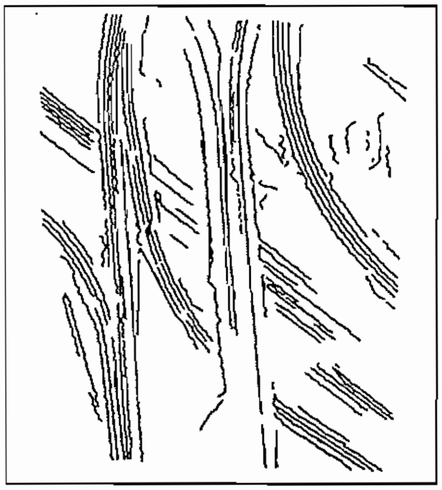

Figure 6. Matched Edges

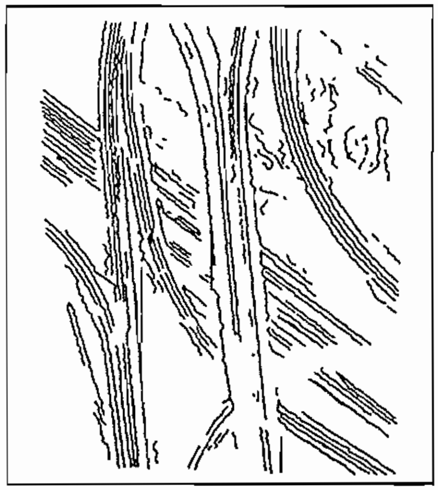

Figure 7. Matched Edges

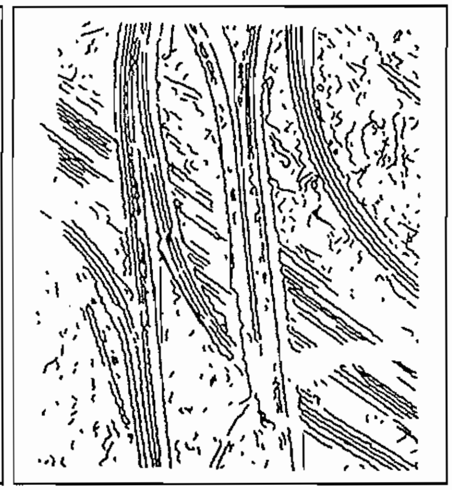

Figure 8. Final Edge Match Results

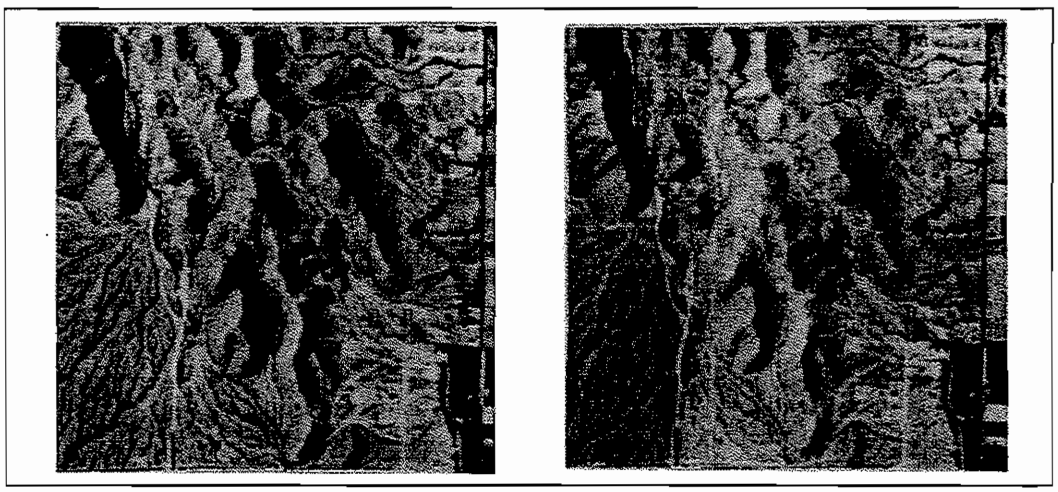

Figure 9. ETL Data Set

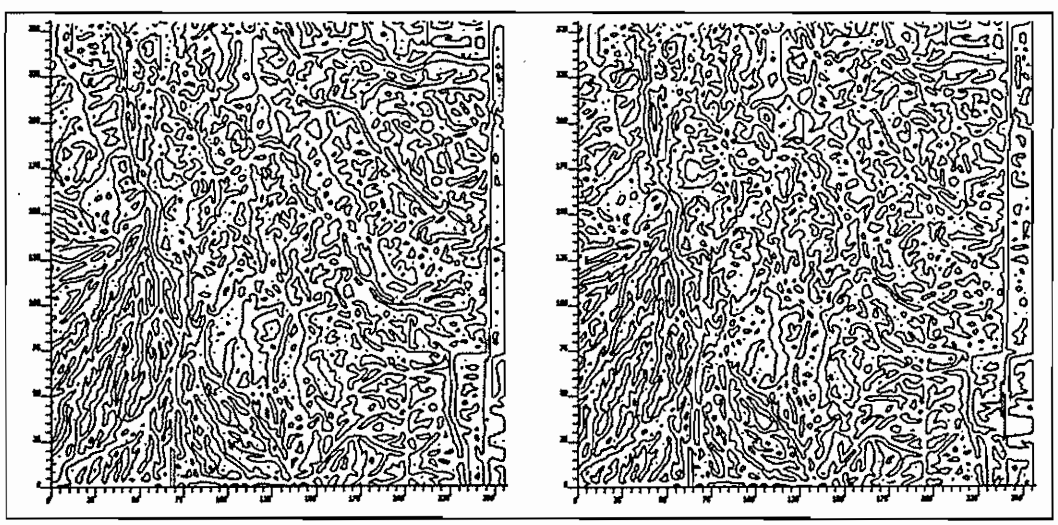

Figure 9. ETL Edges

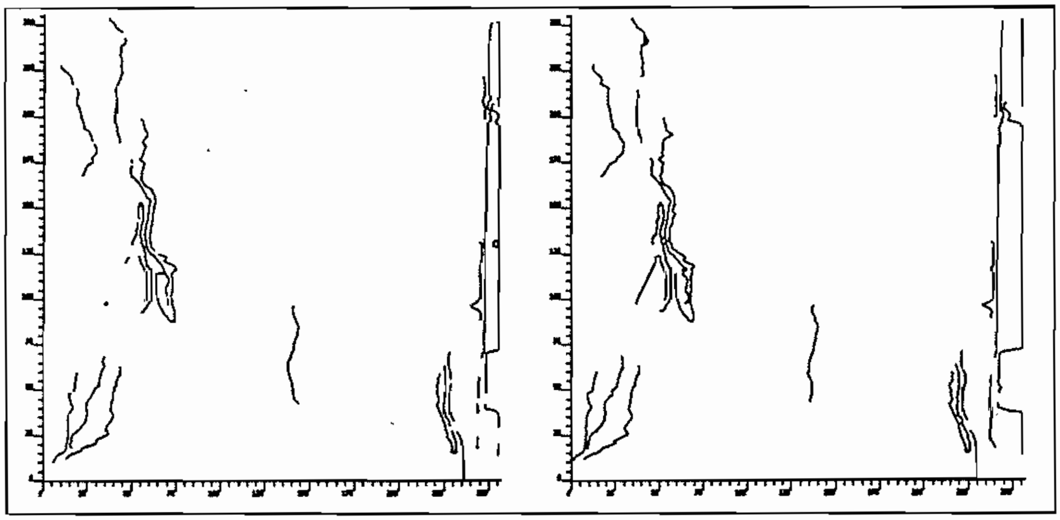

Figure 11. Matches After First Pass

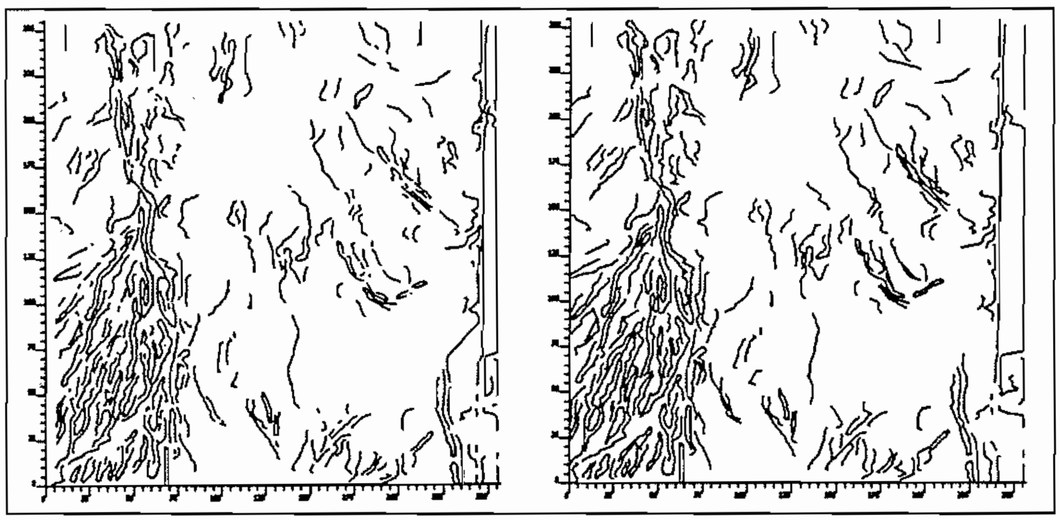

Figure 12. Second Pass

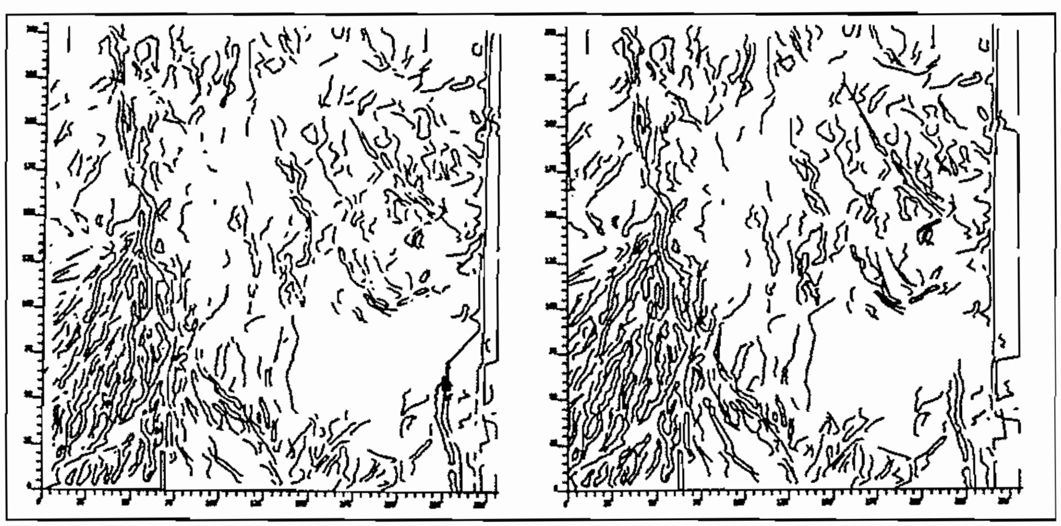

Figure 13. Third Pass

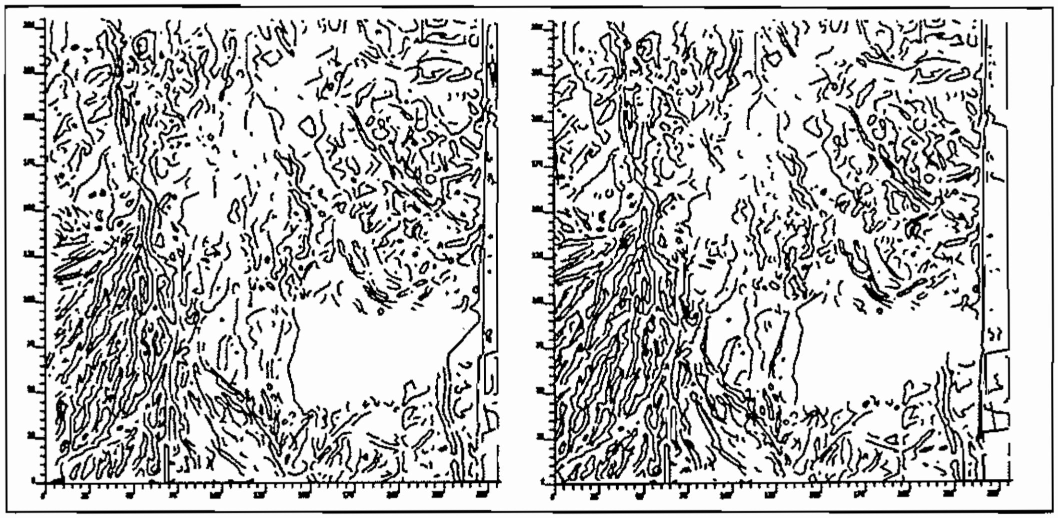

Figure 14. Final Edge-Match Results

The results to this point are a set of edges matched across the two views. This accounts for approximately 10% of the image's pixel area. To provide a more complete set of match points between the images, we perform a further matching similar to that described above, except that image *intensities* are used instead of image *edges* (details are available in [2]). This process applies the constraints provided by the final edge-match results, considers the occlusion clues implied, and employs a similar dynamic programming optimization. Results of the edge and intensity matching are shown in Figure 15.

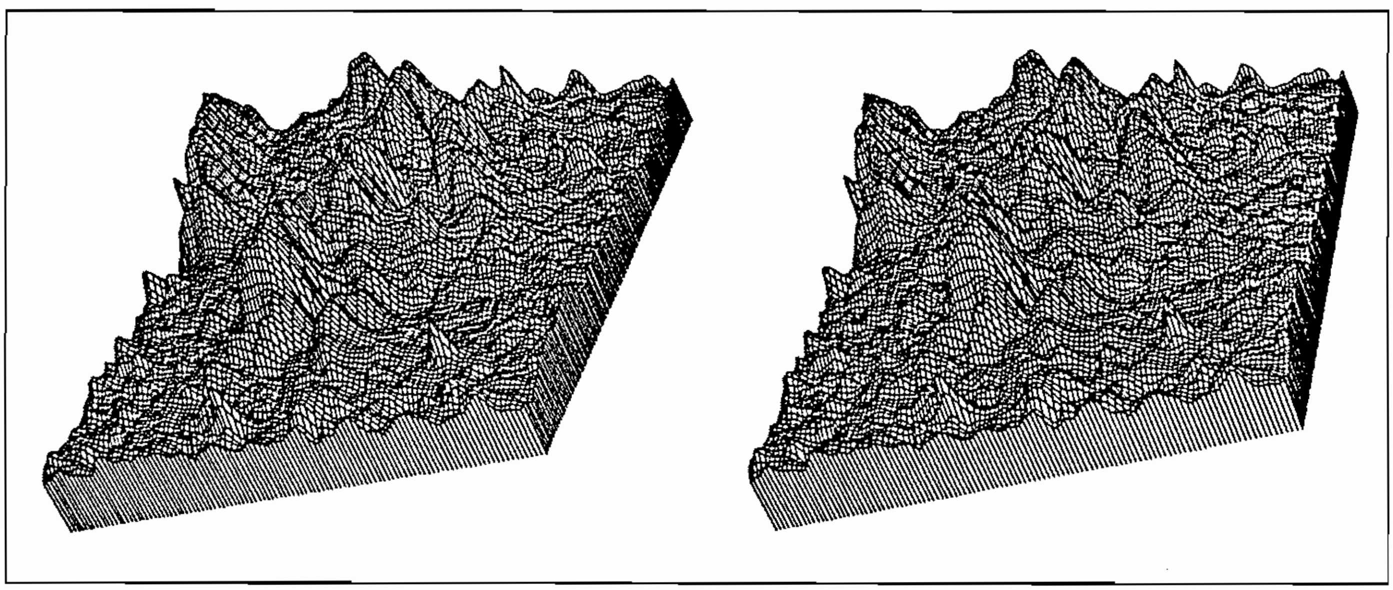

Figure 15. Disparity Plot for Final Edge and Intensity Match Results

Of these matches, 96% agreed with the DIMP results, with the majority of the missing matches being near the ends of contours where zero-crossing positions tend be least stable. A major problem encountered with this data set was due to the poor quality of the photographs: the right image was significantly lighter than the left in a band down its right hand side. The area-based approach, in which normalized correlation was employed, was not impeded by this, but the edge-based matcher, in which global statistics were used, failed where the intensity values varied significantly (for example, the light band). Another point to note is that, since the left and right sides of edges are matched independently, the match record will indicate, among other things, the likely occlusions; these will be at edges where only one side has been put into correspondence.

3. INTEGRATION

A large part of the computation in the edge matcher is devoted to iteration for the purpose of establishing and then refining local disparity estimates (to be used subsequently as matching constraints). If these could be provided to the system, they could enable rapid convergence, more reliable matches, and much higher throughput rates. To explore this possibility, we have done some preliminary experimentation in integrating this stereo system with the baseline SRI STEREOSYS system of Hannah [6]. We passed the results given by STEREOSYS to the edge matcher, and used them as initial seeds – geometric constraint on the matching. The following shows the results of this integration.

First, Figure 16 depicts the edges to be used by the edge-based system (these were obtained at a higher resolution than those of Figure 11). Figure 17 shows the match results from STEREOSYS, with the crosses indicating matched points.

To use these as seeds, we find the edge elements nearest to the matched points and, verifying that either one or both sides are appropriate matches according to the edge-match criteria, we propagate disparity values along the zero-crossing contours. This propagation, an integral part of the edge-based matching described above, has controls that assess the acceptability of the generated matches and determines termination conditions. Figure 18 shows the edge matches obtained by propagation.

This simple propagation increases the number of match points by about an order of magnitude over those furnished as seeds; although not rigorously evaluated, the matches look good. Having established the local disparity constraints here, we can now let the edge-based matcher operate in its normal iterative fashion, with these matches providing the initial conditions. Figures 19 and 20 show the processing over two iterations of the edge-based matcher.

Figure 21 illustrates, for comparison, the final results obtained with the edge-based matcher alone. Clearly, the seeding process leads to better and more dense mapping results.

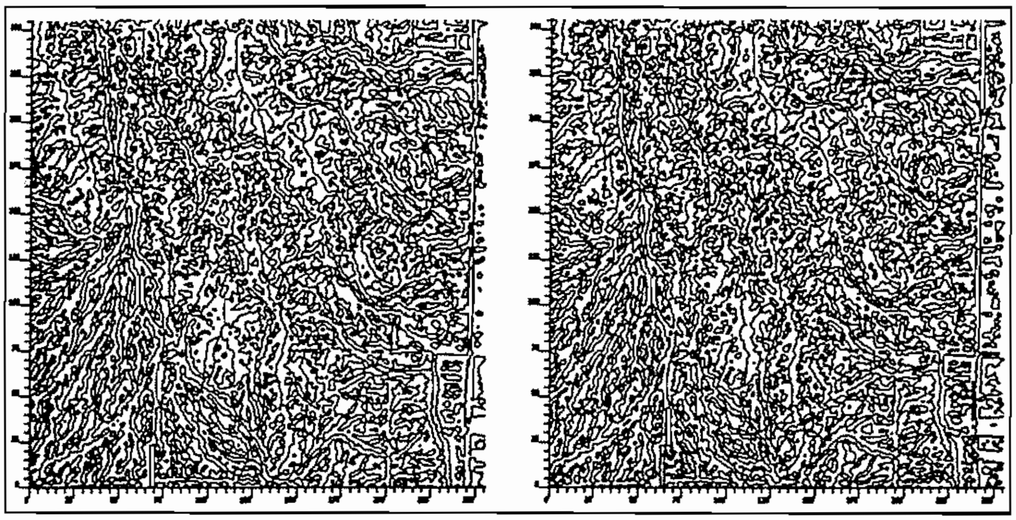

Figure 16. Edges to be Used in Edge-Based Matching

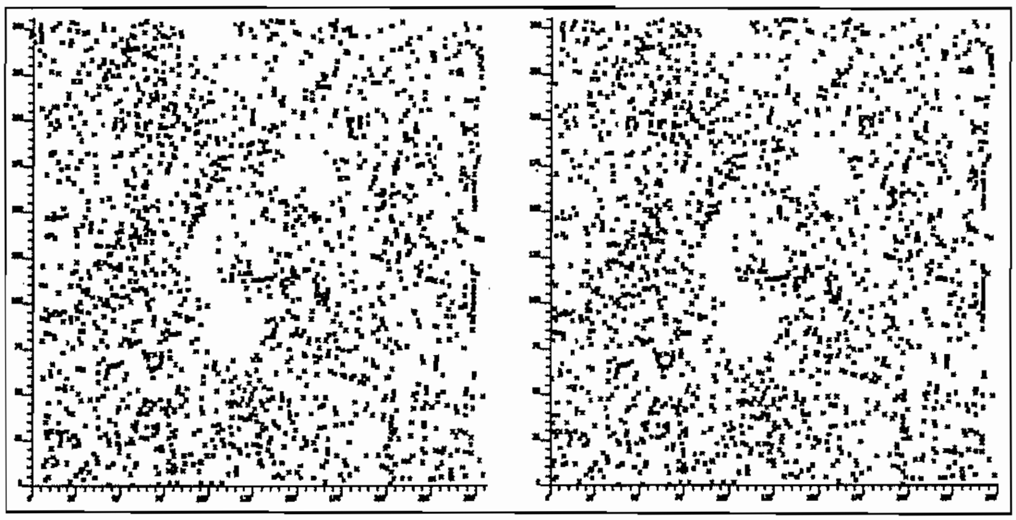

Figure 17. STEREOSYS Match Results

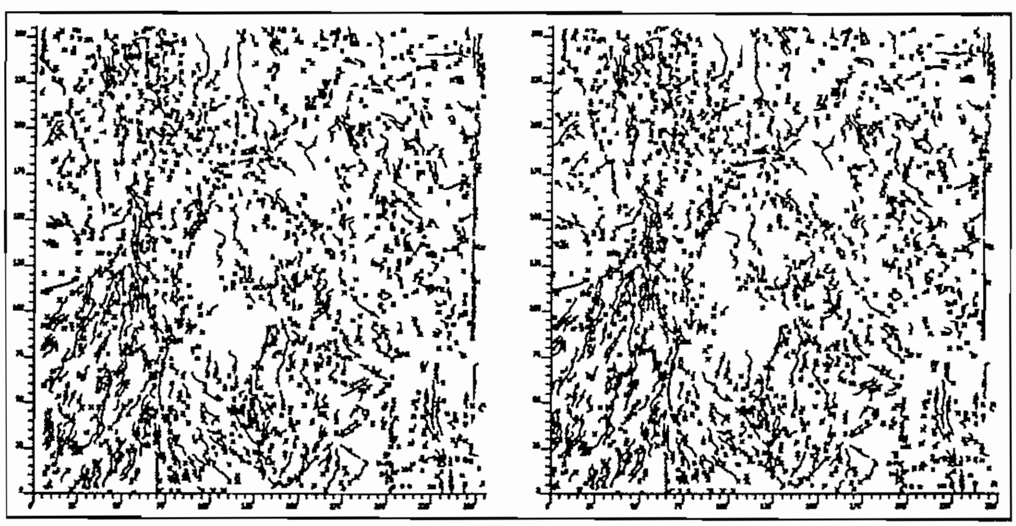

Figure 18. Edge Matches from Propagation

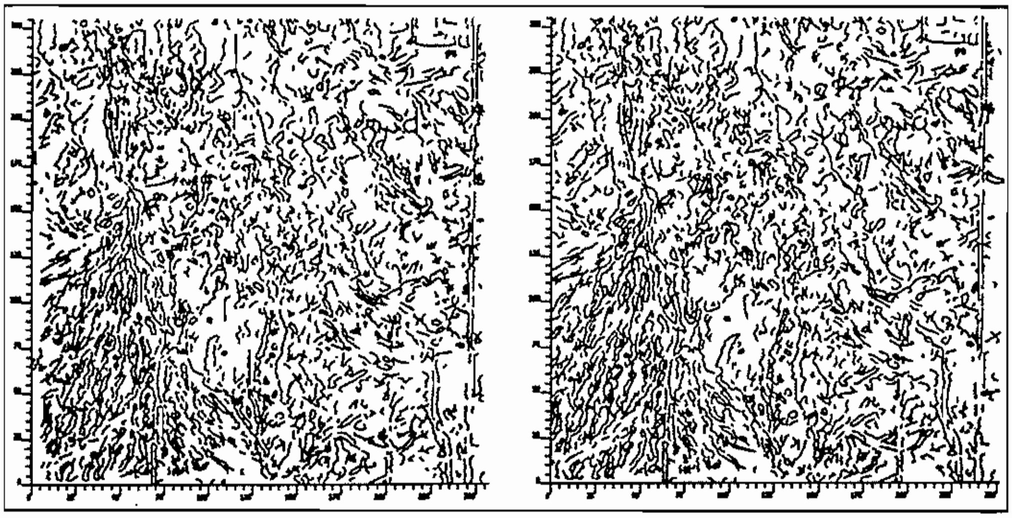

Figure 19. Matches After One Edge-based Pass

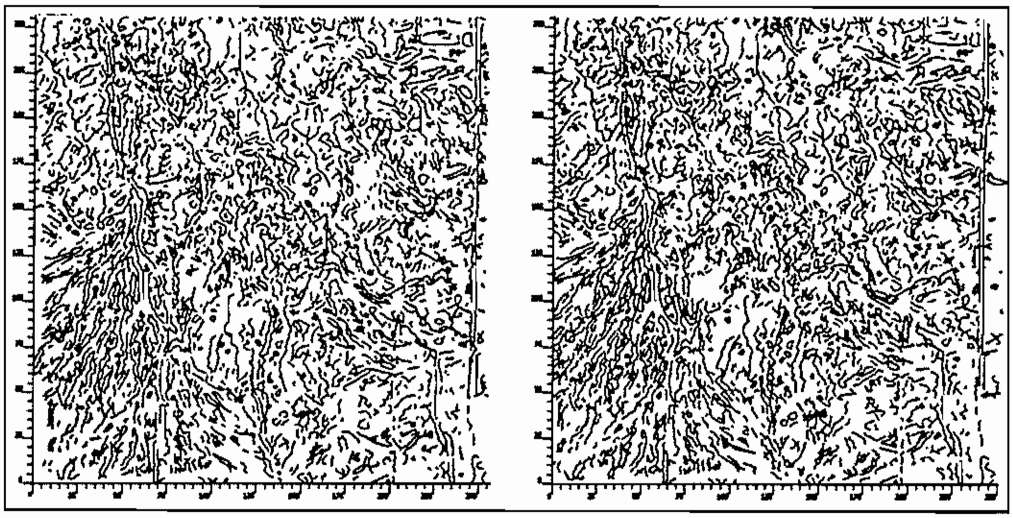

Figure 20. Final Edge-Based Matches

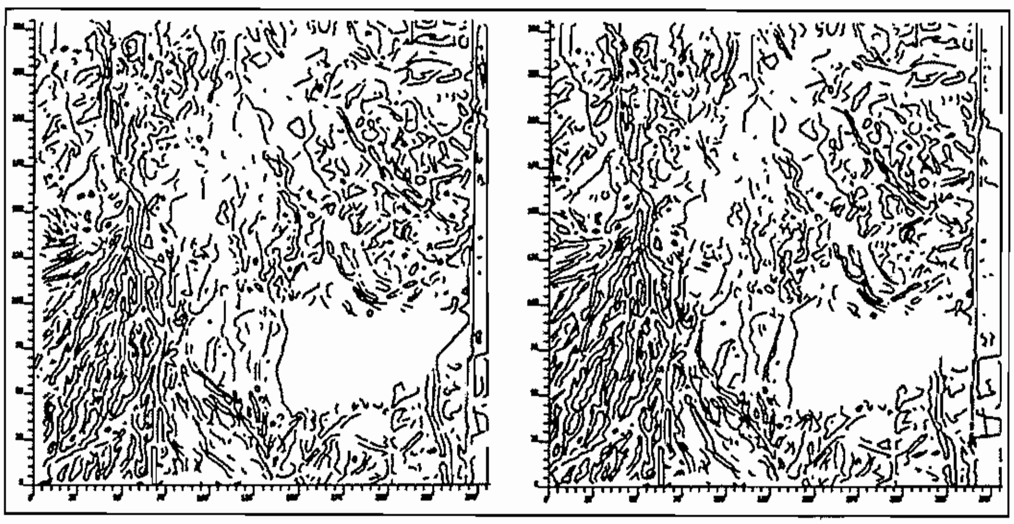

Figure 21. Nonseeded Match Results

4. ASSESSMENT

The imagery referred to as the I5 data set is a roughly one-inch-square portion of a digitized (512×512) three-inch-square subsection of a standard nine-inch photographic negative. Depicting the I5/Spokane Street interchange in Seattle viewed from 24000 feet, it was provided by Boeing Computer Services. The two images are part of a much larger flyover sequence of the Seattle area taken in the mid-1950s. Camera information was not available with the data, nor was ground truth known, preventing a quantitative study of matching results. Manual point selection provided a reasonably accurate camera model; consequently, the results of the iterative edge-based processing were quite good. The few exceptions occurred when moving vehicles disrupted edge continuity (although these tended to be seldom and generally insignificant), and when the repetitive pattern of the parallel freeway lanes (complicated by moving traffic) proved ambiguous.

The main criticism of the results is not concerned with the matching process, as those edge elements that were matched seemed to be the best choice possible, but rather with the limitation inherent in selecting just a single edge frequency for the processing. Zero-crossing contours merge and split as a function of the underlying intensity surface, so that effects caused by occlusion, projection, sampling, and noise can make the edges differ in significant ways between images. The stability of zero-crossing positioning is weakest at those places where a slight change in space constant (σ for the Gaussian convolution) brings about a large change in the topology of the contour. Instability in the coordinates of an edge leads to inaccurate measurement of its disparity at a given scale of analysis. At present, the matcher has no way of knowing about or dealing correctly with this property of edges. Evaluation of the stability and hence the accuracy of a feature's match will likely require treatment of this stability-over-scale property of edges.

In our second demonstration, we processed a 256×256 version of the ETL data set. Working alone, the edge-based matcher performed about as well as could be expected, given that the zero-crossing space constant and resulting contours were relatively arbitrary and the imagery itself had rather poor photometric quality. This edge-based process, unlike the area-based correlator, uses a single gain/bias adjustment for the image set. Furthermore, it does not compensate for nonuniformities in the local intensity surface from one image to the other. Statistics for the gain/bias are collected from the entirety of both images, and applied uniformly over the images during matching. Because of this, a few fairly large regions failed to have matches. Errors greater than one pixel tended to be at the ends of zero-crossing contours, where, as mentioned, point coordinates tend to be least stable. A few errors could be attributed to the ambiguity of repetitive patterns. These were at the edges of the images, where the geometric and photometric constraints used by the analysis are at their weakest. Overall, more than 96% of the matched points were within a pixel of the ETL and STEREOSYS results.

As a preliminary study in seeing how we could integrate the strengths of the two matching approaches, we carried out a further experiment with the edge-based matcher. Here we used the results from STEREOSYS as seeds for the edge-based matcher, applied the connectivity constraints of zero-crossing contours to control a match propagation, and then entered the normal matching iteration. Since establishing disparity constraints is a large part of the edge-based matcher's processing, introducing them directly resulted in significant improvement in the run time. The number of matched points increased by about an order of magnitude over the STEREOSYS results, and the edge-based matches themselves were better and considerably more numerous than in the nonseeded case. Furthermore, the area-based seeds enabled edge-based matching to succeed in the areas of highly textured small patterning to the lower right, where global photometric signal-to-noise estimates proved inappropriate because of film flaws. The edge-based matching enabled substantial improvement over the area-based results in delineating the more obvious structural components of the scene, such as ridge lines along the peaks and drainage flows and arroyos.

References

[1] "Depth from Edge and Intensity Based Stereo," with Thomas O. Binford, *Proceedings of the Seventh International Joint Conference on Artificial Intelligence,* University of British Columbia, Vancouver, Canada, August 1981, 631–636.

[2] H. Harlyn Baker, "Depth from Edge and Intensity Based Stereo," Stanford Artificial Intelligence Laboratory, AIM–347, September 1982.

[3] H. Harlyn Baker, "A System for Automated Stereo Mapping," in *Proceedings of the Commission II Symposium on Advances in Instrumentation for Processing and Analysis of Photogrammetric and Remotely Sensed Data,"*, International Society for Photogrammetry and Remote Sensing, Ottawa, Canada, August 1982, 156–171.

[4] H. Harlyn Baker, "Surfaces from Mono and Stereo Images," invited paper, *Pattern Recognition in Photogrammetry,* Graz, Austria, September 1983, published in **Photogrammetria**, 39 (1984), 217–237.

[5] H. Harlyn Baker, T. O. Binford, J. Malik, J-F. Meller, "Progress in Stereo Mapping," *DARPA Image Understanding Workshop,* Arlington, Virginia, June 1983, 327–335.

[6] Marsha Jo Hannah, "SRI's Baseline Stereo System," *DARPA Image Understanding Workshop,* Miami Beach, Florida, December 1985, 149–155.