SRI International

EVALUATION OF STEREOSYS VS. OTHER STEREO SYSTEMS

Technical Note No. 365

10 October 1985

By: Marsha Jo Hannah, Senior Computer Scientist

Artificial Intelligence Center
Computer Science and Technology Division

SRI Project 5355

The work reported herein was supported by the Defense Advanced Research Projects Agency (DARPA) under Contract No. MDA903-83-C-0027.

SRI International®

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

Evaluation of STEREOSYS vs. Other Stereo Systems

Marsha Jo Hannah

Artificial Intelligence Center, SRI International
333 Ravenswood Ave, Menlo Park, CA 94025

1 Introduction

As previously reported [Fischler, 1984], SRI International is implementing a complete, state-of-the-art stereo system that will produce dense three-dimensional (3-D) data from stereo pairs of intensity images. This system forms a framework for much of our stereo research and will be a base component of our planned expert system for 3-D compilation.

Ideally, we would assess the capabilities of our system by running it on a data set that has known ground truth against which to compare our results. Unfortunately, such data sets do not currently exist because of the extremely high cost of the ground work necessary to measure terrain elevations accurately for a close spacing and to assess the heights of all vegetation and buildings in the area. Lacking such a data set, we can only compare our results against those produced by other stereo systems or against the perceptions of a human looking at the same imagery in stereo on a CRT.

To test our system, currently called STEREOSYS, we have run it on several data sets, including two for which we also have results produced by the DIMP system at the U.S. Army Engineer Topographic Laboratories (ETL). Comparing our matching results to DIMP results or to human perception of what the correct match should be, we have begun to assess the strengths and weaknesses of STEREOSYS's matching techniques, as well as accumulating a catalog of examples of difficult areas for matching [Hannah, 1985]. A description of the experiments that we have conducted and our preliminary conclusions regarding the accuracy of the system are set forth in this report.

2 Description of Systems and Experiments

Our experiments compare the results of our automatic stereo system, STEREOSYS, against the results of the interactive DIMP system at ETL and against the stereo perceptions of an amateur photogrammetrist (the author of STEREOSYS [Hannah, 1984]). These systems and the design of our experiments are briefly described here.

2.1 Description of STEREOSYS

STEREOSYS (an improved version of STSYS, which was described in more detail in Hannah [1984]) is an automatic system for deriving disparity data, hence three-dimensional information, from a pair of aerial images of a scene, taken from moderately different points

of view. This system operates on a hierarchy of different resolution versions of the pair of images, using normalized area correlation as its measure of whether areas in the two images are matched, that is, whether they represent the same point in space. STEREOSYS confines its attentions to image points having high information—the "interesting" points, which tend to be randomly spaced—and operates in several stages, using results from previous stages to constrain the search for points to be filled in at later stages. Overall, the system is very conservative about what constitutes a valid match; it will reject a questionable point at early stages of the processing (possibly filling it in later) in an attempt to produce the most reliable results possible.

2.2 Description of DIMP

The DIMP system, a descendant of the work of Panton [1978] and described in more detail in Norvelle [1981], is an interactively controlled system for deriving disparities. It operates on a single pair of high-resolution images, using normalized area correlation with the areas warped to take the slope of the terrain and the viewing geometry into account. DIMP finds a match for each point on a specified grid within the image, operating in raster scan fashion, with the expected disparity and terrain slope at a point predicted from the matches found at adjoining grid points in the preceding row and column. This system must be initialized manually. Because DIMP must record a match for each grid point (regardless of whether a match exists), and because it uses previous results (regardless of validity) to predict the next match, DIMP has a tendency to get off track, particularly in areas of low or ambiguous information, at places where the elevation or ground slope changes rapidly, or around artifacts in the images. For this reason, DIMP is manually coached—a human monitors its results constantly, interrupting the processing to get DIMP back on track as needed.

2.3 Description of Experiments

Comparing DIMP's grid-based results using warped correlation windows to STEREOSYS's randomly scattered results using ordinary correlation windows is a little like comparing apples and oranges. Matters are further complicated by the fact that, because of the noise properties of the images, STEREOSYS produced its best results in the 1024×1024 versions of these data sets, while DIMP used the 2048×2048 versions. However, we compared them in the following manner.

Comparisons were made only for those points for which STEREOSYS recorded an answer and were done at the resolution of the image in which STEREOSYS had operated. Points were said to have the same answer if the STEREOSYS result and the result at the closest DIMP grid point (scaled into the 1024×1024 image) were within one pixel of having the same disparity. Points about which there was disagreement were examined manually. The operator looked at both results, overlaid on the images at a variety of resolutions, both monocularly and using a stereoscopic viewer. The operator then decided which algorithm appeared to be in error and, based on experience with correlation algorithms, attempted to determine why the mistake had been made.

For data sets with no DIMP results, a much smaller number of points were matched. These were then compared with the human viewer's perception of what were the correct

matches. Only the more blatant mistakes were detected and further analyzed.

3 Evaluations

In this section, we evaluate the performance of STEREOSYS on some of the data sets described in Hannah [1985]. For the first two sets, we have statistics as compared to the DIMP results; for the remaining sets, we give only general impressions of the results as seen by a human viewer.

3.1 The Phoenix Data Set

On the Phoenix data set, STEREOSYS found 5545 "interesting points," of which it thought it could reliably match 4676. Of these, only 43 disagreed significantly with the DIMP results for nearby points. Closer examination showed 15 of these to be uncorrected DIMP errors, 15 were STEREOSYS errors, 5 were points on which both systems appear to have made errors, and 8 were points for which the operator could not determine which system was in error. In most of the cases, the DIMP errors seemed to result from its algorithm having drifted gradually off track (usually starting in an area with little information), and its operator not catching it soon enough; the STEREOSYS approach of first providing a context in which to work, so that the code interpolates disparities, instead of extrapolating them, should remedy this problem. Most of the STEREOSYS errors (and almost all of the points for which the operator could not determine which algorithm was at fault) appeared to have resulted from an inappropriate threshold on the interest value: STEREOSYS was trying to match areas in which there was not enough information to make reliable matches. (The code has since been modified to be more selective about what it uses for "interesting" points.) Some of the STEREOSYS errors were due to not using warped correlation windows to account for the slopes. Most of the information in a window was in a corner of the window, so the disparity that was calculated was that of the corner, not the center of the window; using warped correlation or exponentially weighted correlation windows [Quam, 1984] would solve this problem. A fair number of the mistakes (particularly the ones in which both systems arrived at different wrong answers) were because of artifacts in the data—film grain, scratches, lint, hairs, fiducial marks, and the like; we are a long way from being able to understand, let alone automate, the human ability to identify offending objects and then ignore them in processing stereo data.

3.2 The Canadian Border Data Set

On the Canadian Border data set, STEREOSYS found 1428 "interesting points" (using a more restrictive threshold on interestingness), of which it decided it could reliably match 1262. Of these, 71 disagreed significantly with the DIMP results for nearby points, but only the 27 most blatant disagreements were examined by the operator. Close examination showed 9 of these to be uncorrected DIMP errors, 3 were STEREOSYS errors, 2 were points for which both systems appear to have made errors, and 13 were points for which the operator could not determine which system was in error. The reasons for the errors were highly varied. Most of the cases in which the operator was unable to fix the blame were forested portions of the image: the tree crowns looked sufficiently different in the

two views that a naive human operator was unable to determine the correct match based purely on local context. In the face of this unmatchable data, DIMP had its usual trouble staying on track, particularly since this data set included a lot of discontinuities in depth between trees and ground, which DIMP's surface extrapolation algorithm is not designed to handle. STEREOSYS's errors happened around artifacts in the images, around the depth discontinuity at an overpass, and in an area of trees for which the true match was a subpeak on the correlation function.

3.3 The Moffett-Ames Data Set

Results on the Moffett set were somewhat limited by the lack of detailed camera calibration information to go with this imagery. STEREOSYS has been tuned to depend on having an accurate epipolar line for each point when matching points at later stages in the processing. Unfortunately, the crude relative camera model, which we were able to derive from the first few hierarchically matched points, proved to have significant errors as processing moved away from the center of the image. This meant that, for many points, the search for a match was started out quite far from the true match and frequently did not look far enough: many points failed to match at all, and several locked onto false matches that looked somewhat similar in the clutter of a suburban landscape. Because STEREOSYS was intended for use in a mapping scenario in which accurate camera information is the rule, no attempt has been made to modify it to work more reliably in the absence of accurate camera information.

3.4 The Lexington Reservoir Data Set

The Lexington data set was digitized for another project, which researched algorithms for handling raised objects. Because STEREOSYS is a conventional correlation system, it would not be expected to do well in the presence of depth discontinuities. As predicted, STEREOSYS coped well with the low features in the image and with the shadows of raised objects on the ground, but in areas containing discontinuities, it was unable to find matches that met its criteria for acceptance.

3.5 The Seattle I-5 Data Set

The I-5 data set is a prime example of the type of data on which edge matching triumphs over area matching. The information in the images is almost entirely straight lines resulting from the edges and lanes of the freeway. Most of the places that the "interest" statistic found to be suitable for correlation tended to be either false intersections (where one roadway crossed over another) or cars on the freeway, neither of which had matches in the second image. We were not able to get enough good matches to form even a crude camera model, so were unable to proceed with the processing.

3.6 The International Building Data Set

Despite the fact that STEREOSYS was designed for use on aerial photography, we have tried it on several ground-level pairs of images, just to get a feel for its limitations. We were pleasantly surprised to find that it did relatively well on the International Building

set. Most of the interesting points were on the foreground plants; STEREOSYS coped quite well with these, probably because the background behind them was relatively uniform, so the discontinuities in depth did not cause the appearance of the correlation areas to change much. The only difficulties appeared to be with some false intersections, where two lines that seemed to meet in the image were actually separated in space.

3.7 The Machine Data Set

STEREOSYS also did fairly well with the Machine set. For the most part, it picked out the corners of various things on the machine and seems to have located plausible matches for most of the points it decided that it had matched correctly. There are a couple of questionable matches because of false intersections, which are unavoidable with an area-based matcher.

3.8 The Back Lot Data Set

STEREOSYS also did surprisingly well on the Back Lot data set. Of the points that it decided were well matched, only two were blatantly wrong—a car that is obscured in the second view looks quite similar to the car next to it, so the two interesting points on it are incorrectly matched. A few interesting points that the operator thought should have been easy to match were missed, probably because of interference in the hierarchical matching approach between the disparity of the near-field buildings and that of the background.

4 Conclusions

Our objective in constructing STEREOSYS was to implement a state-of-the-art, area-based system for stereo compilation operating on aerial photography. Along the way, we hoped to remedy some of the obvious problems we had seen with existing systems, such as DIMP's tendency to extrapolate itself off track. In this we have succeeded.

Because STEREOSYS uses fairly independent judgment on each match, it tends to avoid the problems we have seen in the DIMP results; indeed, on the Phoenix data set (and to a lesser degree on the Canadian Border data set), STEREOSYS was able to duplicate DIMP's correct results (for the points tried) and rectify a number of DIMP's mistakes. Although it happens rarely, it is still possible for STEREOSYS to make mistakes in the early stages of its processing, then propagate these mistakes into later matches. To avoid this, more work needs to be done on algorithms for detecting improperly matched points, so they can be removed before further processing.

The major criticism we have heard of STEREOSYS is that it produces matches at randomly spaced points (only where adequate information is present), when what is usually wanted is a closely spaced regular grid of elevation points, regardless of image content. So far, attempts at blindly interpolating the disparity data (ignoring the image data) as reported in Smith [1984] have proven less than satisfying. Marriage of the STEREOSYS techniques with something like DIMP, or with hierarchical warp correlation [Quam, 1984], or with image intensity based interpolation [Smith, 1985] or [Baker, 1982] might be profitable.

We have performed one experiment as a preliminary study in how to integrate the strengths of STEREOSYS with those of an edge-based matcher. The results of STEREO-

SYS were used as seeds for an edge-based matching system [Baker, 1982], which used the connectivity constraints of zero-crossing contours to control match propagation and which then did one iteration of its normal matching process. Because determining disparity constraints is a large part of the edge-based matcher's processing, introducing this information from STEREOSYS's results produced a significant improvement in the runtime of the edge-based matcher. The number of matched points increased by about an order of magnitude over the results of STEREOSYS alone. Although we have not yet finished a quantitative evaluation of these match accuracies, a qualitative analysis indicates that the results from the combined technique are significantly more accurate than the results of the edge-based system alone.

Overall, we have found that STEREOSYS performs credibly on the low-resolution aerial imagery for which it was designed. It has difficulties when processing areas that violate its premises about the continuity of the world, but linking it with an edge-based matcher (which would excel in these types of areas) seems to be a promising approach.

Acknowledgements

The research reported herein was supported by the Defense Advanced Research Projects Agency under Contract MDA903-83-C-0027, which is monitored by the U.S. Army Engineer Topographic Laboratory. The views and conclusions contained in this paper are those of the author and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or of the United States Government.

I would like to thank Robert Bolles, Lynn Quam, Harlyn Baker, and Martin Fischler for their support on this project.

References

Baker, H. Harlyn, 1982. "Depth from Edge and Intensity Based Stereo," Ph.D. Thesis, Stanford University Computer Science Department Report STAN-CS-82-930, September 1982.

Fischler, Martin A., 1984. *Computer Vision Research and Its Applications to Automated Cartography: (Combined) Second and Third Semiannual Technical Reports,* SRI International, Menlo Park, CA, September, 1984.

Hannah, Marsha Jo, 1984. "Description of SRI's Baseline Stereo System", SRI International Artificial Intelligence Center Technical Note 342, October, 1984; also in Fischler [1984].

Hannah, Marsha Jo, 1985. "The Stereo Challenge Data Base", SRI International Artificial Intelligence Center Technical Note 366, October, 1985.

Norvelle, F. Raye, 1981. "Interactive Digital Correlation Techniques for Automatic Compilation of Elevation Data," U.S. Army Engineer Topographic Laboratories Report ETL-0272, October, 1981.

Panton, Dale J., 1978. "A Flexible Approach to Digital Stereo Mapping," *Photogrammetric Engineering and Remote Sensing,* Vol. 44, No. 12, pp. 1499-1512.

Quam, Lynn H., 1984. "Hierarchical Warp Stereo," *Proceedings: Image Understanding Workshop,* New Orleans, LA, October, 1984, pp. 149-156; also in Fischler [1984].

Smith, Grahame B., 1984. "A Fast Surface Interpolation Technique," *Proceedings: Image Understanding Workshop,* New Orleans, LA, October, 1984, pp. 211-215; also in Fischler [1984].

Smith, Grahame B., 1985. "Stereo Reconstruction of Scene Depth," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* San Francisco, CA, June 9-13, 1985.