THE STEREO CHALLENGE DATA BASE

Technical Note No. 366

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

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **10 OCT 1985**	2. REPORT TYPE	3. DATES COVERED **00-10-1985 to 00-10-1985**
4. TITLE AND SUBTITLE **The Stereo Challenge Data Base**		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**		**21**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

The Stereo Challenge Data Base

Marsha Jo Hannah

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

1 Introduction

As previously reported in Fischler [1984] and Hannah [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. 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 stereo system at the U.S. Army Engineer Topographic Laboratories. While comparing our matching results to DIMP results or to human perception of what the correct match should be, we have begun to accumulate a catalog of examples of difficult areas for stereo processing.

In this report, we describe several data sets that we have processed and discuss the types of problems that our matching algorithms have encountered. This information is part of the "stereo challenge data base" we are assembling to test matching algorithms against; the actual data base will contain many more instances of hard-to-match places than are shown in the simple examples illustrated here.

2 Data Sets Processed by STEREOSYS

The following data sets have been processed through STEREOSYS, our stereo compilation program. The areas noted are examples of types of areas that STEREOSYS had incorrectly matched (as compared with other computer algorithms or with human stereo results), ones that STEREOSYS was unable to match well enough to suit its internal criteria, or ones on which STEREOSYS was unable to do anything for lack of information in the imagery.

2.1 The Phoenix Data Set

Most of our area-based processing and analysis to date, as well as some edge-based processing, has been done on a data set that we received from the U.S. Army Engineer Topographic Laboratories (ETL). The imagery consists of a pair of 2048 × 2048 pixel images representing a 2" × 2" portion from two standard 9" × 9" mapping photographs taken over Phoenix South Mountain Park, near Phoenix, Arizona. The data covers approximately a 2-km square of high desert, both plain and steep hills, dotted with brush; the beginnings of an agricultural area is at one edge of the images.

This data set is known locally as the Phoenix set. In addition to the images, this data set also contains camera information in the form of absolute position and orientation data, internal calibrations for the camera, and rectification polynomials to account for the digitization process. We also have a set of results from the interactively coached DIMP stereo compilation system at ETL [Norvelle, 1981] in the form of an array of the matching points for a grid of image points (every 5th pixel) and the arrays of 3-D positions derived from these matched point pairs.

This data set provides a number of challenges to stereo processing algorithms, particularly to those based on area correlation. (Numbers in parentheses refer to the example points in Figure 1 and Table 1.) At least half of the terrain in the imagery is very steep (1), so that an area on the ground frequently projects to windows of different sizes and shapes in the two images; this frequently results in poor correlations or in mismatches. There are some portions of the terrain that have little vegetation, giving correlation algorithms insufficient or unreliable information with which to work (12). The agricultural area contains some very straight roads surrounded by land without distinguishing visual texture (2), causing matches to "slide" along the roads until the noise in the images matches best. Some of the roads contain cars that have moved in the time between the two images (3), rendering those areas difficult to match. The images also include portions of regularly spaced orchards (4, 5, 6), which can lead to local confusion by the matcher, because all the trees look alike and have very similar context. In the agricultural area, a few buildings (7) cause depth discontinuities that can be difficult for the matcher.

The Phoenix data set is made more challenging because the imagery is of somewhat poor quality, with scratches (8), pen marks (9), fiducial marks (10), hairs (11), and the like, which have been digitized into the data. The photographs also appear to have been digitized at the maximum possible resolution—the film grain (12) is apparent in otherwise low-information areas of the imagery, leading to random mismatches.

2.2 The Canadian Border Data Set

We have also done a significant amount of processing on a data set received from the Defense Mapping Agency (DMA). The imagery consists of a pair of 2048 × 2048 pixel images representing a portion of two mapping photographs taken somewhere along the U.S.-Canadian border. The data set covers an area of gently rolling terrain cut by a steep ravine and crossed by a major highway; the ground cover is a mixture of forested areas having sharp boundaries with areas that have been cleared for crop lands; the imagery also contains several farm complexes and a town.

This data set is known locally as the Canadian Border set, or, more simply, the Canada

Point	x	y	Description
1	1136	436	Steep ridge
2	1616	420	Ambiguity along road
3	1972	286	Car moved on road
4	1892	526	Regular pattern in orchard
5	1924	586	Horizontal ambiguity along orchard edge
6	1954	482	Vertical ambiguity along orchard edge
7	1950	722	Discontinuity at building
8	1178	140	Digitized scratch on photo
9	1502	636	Pen mark on photo
10	1236	862	Fiducial mark on photo
11	1726	170	Hair on photo
12	1642	912	Digitized film grain

Table 1: Examples from lower right quarter of Phoenix imagery

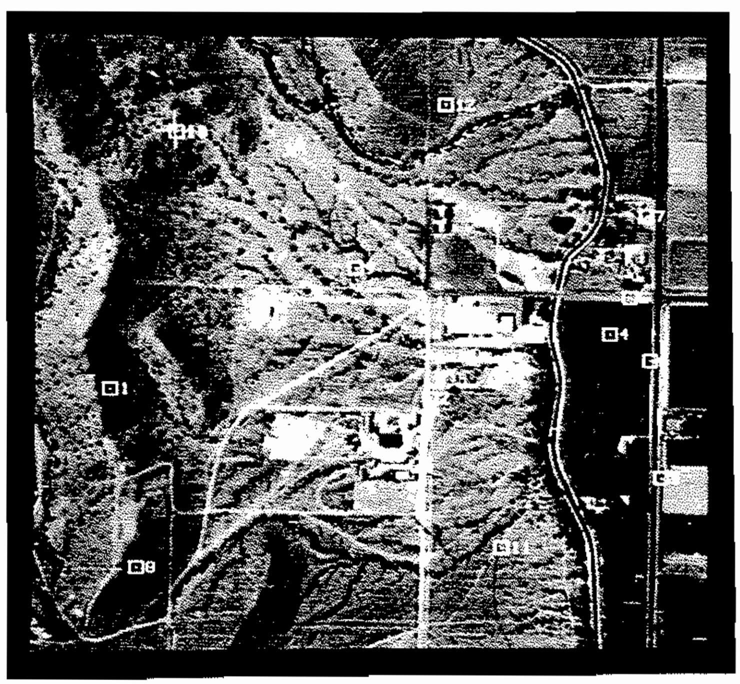

Figure 1: Lower left quarter of Phoenix image at 1024 × 1024 resolution

set. In addition to the images, this data set also contains camera information, in the form of absolute position and orientation data, internal calibrations for the camera, and rectification polynomials to account for the digitization process. We also have a set of results from the interactively coached DIMP stereo compilation system at ETL in the form of an array of the matching points for a grid of image points (every 10th pixel).

This data set is extremely challenging for stereo processing algorithms, whether based on area correlation or edge matching. (Numbers in parentheses refer to the example points in Figure 2 and Table 2.) The major problem encountered in these images is the tree cover. In some areas, the trees are very dense and in full foliage so that the ground cannot be seen at all (1, 2, 3). In other areas, the trees are more sparse so a particular window might contain both tree tops and ground, which match at different disparities (4); this also happens at the edge of a dense forest (5) and where a narrow row of trees lines a field (6). In many cases, the tree tops contain enough detail that they present a much different appearance in the two images making any sort of matching is a problem, let alone separating tree elevation from ground elevation. The steep terrain in the vicinity of the ravine compounds the problem, causing the vegetation to be foreshortened differently in the two views (7). There is a large building complex in the ravine, further complicating the matching problem by introducing partial occlusions along its walls (8). There is also a highway bridge over the ravine (9) and a highway overpass (10), both of which cause similar problems because of occlusions. Straight highways (11), with an occasional car that moved between the times of the two views, cause the usual problems, as do agricultural fields (12) with little internal visual information. As with the Phoenix set, film grain and various artifacts such as hairs, scratches (13), and pen marks (12) all have negative effects on matching algorithms.

2.3 The Moffett-Ames Data Set

We have also processed an urban data set received from the Defense Mapping Agency. The imagery consists of a pair of 1024 × 1024 pixel images representing a portion of two mapping photographs taken over the Moffett Field Naval Air Station and the NASA Ames Research Center including portions of the cities of Mountain View and Sunnyvale, California. The data covers an area of generally level terrain adjoining San Francisco Bay; in addition to the airfield and hangers, the area includes salt evaporator ponds, agricultural fields, housing developments, and office complexes and is crossed by a major highway.

This data set is known locally as the Moffett-Ames set or, or more simply, the Moffett set. This data set came with camera information (absolute position and orientation data, internal calibrations for the camera, and rectification polynomials to account for the digitization process), but we have been advised that this information contains errors, so have not attempted to use it. At present, we have no other matching results for this data set, although it is rumored that some form of ground truth exists.

This data set has a number of challenging features for stereo processing algorithms, whether based on area correlation or edge matching. (Numbers in parentheses refer to the example points in Figure 3 and Table 3.) Most of the features in the images are man-made structures of one form or another; this leads to strong linear edges along roads (1) and airfield runways (2), which are troublesome for area correlation. There are a number of large buildings in the area, including Moffett's blimp hanger (3), NASA's wind tunnel (4), and

Point	x	y	Description
1	698	752	Dense trees with dark foliage
2	334	1822	Dense trees with medium-intensity foliage
3	850	662	Dense trees with light foliage
4	396	444	Mixed trees and ground
5	808	862	Edge of dense trees
6	196	1606	Row of trees between fields
7	888	1632	Trees in ravine
8	2000	1182	Large buildings in ravine
9	968	1580	Highway bridge over ravine
10	1058	1208	Highway overpass
11	1592	794	Ambiguity along highway
12	1162	86	Pen marks in field
13	420	1992	Scratches on photo

Table 2: Examples from Canada imagery

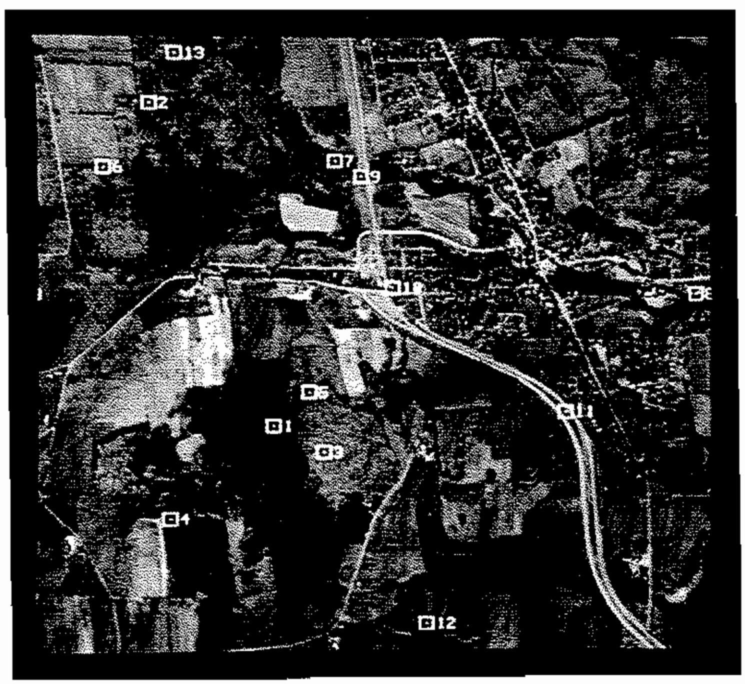

Figure 2: Canada image at 512 × 512 resolution

Point	x	y	Description
1	736	675	Edge of US-101
2	589	665	Edge of Moffett runway
3	338	662	One of Moffett's blimp hangers
4	463	322	NASA's wind tunnel
5	681	945	Lockheed's Building 001
6	438	206	Trailer park
7	628	438	Rows of barracks at the naval station
8	676	186	Similar blocks of regularly spaced houses
9	881	855	Rows of identical light industrial buildings
10	557	935	Parking lots with regular patterns of cars
11	677	735	Agricultural fields
12	76	760	Salt ponds
13	186	838	Specular reflection on salt pond
14	238	909	Specular reflection on salt pond

Table 3: Examples from Moffett imagery

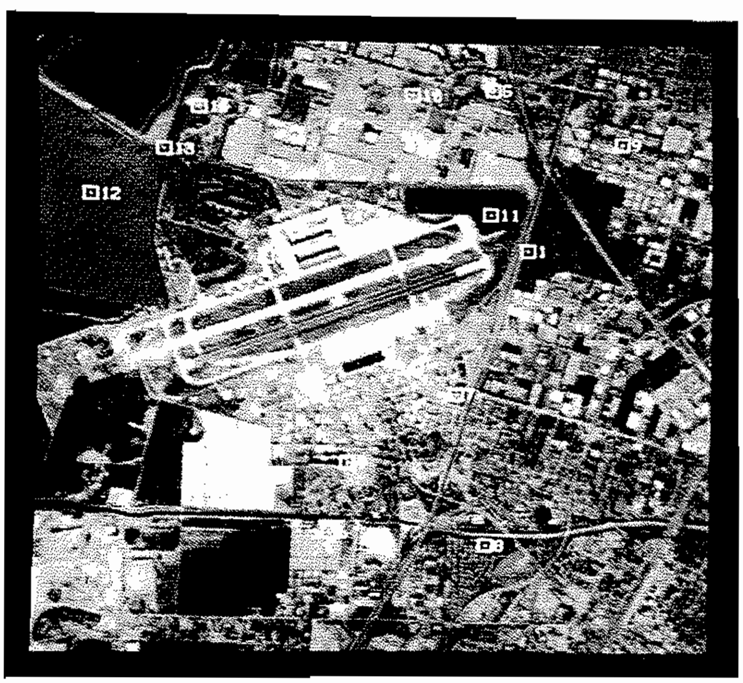

Figure 3: Moffett image at 512 × 512 resolution

Lockheed's Building 001 (5), which present the usual problems with partial occlusions. The imagery includes a variety of suburban housing, whose fine detail will be difficult for edge matching algorithms to handle. In addition, there are several repetitive patterns in these images, such as rows of trailers in a trailer park (6), rows of barracks at the naval station (7), blocks of regularly spaced houses (8), rows of identical light industrial buildings (9), and parking lots with regular patterns of cars (10). There are the usual problems with large blank areas such as the agricultural fields (11) and the salt ponds (12). The salt ponds are particularly troublesome, because the motion of the camera along the flight path causes some of these ponds (13, 14) to show specular reflections in one image, but not in the other; this causes contrast reversals with the surrounding dams, which will confound most area and edge matchers.

On the positive side, this data set appears to be relatively clean; that is, it is free from the scratches, lint, hairs, pen marks, and other artifacts that frequently compound the problem with aerial imagery. However, the lack of precise camera information severely handicapped our processing of this imagery, because the images appear to have a significant distortion near their edges. The crude relative camera model calculated from the first few matched points was significantly in error (i.e., human-indicated matching points were several pixels away from the predicted epipolar lines) over much of the image; this resulted in many points which failed to match at all, as well as a number of falsely accepted mismatches, because of the ambiguities inherent in urban scenes.

2.4 The Lexington Reservoir Data Set

We have partially processed a data set that we digitized ourselves from aerial images received from the Defense Mapping Agency. The imagery consists of a pair of 512×512 pixel images representing a small portion of two mapping photographs taken along Highway 17 in the vicinity of Lexington Reservoir near Los Gatos, California. The data is a high-resolution view of a relatively small area, including a part of the freeway, a small water storage tank, part of a large tank, a small building, a few trees, and a hill.

This data set is known locally as the Lexington Reservoir set or, more simply, the Lexington set. We do not have camera information for this data set, nor do we have other matching results for it.

This data set provides a severe challenge for ordinary matching algorithms. (Numbers in parentheses refer to the example points in Figure 4 and Table 4.) Large areas of the data have no visual information, such as the concrete aprons around the tanks (1), asphalt service roads (2), or grassy hillsides (3). The tops of the trees (4, 5) are seen from much different perspectives and so have radically different appearances. The linear edges between the bland areas cause the usual problems, as does the highway itself (6); the car (7) that has moved between the two views also causes matching problems. Because the images are such high resolution, the discontinuities in the image around the small tank (8) and the building (9) are a significant problem. For the ultimate challenge, there is also an isolated power pole (10) to attempt to match.

On the positive side, this data set appears to be relatively free from the scratches, lint, and other artifacts that frequently compound the problem with aerial imagery. However, the high resolution was obtained by digitizing down to the film grain, so many of the "features" found by the interest operator are really noise in otherwise blank areas.

Point	x	y	Description
1	496	366	Concrete apron around a tank
2	410	404	Asphalt service road
3	228	162	Grassy hillside
4	126	258	Tree top
5	42	348	Tree top
6	178	28	Highway 17
7	80	38	Car that has moved between the two views
8	230	272	Small tank, up on stilts
9	396	320	Building
10	384	146	Power pole

Table 4: Examples from Lexington imagery

Figure 4: Lexington image at 512 × 512 resolution

2.5 The Seattle I-5 Data Set

We have partially processed a data set acquired from Boeing. The imagery consists of a pair of 200×200 pixel images from mapping photographs taken over the interchange of Interstate 5 with Spokane Street in Seattle, Washington. The data is a medium-resolution view of a relatively small area, featuring part of this major freeway interchange.

This data set is known locally as the Seattle I-5 set or, more simply, the I-5 set. We do not have camera information for this data set, nor do we have matching results other than those area- and edge-based matches we have produced on it.

This data set provides many good features for edge matching, but a severe challenge for area-based matching algorithms. (Numbers in parentheses refer to the example points in Figure 5 and Table 5.) The vast majority of the information in the images lies along the various roadways, both in their external edges (1) and in the internal edges between lanes (2). Our "interest" operator will not select areas containing only linear structures, but readily selects places where one linear structure intersects another. Unfortunately, such points occur mainly where one roadway crosses over another (3, 4). Because these are not true intersections (i.e., the freeway and its overcrossing do not actually intersect, but merely appear to do so in most views), such points rarely have a proper match in a different view of the scene. Unfortunately, they do have very well-correlated false matches, which occur where the two linearly-ambiguous structures falsely intersect in the second photo. Also highly "interesting" are points where the linear pattern of the road is obscured by a car (5), which, of course, has a different position in the other image. In addition to the problems of obscuration caused by the discontinuities between the levels of the roadway (6), there are also the usual problems with foreshortening on the steep banks leading from one level of the interchange to another (7) and with the relatively blank areas of landscaping in some of the adjoining areas (8).

As presently implemented, our stereo system was unable to do much with these images. So many of the points were either unmatchable or had false matches that we were unable to obtain even a crude relative camera model for these images; hence, we were unable to proceed. An edge-matching algorithm, started with carefully hand-picked initial matching points, was able to derive the model it needed and process most of the image, although it had difficulties with the ambiguities inherent in the similar, parallel lanes of the freeway.

2.6 The International Building Data Set

We have also processed several ground-level stereo data sets digitized locally from pictures taken with a hand-held 35-mm camera. The first of these sets consists of a pair of 450×450 pixel images taken in the patio of the International Building at SRI in Menlo Park, California. In the foreground are three large pots containing a small tree, a bush, and some succulents; in the background are a few chairs in front of a wall of the building.

This data set is known locally as the International Building set. We do not have camera information for this data set, nor do we have matching results other than those we have produced on it.

This data set provides some very interesting challenges for all types of matching algorithms. (Numbers in parentheses refer to the example points in Figure 6 and Table 6.) The little tree in the foreground (1) is quite diffuse, so almost any window within the tree

Point	x	y	Description
1	46	158	Edge of I-5
2	13	68	Edges between lanes of I-5
3	65	128	Pseudo-intersection of two roadways
4	114	95	Pseudo-intersection of two roadways
5	24	144	Car which moved between images
6	138	82	Discontinuities between levels of the roadway
7	190	141	Foreshortening on steep banks
8	31	56	Featureless areas of landscaping

Table 5: Examples from I-5 imagery

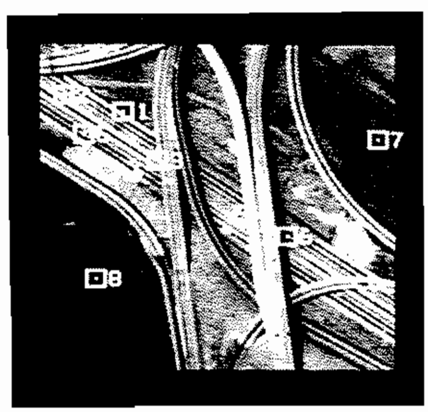

Figure 5: I-5 image at 200 × 200 resolution

Point	x	y	Description
1	288	275	Diffuse foreground tree
2	226	286	Background behind tree
3	80	336	Reflection in window
4	39	196	Near-field occlusions
5	425	203	Pseudo-intersections
6	375	205	Linear column edge
7	104	419	Blank ceiling

Table 6: Examples from International Building imagery

Figure 6: International Building image at 450 × 450 resolution

will also contain pixels from the background (2); the trick is to separate them. The large windows in the middle ground (3) contain very clear reflections of objects out of the field of view of the images; these objects are matchable, but will receive spurious depths, because the depth triangulation calculations assume that lines of sight are straight. Extreme near-field objects will cause the usual problem with occlusions (4) and pseudo-intersections (5). Of course, area-based measures will have their usual difficulties with linear features such as the columns (6) and blank areas such as the ceiling (7).

2.7 The Machine Data Set

Another of the ground-level stereo data sets we have processed was also digitized locally from pictures taken with a hand-held 35-mm camera. This set consists of a pair of 500×500 pixel images taken in one of the parking lots at SRI in Menlo Park, California. In the foreground is a large piece of machinery (probably a diesel-powered generator) sitting on blocks, and behind it is an oblique view of a building with a few small trees planted along it and part of a row of cars parked in front of it.

This data set is known locally as the Machine set. We do not have camera information for this data set, nor do we have matching results other than those we have produced on it.

This data set provides some interesting challenges for matching algorithms. (Numbers in parentheses refer to the example points in Figure 7 and Table 7.) The radiator of the machine (1) is seen at a rather oblique angle, so is foreshortened differently in the two views; the digitization also brought out interesting moire patterns, which differ in the two views. The electric truck behind the machine (2) has been driven away between the times of two views, complicating matches in that area. The exhaust stacks on the machine (3) create pseudo-intersections with the building, which will cause difficulties for most matchers. The car fender (4) is occluded by the machine in the second view. The machine contains a great deal of fine detail, such as wiring (5), whose narrowness presents problems for the matcher. Much of the detail on the building (6) is linear and very nearly parallel with the epipolar line, so is difficult for area- or edge-based matchers to handle properly. The building itself (7) and the asphalt of the parking lot (8) both contain little information, with just enough noise introduced by the digitization to cause trouble.

2.8 The Back Lot Data Set

Another of the low-angle stereo data sets we have processed was also digitized locally from pictures taken with a hand-held 35-mm camera. This set consists of a pair of 254×254 pixel images taken from the roof of one of the buildings at SRI in Menlo Park, California. The scene is framed by two large buildings at each side of the imagery; seen between the buildings are two rows of cars parked along a street with a low building behind them and lots of trees behind that.

This data set is known locally as the Back Lot set, or more simply, the Lot set. We do not have camera information for this data set, nor do we have matching results other than those we have produced on it.

This data set provides some interesting challenges for matching algorithms. (Numbers in parentheses refer to the example points in Figure 8 and Table 8.) The most difficult problem posed by this data set is how to deal with points that are unmatchable, because

Point	x	y	Description
1	139	288	Radiator foreshortened, with moire pattern
2	88	289	Truck moves between frames
3	216	397	Exhaust stack pseudo-intersects building
4	106	336	Fender occluded
5	324	245	Wiring detail on machine
6	457	421	Linear feature, paralleling epipolar lines
7	168	435	Blank wall
8	80	88	Blank pavement

Table 7: Examples from Machine imagery

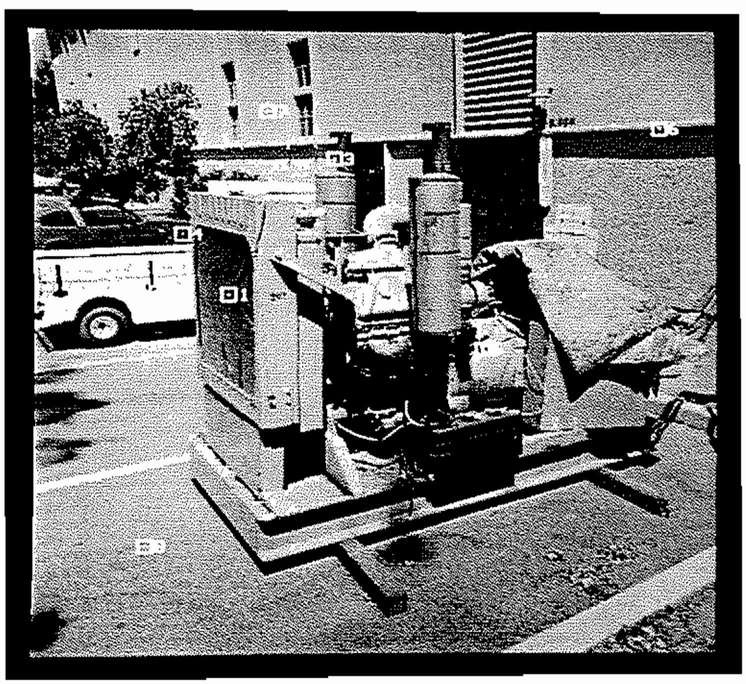

Figure 7: Machine image at 500 × 500 resolution

Point	x	y	Description
1	172	32	Front wheel of car obscured in 2nd image
2	170	91	Car obscured in 2nd image
3	91	62	Cars foreshortened differently
4	124	224	Tree structure ambiguous
5	168	227	Tree nearly obscured
6	183	76	Linear building edge
7	157	118	Linear roof line, paralleling epipolar lines
8	221	64	Blank wall
9	101	40	Blank ground

Table 8: Examples from Back Lot imagery

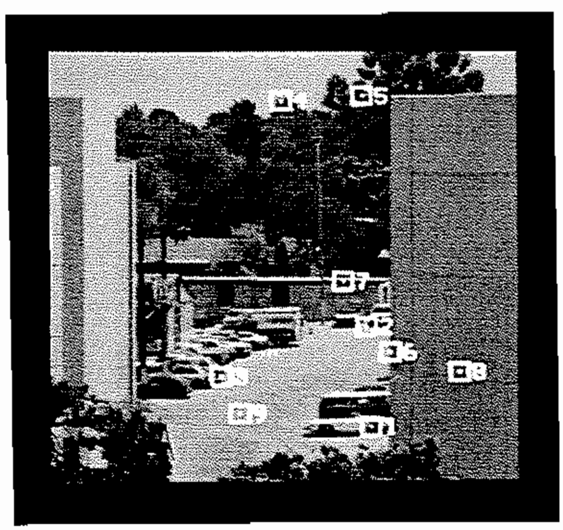

Figure 8: Back Lot image at 254 × 254 resolution

of occlusions. The strip of data just to the left of the edge of the right-hand building does not appear in the second image, because of the change in point of view. This means that the front wheel of the first car in that row (1) and the partially visible car in the back row (2) do not have valid matches in the second image, but a window containing the front wheel of the first car (1) looks quite like a window containing the back wheel of that car, leading to a mismatch with a fairly good correlation; similarly, the car in the back (2) looks enough like the car next to it to cause a persistent mismatch. The cars in the other row (3) are foreshortened or occluded just enough to make matching difficult. The humps and bumps in the skyline tree edge (4) are sufficiently similar to cause mismatches. Hierarchical techniques did not work well on the tree (5) behind the building at the right, seeming to lock onto the building corner instead of the tree in the low resolution versions of the image. There were the usual problems with linear edges (6), especially the ones parallel to the epipolar lines (7), as well as problems with areas that had marginal information, such as the buildings (8) and the parking lot (9).

3 Other Data Sets

We have available several more data sets that we have not processed as yet. From our experience, however, we feel that each of these data sets provides some interesting challenges for stereo processing. We note these in passing.

3.1 The Washington Monument Data Set

We have a pair of 512 × 512 pixel images acquired from Carnegie-Mellon University; these were taken over the Washington Monument in Washington, DC (see Figure 9). This is a fairly wide-angle pair so that many of the buildings have one vertical face shown in one image and the opposing face shown in the other; these occlusions will significantly complicate matching. A fair amount of traffic on the streets has moved in the time between the two images. The strong linear patterns of the streets and the blank roof tops will cause the usual problems for area-matching algorithms; the detail on some of the building sides may confuse edge-based methods.

3.2 The Fort Belvoir Doublet Data Set

We have a pair of 512 × 512 pixel images received from the Defense Mapping Agency; these were taken near Fort Belvoir, Virginia (see Figure 10). The images show part of a freeway with the usual moving traffic as well as a petroleum tank farm. Because this is a fairly wide-angle pair, the amount of visible tank face varies between the images. In a number of areas, the trees have apparently shed their leaves for the winter, as the shadows of the trunks are visible on the ground through a "haze" of upper branches—a difficult situation for area- and edge-based matchers alike. The images are "contaminated" with a large black triangle, which was apparently drawn on the original photograph before it was digitized. Camera information is reputed to be available for these images, but is rumored to contain errors.

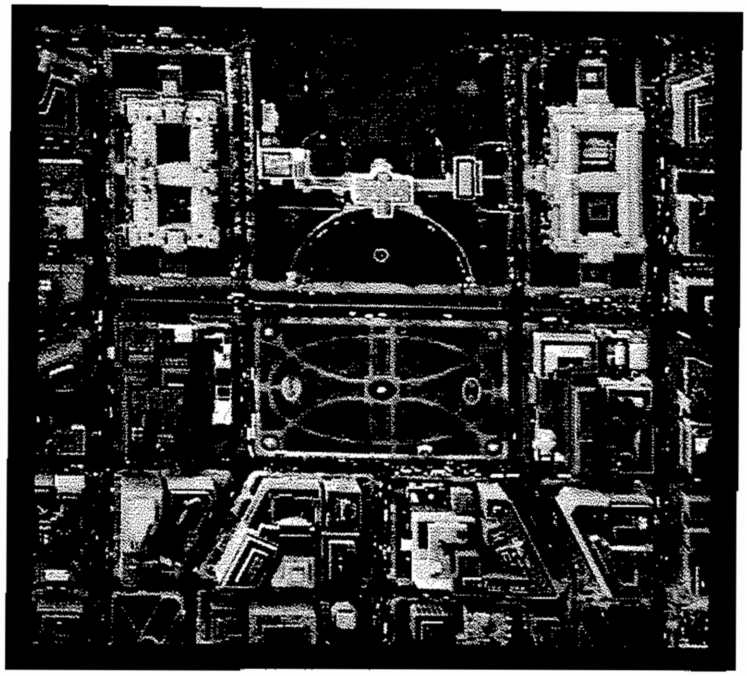

Figure 9: Washington Monument image at 512 × 512 resolution

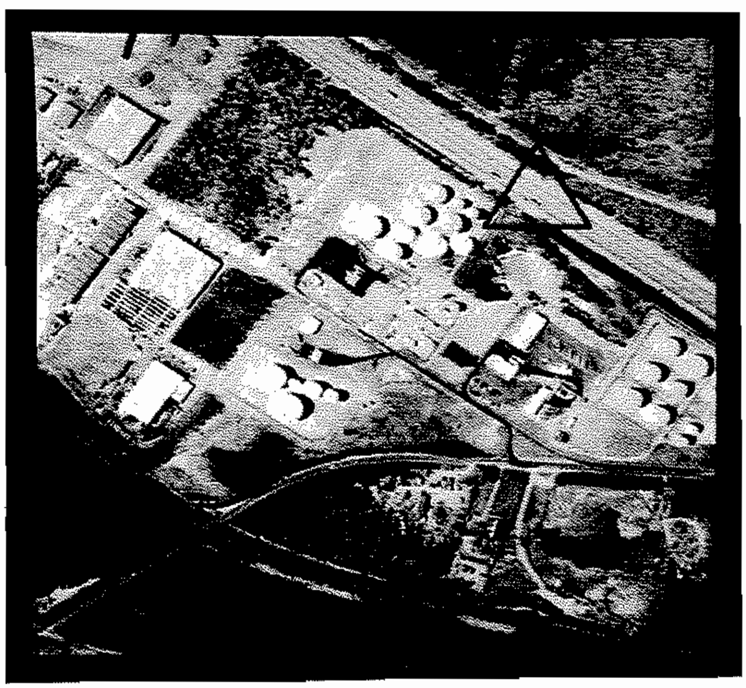

Figure 10: Fort Belvoir Doublet image at 512 × 512 resolution

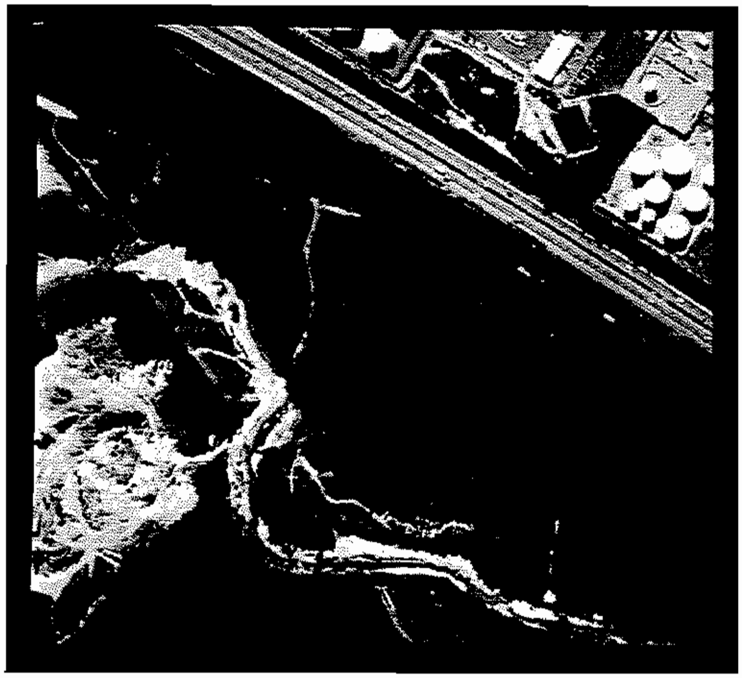

Figure 11: Fort Belvoir Triplet image at 512 × 512 resolution

3.3 The Fort Belvoir Triplet Data Set

We also have a trio of 512 × 512 pixel images received from the Defense Mapping Agency; these were taken near Fort Belvoir, Virginia (see Figure 11). The images show part of a freeway with the usual moving traffic, as well as a large area of forest, a steep ravine, a gravel quarry, and what appears to be an office complex under construction; a portion of the petroleum tank farm featured in the Fort Belvoir Doublet also appears in a corner of some of the images. Most of the area of the images is covered with trees, which are in full leaf; the crowns provide a relatively bland area with detail differing greatly in the two views. An interesting challenge is matching the high-tension power transmission towers, which appear at various places across the images. The images are "contaminated" with some of the edge markings on the original photographs, because the edges were not clipped before digitization. Also, the contrast and brightness of the images is not constant—the third image differs significantly from the other two, which may confound some matching algorithms. Camera information is reputed to be available for these images, but is rumored to contain errors.

3.4 The Phone Data Set

We also have a pair of 256 × 256 pixel images of a telephone sitting on a desk top (see Figure 12), which forms quite a challenge for stereo processing. On the desk, in addition to the phone, there is a decorated porcelain coffee mug containing a pencil. The background behind the scene is slightly out of focus and contains a sparse, but highly ambiguous pattern,

Figure 12: Phone image at 256 × 256 resolution

which most stereo algorithms match incorrectly. The change in point of view results in a significant rotation of the scene, so most of the objects are foreshortened differently between the two views.

3.5 The Chair Data Set

We also have a trio of 256 × 192 pixel images taken of two chairs (see Figure 13). The two chairs, one a secretarial swivel chair, the other a conference room stackable chair, each contain relatively little detail, and their background is a wall that is almost the same intensity as the chairs. Other objects in the scene include a chart of some type hanging askew on the wall, a large soft-drink cup on the secretarial chair, a small oscilloscope on the stackable chair, and a tablelike object in the foreground with two unidentified objects on it. Both of the chairs have reflections of the ceiling light fixtures on their vinyl coverings, and there is an artifact common to the 3 images in the lower left corner: a black corner with a white bar across it. The lack of features and the indistinct edges will make this a challenging data set for most stereo algorithms.

3.6 The Motion Data Sets

We also have available some motion sequences of images taken in the robotics laboratory, which had been cluttered with a variety of house plants and other objects to make the problem more interesting (Figure 14 shows a typical scene). These images were taken

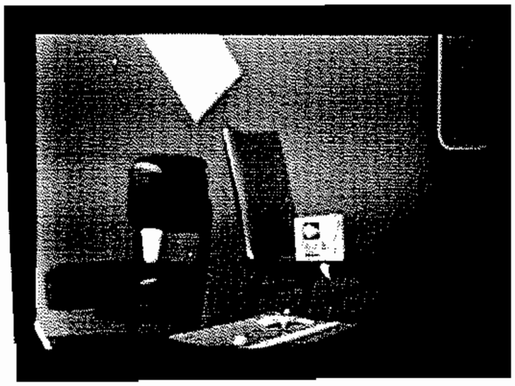

Figure 13: Chair image at 256 $\times$ 192 resolution

Figure 14: A typical Motion image at 490 $\times$ 480 resolution

with a CCD video camera mounted on an x-y table, which was moved in 125 steps of 0.2" each, in a straight line either laterally or forward; because the camera was precisely controlled, it should be possible to recover the camera information. All of the scenes are quite complicated, with near-field objects that change relative positions with respect to objects in the background from frame to frame, some areas of nearly constant intensity, and many pseudo-intersections, where edges that do not meet in the real world appear to intersect in the images. The large number of images (currently available on the LISP-Machines, but a few may be transferred to the VAX for more study) makes it possible to experiment with optic flow techniques, stereo at a variety of baseline lengths, stereo combined with motion, and the like.

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].

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.