TEC-0016

Built-Up Area Feature Extraction: Final Report

Digital Mapping Laboratory
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

October 1991

Prepared for:
U.S. Army Corps of Engineers
Topographic Engineering Center
Fort Belvoir, Virginia 22060-5546

US Army Corps of Engineers
Topographic Engineering Center

T E C

Destroy this report when no longer needed.
Do not return it to the originator.

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

The citation in this report of trade names of commercially available products does not constitute official endorsement or approval of the use of such products.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 1991	Final Technical Report July 1987 - June 1990

4. TITLE AND SUBTITLE

Built-Up Area Feature Extraction: Final Report

5. FUNDING NUMBERS

DACA72-87-C-0001

6. AUTHOR(S)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Digital Mapping Laboratory
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Topographic Engineering Center
Fort Belvoir, VA 22060-5546

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

TEC-0016

11. SUPPLEMENTARY NOTES

Effective 1 October 1991, the U.S. Army Engineer Topographic Laboratories (ETL) became the U.S. Army Topographic Engineering Center (TEC).*

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This final technical report describes the overall research progress during a three year period. In July 1987, the Digital Mapping Laboratory at Carnegie Mellon University began work with the U.S. Army Topographic Engineering Center (TEC)* to explore the detailed analysis of aerial imagery with particular emphasis on built-up areas containing large numbers of complex man-made structures. During the past three years research was done in several important areas including scene registration, stereo analysis, shadow analysis, and building detection. Each of these areas addresses an important set of issues toward the development of automated tools for cartographic feature extraction.

14. SUBJECT TERMS

Computer Vision, Automated Cartography, Digital Mapping

15. NUMBER OF PAGES

67

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

Table of Contents

List of Figures

List of Tables

PREFACE

This report was prepared under contract DACA72-87-C-0001 for the U.S. Army Topographic Engineering Center, Fort Belvoir, Virginia 22060-5546 by the Digital Mapping Laboratory, School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213. The Contracting Officer's Representative was Mr. George Lukes.

1. Executive Summary

In July 1987, the Digital Mapping Laboratory at Carnegie Mellon University began work on a three year contract, DACA 72-87-C-0001, with the U.S. Army Engineer Topographic Laboratories to explore the detailed analysis of aerial imagery with particular emphasis on built-up areas containing large numbers of complex man-made structures. During the past three years we have performed research in several important areas including scene registration, stereo analysis, shadow analysis, and building detection. Each of these areas addresses an important set of issues toward the development of automated tools for cartographic feature extraction. This is the final technical report under contract DACA 72-87-C-0001 and describes our overall research progress during this three year period.[1]

1.1. Background

In previous reports we have described our research in monocular analysis for buildings detection and delineation. This research developed the use of intensity based cues relying on the detection of nearly right-angle corners that can be aggregrated into rectilinear shapes using lines, corners, and structures. Systems based on such techniques tend to rel y on good contrast between buildings and the adjacent terrain, as well as shape assumptions based upon composites of rectangles. It is clear that such techniques require additional information in order to be robust across a variety of image acquisition and spatial resolution conditions.

As a result, during the second contract year we began research on the detailed analysis of shadows cast by man-made structures. Our shadow analysis research has resulted in three techniques for the interpretation of monocular imagery: building prediction, grouping of related building hypotheses, and building hypothesis verification. In addition we have implemented an technique to acquire estimates of building heights using the lengths of cast shadows. Height estimation of man-made structures can be accomplished even using monocular imagery.

Previous work in stereo image analysis focused on the development of a new feature-based matching algorithm based upon hierarchical waveform analysis. Our work in stereo analysis complements the monocular feature extraction research and provides a basis for the integration of explicit three-dimensional information into built-up area analysis. During the first two years we began a major initiative to explore automatic methods for scene registration in complex aerial imagery. This research has progressed, with improved results generated using a variety of different image features.

In keeping with the theme of the use of multiple cues to provide additional information from which a more robust estimate of the building height and/or location could be derived we began

[1]This research was sponsored by the U.S. Army Engineer Topographic Laboratories under Contract DACA72-87-C-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Engineer Topographic Laboratories or of the United States Government.

research on fusion of stereo estimates. This research focused on the combination of an area-based and a feature-based method to attempt to achieve better overall height estimates in the presence of occlusions, large depth discontinuities, and in complex matching conditions. We also began to develop tools and techniques for automated performance evaluation using a manually derived three-dimensional ground-truth database. Such quantitative performance evaluation is critical for understanding the incremental performance of changes to various matching techniques, the effects of parameter selection, and in head-to-head comparisons of various end-to-end stereo systems.

Finally, we have supported a modest effort to investigate the utility of share-memory multi-processors for high-level vision. Our focus has been the exploration of task-level parallelism for a knowledge-based system that has been used to interpret airport and suburban house scenes. We have achieved near linear speedups on an Encore multimax processor for the most computationally intensive component of the system.

In Sections 2 through 6 we summarize the most recent accomplishments in each of these areas achieved over the last year. We believe that progress has been steady and that the work in shadow analysis, monocular fusion, scene registration, stereo analysis and refinement has greatly improved our suite of techniques for built-up area analysis. In the remainder of this section we summarize various technical talks, published papers, and other tangible accomplishments funded under this research contract.

1.2. Accomplishments

Our primary effort under this contract was to investigate the use of knowledge-intensive techniques for the detailed analysis of remotely sensed imagery by developing scene interpretation systems for complex urban areas. Our research has resulted in the design and implementation of several cartographic feature extraction components/systems as well as supporting work in stereo matching, information fusion, and tools for database utilization. In the process, a variety of basic research issues in computer vision and cartography have been addressed.

- Developed an information fusion paradigm based on using multiple scene domain cues to support a variety of tasks in cartographic feature extraction. These include monocular fusion of building boundary cues, refinement of stereo disparity estimated using intensity/surface material information. The common thrust of this work is to find and exploit multiple information sources, extracted from common imagery, that may contain redundant cues concerning the geometric structure of the scene.
- Developed several techniques for shadow analysis including building hypothesis generation, building hypothesis verification, and techniques to group buildings based upon their consistency with detected shadow boundaries.
- Developed an automatic scene registration capability, with improved accuracy results using a variety of different image features. Currently we are able to perform a relative orientation between stereo image pairs whose accuracy is close to human-

level performance in accuracy using manually derived ground control points.

- Our work with area-based and waveform-based stereo algorithms has continued, producing improvements in the individual results, and development of a technique for merging the results of the two methods, as well as methods for performance evaluation. Our feature-based algorithm was improved particularly with respect to waveform approximation and the use of inter-scanline consistency to detect and correct mismatches. Our performance analysis on a variety of stereo pairs (currently 10 datasets) with various disparity ranges including significant terrain relief has given us a better understanding issues in robust stereo analysis.
- Making simple assumptions about the intensity of smooth surfaces has resulted in a technique for fusing information in the disparity map with edge and intensity information to generate much improved disparity map segmentations. Analysis of the disparity image guided by the intensity image appears to be a promising technique to reject mismatches and to generate a refined disparity map that lends itself to further interpretation. This technique appears to be superior to many interpolation based methods because it explicitly takes into account the nature of surface patches with similar albedo.
- We began a modest effort to integrate ITD cartographic data into our spatial database system, CONCEPTMAP. This has resulted in the development of a flexible query system, as well as a powerful window-based user interface. This initial work has pointed the way to many issues in the efficient access to spatial data for planning, navigation, and incorporation into cartographic feature extraction systems.

1.3. Publications

Over the period of our research contract our research contract in Built-Up Area Feature Extraction we have published our results in refereed journals and conferences, and presented progress reports at various meetings. This section details the most significant publications and presentations supported under this contract.

1.4. Publications

- D. M. McKeown, Jr. (1990). "Toward Automatic Cartographic Feature Extraction", in Mapping and Spatial Modeling For Navigation, NATO ASI Series F: Computer and Systems Sciences, Vol. 65, Springer-Verlag, Edited by L. Pau, 1990, pp 149-180.
- D. McKeown, F. Perlant, and J. Shufelt, (1990). "Information Fusion in Cartographic Feature Extraction from Aerial Imagery" in *Proceedings of ISPRS Symposium on Global and Environmental Monitoring: Techniques and Impacts*, Victoria, British Columbia, Canada, September, 1990., pp. 140-144.
- J. Shufelt and D. M. McKeown, "Fusion of Monocular Cues to Detect Man-Made Structures in Aerial Imagery" in Proceedings of *IAPR Workshop on Multisource Data Integration in Remote Sensing* June, 1990.
- Y. Hsieh, F. P. Perlant, and D. M. McKeown, "Recovering 3D Information from Complex Aerial Imagery" in Proceedings of *10th International Conference on Pattern Recognition* Atlantic City, New Jersey, June, 1990. pp. 136-146.

- F. P. Perlant and D. M. McKeown, "Improved Disparity Map Analysis Through the Fusion of Monocular Image Segmentations" in Proceedings of *IAPR Workshop on Multisource Data Integration in Remote Sensing* June, 1990.
- F. P. Perlant, and D. M. McKeown (1990) "Scene Registration in Aerial Image Analysis", in Journal of Photogrammetric Engineering and Remote Sensing, Volume 56, No. 4, April, 1990, pp. 481-493.
- R. B. Irvin, and D. M. McKeown, (1989) "Methods for exploiting the relationship between buildings and their shadows in aerial imagery" in IEEE Transactions on Systems, Man and Cybernetics Volume 19, Number 6, November/December 1989, pp. 1564-1575.
- W. Harvey, D. Kalp, M. Tambe, D. McKeown, A. Newell, "Measuring the Effectiveness of Task-Level Parallelism for High-Level Vision" in Proceedings of *DARPA Image Understanding Workshop*, Palo Alto, California, May 23-26, 1989. Morgan Kaufmann Publishers., pp. 916-933.
- D. M. McKeown, Jr., Harvey, W.A., and Wixson, L. "Automating Knowledge Acquisition For Aerial Image Interpretation" Computer Vision, Graphics and Image Processing Volume 46, Number 1, April, 1989, pp 37-81.
- R. B. Irvin, and D. M. McKeown, "Methods for exploiting the relationship between buildings and their shadows in aerial imagery" in Proceedings of *SPIE Conference on Image Understanding and the Man-Machine Interface II* Los Angeles, Calif, January 17-18, 1989., Volume 1076, pp. 156-164.
- F. P. Perlant, and D. M. McKeown, "Scene Registration in Aerial Image Analysis" in Proceedings of *SPIE Conference on Reconnaissance, Astronomy, Remote Sensing and Photogrammetry* Los Angeles, Calif, January 19-20, 1989., Volume 1070, pp. 88-99.

1.5. Invited Presentations

- Keynote Speaker: "Knowledge-Based Systems for Remote Sensing" Workshop on Environmental Remote Sensing at Research Institute for Applied Knowledge Processing, FAW, Ulm, Germany, October 1-5, 1990.
- Session Chairman, "Knowledge-Based Techniques/Systems for Data Fusion", at *ISPRS Symposium on Global and Environmental Monitoring: Techniques and Impacts*, Victoria, British Columbia, Canada, September 17-21, 1990.
- "Knowledge-Based Vision, Airports, and SAR" JPL/Caltech Image Recognition Workshop, Pasadena, CA., May 17-18, 1990.
- "Toward Automatic Cartographic Feature Extraction", Machine Vision - Image Understanding Workshop, AFOSR/AFWL, Albuquerque, NM., May 15-16, 1990.
- "Automated Feature Extraction Research", Imagery Perspective Transformation Symposium, Bolling AFB, Washington D.C. May 1-2, 1990.
- Participant: DARPA IUS Working Group Meeting, Scottsdale, AZ. February 26-28, 1990.
- "Progress in Automated Cartographic Feature Extraction", U.S. Army Engineer

Topographic Laboratories, Fort Belvoir, VA., January 25, 1990.

- Tutorial: "Data Fusion Techniques for GIS and Remote Sensing", Workshop on Advances in Spatial Information Extraction and Analysis for Remote Sensing, International Society for Photogrammetry and Remote Sensing, Orono, Maine, January 15, 1990.
- "Knowledge-Based Techniques for Geographic Information Systems", AIST, Northeast AI Consortium, Syracuse, N.Y., October 23, 1989.
- "Automated Feature Extraction in Urban Areas" Project 2851 Mission Rehearsal Special Interest Group, Defense Mapping Agency Aerospace Center, St. Louis, MO, September 28, 1989.
- "Trends in Automated Cartographic Feature Extraction" NATO Advanced Research Workshop on *Mapping and Spatial Modeling For Navigation*, Fano, Denmark, August 21-25, 1989.
- Participant/Panel Leader: Specialist Meeting on Large Spatial Databases, NSF-National Center for Geographic Information and Analysis, Santa Barbara, Cal., July 19-22, 1989.
- "Artificial Intelligence in the Analysis of Aerial Imagery", IEEE Computer Society Workshop on Artificial Intelligence for Computer Vision, San Diego, Cal., June 5, 1989.
- "Product Opportunities in Cartography and Remote Sensing", DARPA IU Program Meeting, Institute for Defense Analysis, Alexandria, VA., March 13, 1989.
- Participant: DARPA Program Review Meeting, SCORPIUS Image Understanding Program, El Segundo, Cal., January 18, 1989.
- Participant: DARPA/USAETL Program Review Meeting on Spatial Databases for ADRIES/TACNAT, Fort Belvoir, VA., September 13-14, 1988.
- "Automated Feature Extraction From Aerial Imagery", EXRAND Committee Meeting, Washington, D.C., February, 9, 1988.
- Tutorial: "Spatial Interpretation of Aerial Imagery" at *IEEE Workshop on Applied Imagery Pattern Recognition*, Washington, D.C., October 28, 1987.

1.6. Researchers Supported

The following members of the Digital Mapping Laboratory, School of Computer Science, Carnegie Mellon University were fully or partially supported during the period of this research contract.

- David M. McKeown, Jr.
  Research Computer Scientist
- Aviad Zlotnick
  Post Doctoral Research Associate
- Wilson A. Harvey
  Senior Research Programmer

- Frederic P. Perlant
  Visiting Scientist
- Jean-Christophe Dhellemmes
  Visiting Scientist
- Yuan Hsieh
  Research Programmer
- Matthew Diamond
  Research Programmer
- Steven Ford
  Research Programmer
- Undergraduate Research Assistants
  Emily Burke, Bruce Irvin, Jeffrey Shufelt, Lambert Wixson

1.7. Acknowledgements

During the course of this research program we have benefited from detailed technical interactions with personnel from the U.S. Army Engineer Topographic Laboratories, Fort Belvior, VA.. Dr. Fred Rhode, Edmundo Simental, Dan Edwards, and George Lukes each acted as COTR during various phases of this contract. Each helped by providing good feedback on the relevance of our research program to the U.S. Army, and in maintaining continuity of funding and focus. We had many interesting technical discussions with various members of the Research Institute. In particular, Ray Norvelle and Dan Edwards were helpful on issues including stereo matching techniques and cartographic databases.

1.8. Organization of this Report

In the body of this final report we provide a detailed technical description of our research supported under this contract. Section 2 discusses our work in the fusion of multiple building hypothesis obtained using different feature extraction techniques into an improved set of building estimates. This represents the integration of our work on building detection using intensity cues with our work on shadow analysis for building hypothesis generation, verification and grouping. An quantitative evaluation of the various feature extraction systems and the improved results using our fusion technique is presented.

Section 3 discusses our work in automated scene registration to support stereo analysis. It builds on our research in monocular cue analysis in that it uses features such as shadow corners and building structures to provide matchable features for the registration process. We provide quantitative results that compare registration accuracy using five different feature extraction techniques with that achieved using manual matching.

Section 4 describes our results in stereo analysis using both area-based and feature-based approaches. We briefly discuss some modifications to the stereo algorithms, S1 and S2, and focus on detailed performance analysis using a 3-dimensional ground-truth disparity map

generated for several test scenes. We introduce several metrics for stereo accuracy that are relevant for the built-up area task: average error, percentage of point within +/- 1 pixel of true disparity, building height accuracy, and building delineation accuracy.

Section 5 describes research in the use of image intensity patch information with stereo height estimates to provide a basis for disparity refinement. This refinement technique can be used to improve disparity estimates by associating surface patches in the intensity image with collections of disparity points. The assumption is that these patches reflect surfaces and partial surfaces in the scene that should have a nearly homogeneous height. Statistical analysis of the disparity estimates within these regions can detect gross mismatches as well as incorrect matches due to occlusion.

Section 6 describes some preliminary research in providing user-interface support for large scale spatial databases. We began with a DMA ITD database of Fort Hood and developed tools to decode and reformat the spatial data to allow for random queries based on geographic location and/or partial matching of feature attributes. An Xwindow interface was developed for feature display and query processing.

Finally, Section 7 provides some brief conclusions on our program of research in the area of automated cartographic feature extraction.

2. Fusion of Monocular Building Hypotheses

The extraction of buildings from aerial imagery is a complex problem for automated computer vision. It requires locating regions in a scene that possess properties distinguishing them as man-made objects as opposed to naturally occurring terrain features. The building extraction process requires techniques that exploit knowledge about the structure of man-made objects. Techniques do exist that take advantage of this knowledge: various methods use edge-line analysis, shadow analysis, and stereo imagery analysis to produce building hypotheses. It is reasonable, however, to assume that no single detection method will correctly delineate or verify buildings in every scene. As an example, a feature extraction system that relies on analysis of cast shadows to predict building locations is likely to fail in cases where the sun is directly above the scene.

It seems clear that a cooperative-methods paradigm is useful in approaching the building extraction problem. Using this paradigm, each extraction technique provides information which can then be added or assimilated into an overall interpretation of the scene. Thus, our research focus is to explore the development of a computer vision system that integrates the results of various scene analysis techniques into an accurate and robust interpretation of the underlying three-dimensional scene.

This section describes research performed under DACA 72-87-C-0001 on the problem of building hypothesis fusion generated using monocular imagery. First, our building extraction techniques are briefly surveyed, including four building extraction, verification, and clustering systems that form the basis for the work described here. A method for fusing the symbolic data generated by these systems is described, and applied to monocular image and stereo image data sets. Evaluation methods for the fusion results are described, and the fusion results are analyzed using these methods.

2.1. Building extraction techniques

Under this research contract, we have developed several techniques for the extraction of man-made objects from aerial imagery. One common goal of these techniques is to organize the image into manageable parts for further processing, by using external knowledge to organize these parts into regions. A set of four monocular building detection and evaluation systems were used. Three of these were shadow-based systems; the fourth was line-corner based. The shadow based systems are described more fully by Irvin and McKeown [1], and the line-corner system is described by Aviad, McKeown, and Hsieh [2]. A brief description of each of the four detection and evaluation systems follows.

BABE (Builtup Area Building Extraction) is a building detection system based on a line-corner analysis method. BABE starts with intensity edges for an image, and examines the proximity and angles between edges to produce corners. To recover the structures represented by the corners, BABE constructs chains of corners such that the direction of rotation along a chain is either clockwise or counterclockwise, but not both. Since these chains may not necessarily form closed segmentations, BABE generates building hypotheses by forming boxes out of the individual lines

that comprise a chain. These boxes are then evaluated in terms of size and line intensity constraints, and the best boxes for each chain are kept, subject to shadow intensity constraints [3], [4].

SHADE (SHAdow DEtection) is a building detection system based on a shadow analysis method. SHADE uses the shadow intensity computed by BABE as a threshold for an image. Connected region extraction techniques are applied to produce segmentations of those regions with intensities below the threshold, i.e., the shadow regions. SHADE then examines the edges comprising shadow regions, and keeps those edges that are adjacent to the buildings casting the shadows. These edges are then broken into nearly straight line segments by the use of an imperfect sequence finder [5]. Those line segments that form nearly right-angled corners are joined, and the corners that are concave with respect to the sun are extended into parallelograms, SHADE's final building hypotheses.

SHAVE (SHAdow VErification) is a system for verification of building hypotheses by shadow analysis. SHAVE takes as input a set of building hypotheses, an associated image, and a shadow threshold produced by BABE. SHAVE begins by determining which sides of the hypothesized building boxes could possibly cast shadows, given the sun illumination angle, and then performs a walk away from the sun illumination angle for every pixel along a building/shadow edge to delineate the shadow. The edge is then scored based on a measure of the variance of the length of the shadow walks for that edge. These scores can then be examined to estimate the likelihood that a building hypothesis corresponds to a building, based on the extent to which it casts shadows.

GROUPER is a system designed to cluster, or group, fragmented building hypotheses, by examining their relationships to possible building/shadow edges. GROUPER starts with a set of hypotheses and the building/shadow edges produced by BABE. GROUPER back-projects the endpoints of a building/shadow edge towards the sun along the sun illumination angle, and then connects these projected endpoints to form a region of interest in which buildings might occur. GROUPER intersects each building hypothesis with these regions of interest. If the degree of overlap is sufficiently high (the criteria is currently 75% overlap), then the hypothesis is assumed to be a part of the structure which is casting the building/shadow edge. All hypotheses that intersect a single region of interest are grouped together to form a single building cluster.

2.2. A simple hypothesis merging technique

Building hypotheses typically take the form of geometric descriptions of objects in the context of an image. One can imagine "stacking" sets of these geometric descriptions on the image: in the process, those regions of the image that represent man-made structure in the scene should accumulate more building hypotheses than those regions of the image that represent natural features in the scene. The merging technique developed here exploits this idea.

The method takes as input an arbitrary collection of polygons. An image is created that is

sufficiently large to contain all of the polygons, and each pixel in this image is initialized to zero. Each polygon is scan-converted into the image, and each pixel touched during the scan is incremented. The resulting image then has the property that the value of each pixel in the image is the number of input polygons that cover it.

Segmentations can then be generated from this "accumulator" image by applying connected region extraction techniques. If the image is thresholded at a value of 1 (i.e. all non-zero pixels are kept), the regions produced by a connected region extraction algorithm will simply be the geometric unions of the input polygons. It is the case, however, that the image could be thresholded at higher values. We motivate thresholding experiments in Section 2.3.4.

2.3. Merging multiple hypothesis sets

We briefly describe some of the experiments performed with the scan-conversion hypothesis fusion technique. The procedure used to apply this technique to the results of four building detection and evaluation systems (BABE, SHADE, SHAVE, and GROUPER) is described. A technique for quantitative evaluation of building hypotheses is described, and applied to the hypothesis fusion results. These results are analyzed to suggest improvements to the fusion technique.

2.3.1. The merging technique applied to four extraction systems

There were two merging problems under consideration. The first of these was the creation of a single hypothesis out of a collection of fragmented hypotheses believed to correspond to a single man-made structure. This problem was addressed by applying the scan-conversion technique to the fragmented clusters produced by GROUPER. The technique was applied to each cluster individually, and the resulting accumulator image was thresholded at 1, and connected region extraction techniques were applied to provide the geometric union of each cluster. These clusters were then used as the building hypotheses produced by GROUPER.

The second problem was the fusion of each of these monocular hypothesis sets into a single set of hypotheses for the scene. Again, the scan-conversion technique was applied. The four hypothesis sets were scan-converted, and the resulting accumulator image was thresholded at 1. Connected region extraction techniques were applied to produce the final segmentation for the image.

Figure 2-1 shows a section of a suburban area in Washington, D.C. Figure 2-2 shows the SHADE results for this scene. Figure 2-3 shows the SHAVE results. Figure 2-4 shows the GROUPER results, and Figure 2-5 shows the BABE results. Figure 2-6 shows the fusion of these four monocular hypothesis sets.

2.3.2. Evaluation of the technique

To judge the correctness of an interpretation of a scene, it is desirable to have some mechanism for quantitatively evaluating that interpretation. One approach is to compare a given set of hypotheses against a set that is known to be correct, and analyze the differences between the given set of hypotheses and the correct ones. In performing evaluations of the fusion results, we use *ground-truth segmentations* as the correct detection results for a scene. Ground-truth segmentations are manually produced segmentations of the buildings in an image.

Figure 2-1: DC37 image with ground-truth segmentation

There exist two simple criteria for measuring the degree of similarity between a building hypothesis and a ground-truth building segmentation: the mutual area of overlap and the difference in orientation. A correct building hypothesis and the corresponding ground-truth segmentation region should cover roughly the same area, and should have roughly the same

alignment with respect to the image. A scoring function can be developed that incorporates these criteria. A region matching scheme such as this, however, suffers from the fact that multiple buildings in the scene are segmented by a single region in the hypothesis set. In these cases, the building hypothesis will have low matching scores with each of the buildings it contains, due to the differences in overlap area.

A simpler coverage-based global evaluation method was developed. This evaluation method works in the following manner. H, a set of building hypotheses for an image, and G, a ground-truth segmentation of that image, are given. The image is then scanned, pixel by pixel. For any pixel P in the image, there are four possibilities:

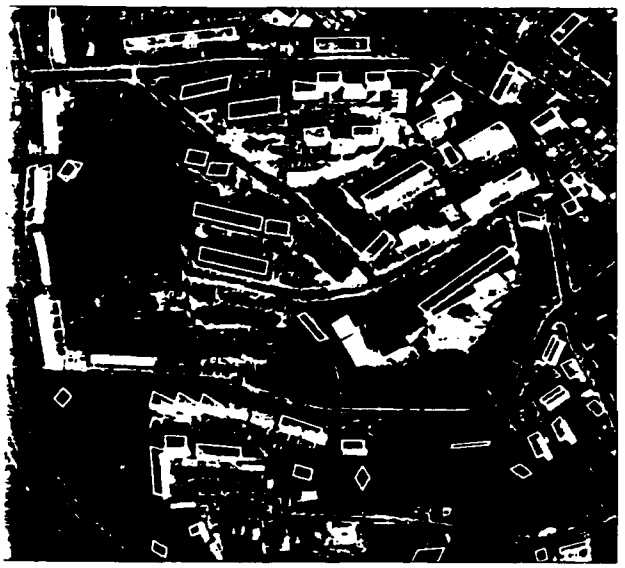

Figure 2-2: DC37 SHADE results

Figure 2-4: DC37 GROUPER results

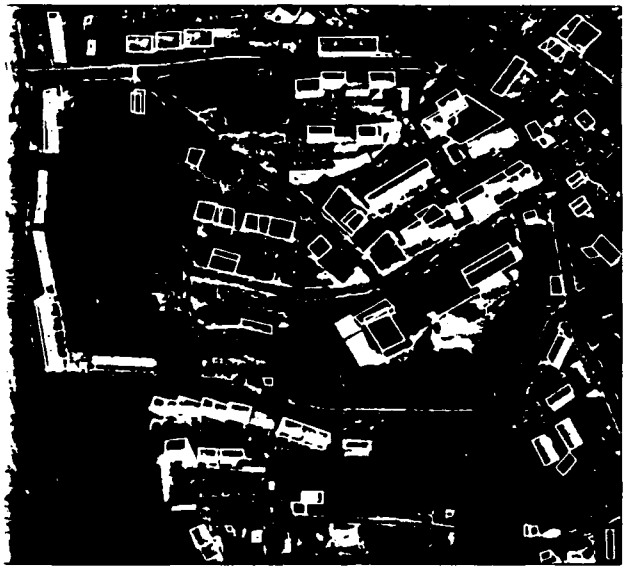

Figure 2-3: DC37 SHAVE results

Figure 2-5: DC37 BABE results

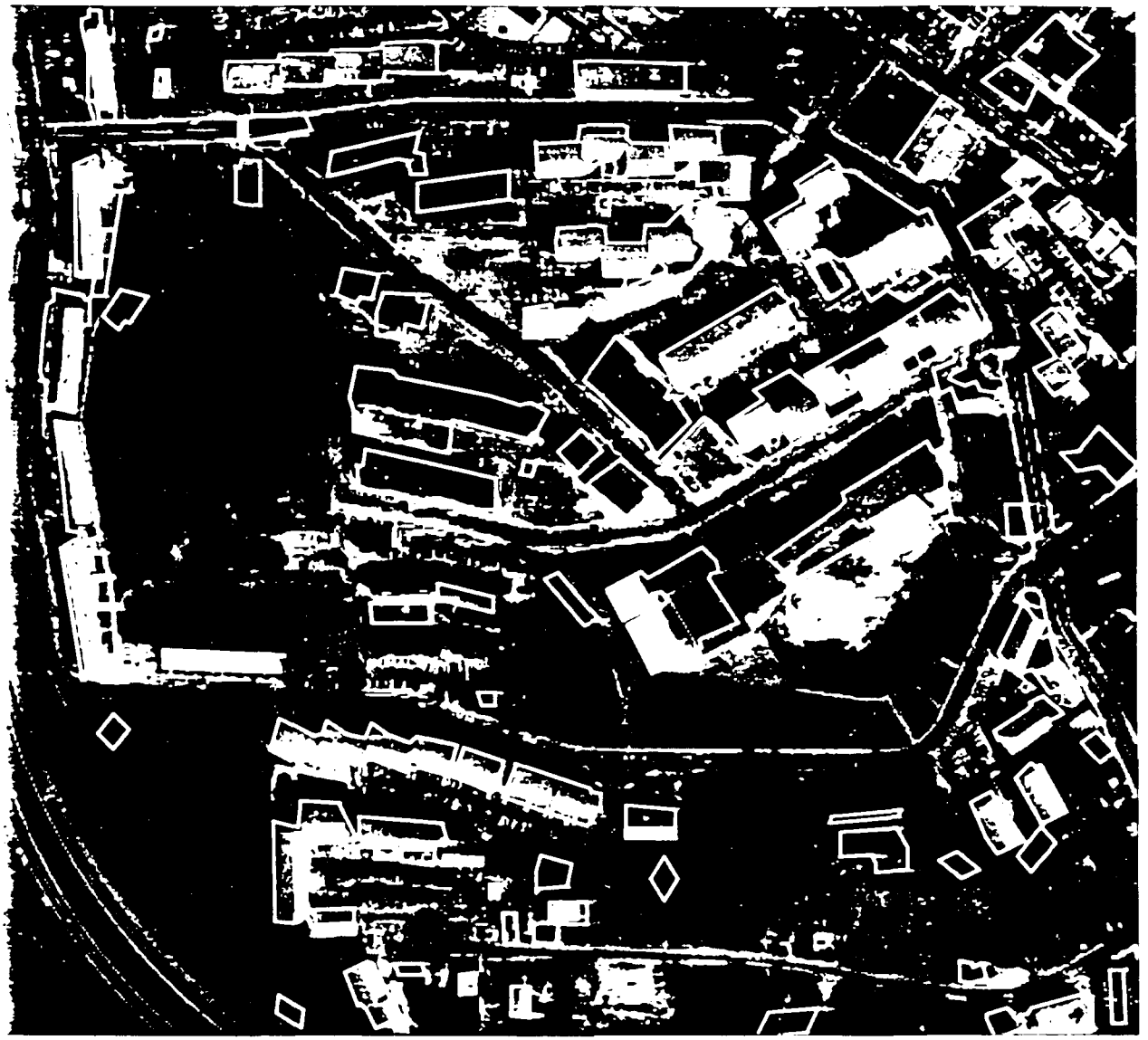

Figure 2-6: Monocular hypothesis fusion for DC[37]

Neither a region in H nor a region in G covers P. This is interpreted to mean that the system producing H correctly denoted P as being part of the background or natural structure of the scene.

- No region in H covers P, but a region in G covers P. This is interpreted to mean that the system producing H did not recognize P as being part of a man made structure in the scene. In this case, the pixel is referred to as a "false negative."

- A region (or regions) in H cover P, but no region in G covers P. This is interpreted to mean that the system producing H incorrectly denoted P as belonging to some man made structure, when it is in fact part of the scene's background. In this case, the pixel is referred to as a "false positive."

- A region (or regions) in H and a region in G both cover P. This is interpreted to

mean that the system producing H correctly denoted P as belonging to a man-made structure in the scene.

By counting the number of pixels that fall into each of these four categories, we may obtain measurements of the percentage of building hypotheses that were successful (and unsuccessful) in denoting pixels as belonging to man-made structure, and the percentage of the background of the scene that was correctly (and incorrectly) labeled as such. Further, we may use these measurements to define a *building pixel branching factor*, which will represent the degree to which a building detection system overclassifies background pixels as building pixels in the process of generating building hypotheses. The building pixel branching factor is defined as the number of false positive pixels divided by the number of correctly detected building pixels.

2.3.3. Results and analysis

The fusion process was run on other scenes in addition to the DC37 scene: DC36A, DC36B, and DC38, three more scenes from the Washington, D.C. area; and LAX, a scene from the Los Angeles International Airport. The coverage-based evaluation program was then applied to generate Tables 2-1 through 2-5. Each table gives the statistics for a single scene. The first column represents a building extraction system. The next two columns give the percentage of building and background terrain correctly identified as such. The fourth and fifth columns show incorrect identification percentages for buildings and terrain. The next two columns give the breakdown (in percentages) of incorrect pixels in terms of false positives and false negatives. The last column gives the building pixel branching factor.

Evaluation results for the fusion process on DC37							
System	% Bld Detected	% Bkgd Detected	% Bld Missed	% Bkgd Missed	% False Pos.	% False Neg.	Br Factor
SHADE	37.5	98.2	62.5	1.8	15.0	85.0	0.294
SHAVE	47.2	96.8	52.8	3.2	26.8	73.2	0.408
GROUPER	48.7	95.8	51.3	4.2	32.6	67.4	0.508
BABE	58.9	97.2	41.1	2.8	28.5	71.5	0.278
FUSION	77.7	92.0	22.3	8.0	68.0	32.0	0.611
99 regions in ground truth							

Table 2-1: Evaluation statistics for DC37 hypothesis fusion

We note that the quantitative results generated by the new evaluation method accurately reflect the visual quality of the set of building hypotheses. Further, the building pixel branching factor provides a rough estimate of the amount of noise generated in the fusion process. Judging by these measures, we note that the final results of the hypothesis fusion process significantly improve the detection of buildings in a scene. In all of the scenes, the detection percentage for the final fusion is greater than the same percentage for any of the individual extraction system hypotheses, although the building pixel branching factor also increases due to the accumulation

Evaluation results for the fusion process on DC36A							
System	% Bld Detected	% Bkgd Detected	% Bld Missed	% Bkgd Missed	% False Pos.	% False Neg.	Br Factor
SHADE	53.8	97.0	46.2	3.0	30.7	69.3	0.381
SHAVE	63.6	96.2	36.4	3.8	41.8	58.2	0.411
GROUPER	58.0	95.8	42.0	4.2	40.6	59.4	0.495
BABE	51.0	97.9	49.0	2.1	22.1	77.9	0.273
FUSION	80.9	91.9	19.1	8.1	74.3	25.7	0.682
51 regions in ground truth							

Table 2-2: Evaluation statistics for DC36A hypothesis fusion

Evaluation results for the fusion process on DC36B							
System	% Bld Detected	% Bkgd Detected	% Bld Missed	% Bkgd Missed	% False Pos.	% False Neg.	Br Factor
SHADE	29.8	93.8	70.2	6.2	46.3	53.7	2.034
SHAVE	28.4	96.7	71.6	3.3	31.3	69.7	1.146
GROUPER	10.3	96.8	89.7	3.2	25.9	74.1	3.027
BABE	9.9	98.8	90.1	1.2	11.3	88.7	1.159
FUSION	49.8	89.2	50.2	10.8	67.8	32.2	2.126
133 regions in ground truth							

Table 2-3: Evaluation statistics for DC36B hypothesis fusion

Evaluation results for the fusion process on DC38							
System	% Bld Detected	% Bkgd Detected	% Bld Missed	% Bkgd Missed	% False Pos.	% False Neg.	Br Factor
SHADE	51.3	97.4	48.7	2.6	13.2	86.8	0.144
SHAVE	43.1	95.3	56.9	4.7	19.1	80.9	0.311
GROUPER	54.6	95.8	45.4	4.2	21.0	79.0	0.221
BABE	44.7	96.0	55.3	4.0	17.3	82.7	0.260
FUSION	74.7	90.6	25.3	9.4	51.5	48.5	0.360
53 regions in ground truth							

Table 2-4: Evaluation statistics for DC38 hypothesis fusion

of delineation errors from the various input hypotheses.

It is worth noting that the results for the DC36B scene (Table 2-3) are substantially worse than those of the other scenes. This is in large part due to the fact that the DC36B scene has a low

Evaluation results for the fusion process on LAX							
System	% Bld Detected	% Bkgd Detected	% Bld Missed	% Bkgd Missed	% False Pos.	% False Neg.	Br Factor
SHADE	34.4	99.0	65.6	1.0	10.1	89.9	0.213
SHAVE	54.1	94.9	45.9	5.1	43.6	56.4	0.655
GROUPER	46.0	98.5	54.0	1.5	16.5	83.5	0.232
BABE	63.3	98.8	36.7	1.2	18.3	81.7	0.130
FUSION	73.0	92.9	27.0	7.1	65.0	35.0	0.687
26 regions in ground truth							

Table 2-5: Evaluation statistics for LAX hypothesis fusion

dynamic range of intensities, and the component systems used for these fusion experiments are inherently intensity-based. The building pixel branching factors reflect the poor performance of the component systems; in GROUPER's case, over 3 pixels are incorrectly hypothesized as building pixels for every correct building pixel. The fusion process, however, improved the building detection percentage noticeably over the percentages of the component systems.

We also note that several difficulties are attributable to performance deficiencies in the systems producing the original building hypotheses. The shadow-based detection and evaluation systems, SHADE and SHAVE, both use a threshold to generate "shadow regions" in an image. This threshold is generated automatically by BABE, a line-corner based detection system. In some cases, the threshold is too low, and the resulting shadow regions are incomplete, which results in fewer hypothesized buildings.

GROUPER, the shadow-based hypothesis clustering system, clusters fragmented hypotheses by forming a region (based on shadow-building edges) in which building structure is expected to occur. This region is typically larger than the true building creating the shadow-building edge, and incorrect fragments sometimes fall within this region and are grouped with correct fragments. The resulting groups tend to be larger than the true buildings, and thus produce a fair number of false positive pixels.

SHAVE scores a set of hypotheses based on the extent to which they cast shadows, and then selects the top fifteen percent of these as "good" building hypotheses. In some cases, buildings whose scores fell in the top fifteen percent actually had relatively low absolute scores. This resulted in the inclusion of incorrect hypotheses in the final merger.

SHADE uses an imperfect sequence finder to locate corners in the noisy shadow-building edges produced by thresholding. The sequence finder uses a threshold value to determine the amount of noise that will be ignored when searching for corners. In some situations, the true building corners are sufficiently small that the sequence finder regards them as noise, and as a result, the final building hypotheses can either be erroneous or incomplete.

2.3.4. Thresholding the accumulator image

As part of the scan-conversion fusion process, an accumulator image is produced which represents the "building density" of the scene. More precisely, each pixel in the image has a value, which is the number of hypotheses that overlapped the pixel. Pixels with higher values represent areas of the image that have higher probability of being contained in a man-made structure. Theoretically, thresholding this image at higher values and then applying connected region extraction techniques would produce sets of hypotheses containing fewer false positives, and these hypotheses would only represent those areas that had a high probability of corresponding to structure in the scene.

To test this idea, the accumulator images for each of the six scenes were thresholded at values of 2, 3, and 4, since four systems were used to produce the final hypothesis fusion. Connected region extraction techniques were then applied to these thresholded images to produce new hypothesis segmentations. The new evaluation method was then applied to these new hypotheses.

In each of the scenes, increasing the threshold from its default value of 1 to a value of 2 causes a reduction of roughly 20 percent in the number of correctly detected building pixels. This suggests that a fair number of hypothesized building pixels are unique; i.e., several pixels can only be correctly identified as building pixels by one of the detection methods. Another interesting observation is that the building pixel branching factor roughly doubles every time the threshold is decremented. These observations suggest that thresholding alone may eliminate unique information produced by the individual detection systems, and that more work will need to be done to limit the number of false positives (and erroneous delineations) produced by each system, and by the final fusion as a whole.

3. Automated Scene Registration

The primary goal of stereo photogrammetry is to determine the three-dimensional position of any object point that is located in the overlap area of two images taken from two different camera positions. The determination of the orientation of each camera at the moment of exposure and the relationship between the cameras is a necessary step in the photogrammetric process. The camera orientation determines the relationship between the image points and ground points in the scene. The classical epipolar geometry for stereo imagery establishes a very simple spatial relationship between corresponding points in the left and right images. The solution to the general camera orientation problem has four components: the interior orientation, the exterior orientation, the relative orientation, and the absolute orientation. In this section we describe our research progress towards a fully automated scene registration system that provides a relative orientation between two stereo images. This orientation allows us to resample the right image of the stereo pair into epipolar geometry so that stereo matching can proceed.

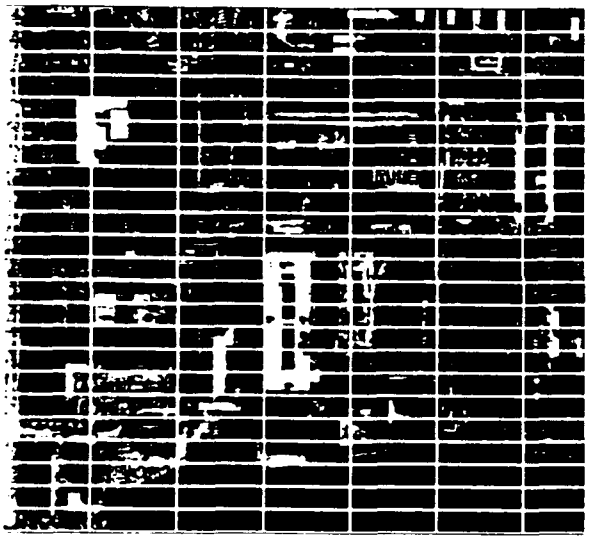

Figure 3-1: Left image DC38008 with CONCEPTMAP database registration

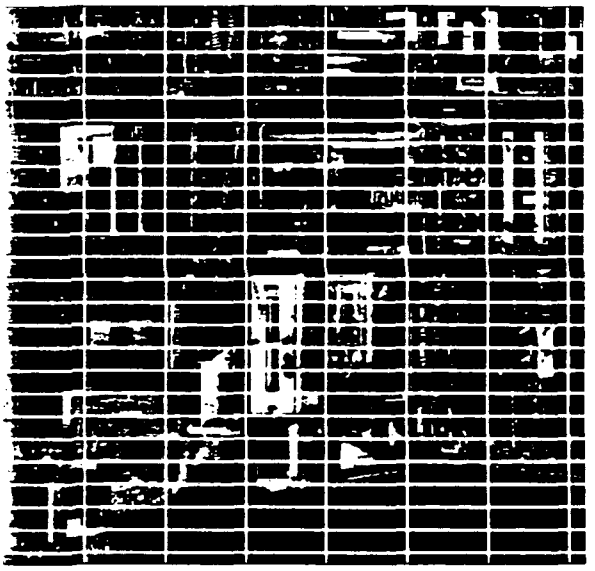

Figure 3-2: Right image DC38007 with CONCEPTMAP database registration

Figure 3-1 and 3-2 show a stereo image pair of an industrial area taken from the CONCEPTMAP database. These images were digitized from standard nine inch format mapping photography taken at the altitude of 2000 meters using a camera with a 153 millimeter lens. One pixel in the image approximately corresponds to 1.3 meters on the ground. The left image is a 512 x 512 sub-area selected from a 2300 x 2300 image. The right image sub-area was generated by calculating the <latitude,longitude> for the corner points of the left image and projecting those points onto the complete right image. This projection is then used to extract the image sub-area from the complete right image. We have superimposed a set of gridlines on both images in order to make it easier to see the actual misregistration.

3.1. Automatic selection using different features

Clearly, one requirement for automated registration is the automatic selection of corresponding points in the stereo pair images. There are two problems that must be solved. First we must automatically detect potential landmarks in each image, and then we must determine those landmarks that have been found in both images. General landmark matching is an unsolved problem and most automatic registration techniques rely on the matching of characteristic points [6] that often have no physical significance or relationship with the landmarks.

There are some important criteria for automated control point selection. First, since the elevation of the control points is not known and we are using a simple geometrical model, it is important that the set of selected control points lie approximately in the same elevation plane. Second, the selection of control points should not rely on a single type of scene domain feature, such as road intersections, since not all control point features are abundant in all scenes. For example, in urban scenes there are often many buildings and shadow regions available as candidate control points, and they are usually well distributed throughout the imagery. However, in airport scenes elongated line pairs and uniform intensity regions appear to be a better choice. In any case we use an iterative selection algorithm [7] that converges to a consistent set of control points that are usually a small subset of all of the possible matches in the stereo pair.

Another advantage of using multiple features for control point estimation is that the results of feature matching can be used to estimate the disparity range of the scene. Once the scene is registered, all matched features can be remapped to the new coordinate frame. It is then possible to calculate the disparity of each feature. Since all features are not at the same height, we automatically obtain a rough estimate of the disparity range for this scene. This disparity range estimate is directly used by the stereo matching algorithms to control search for corresponding points and can greatly reduce initial matching errors. In most research stereo systems the disparity range is either manually provided or it is set to what is considered to be a "sufficiently large" value. The drawback of the former approach is that it introduces a difficult manual step in that the entire stereo model must be searched to find the minimum and maximum disparity points. The latter situation can influence the accuracy of the resulting stereo matching algorithm by causing some matches to be never considered, or decrease the efficiency by allowing large areas to be searched for which correct matches are impossible.

For this experiment, we assume that a coarse registration of the two images has already been performed. Using this coarse correspondence, we are able to limit the search to find corresponding features in the images. Most of the remaining error is translational rather than rotational which simplifies the determination of corresponding points. Candidates for automatic control point generation include shadow corners, shadow regions, BABE monocular building hypotheses, uniform intensity regions, and elongated line structure pairs:

Shadow corners: Shadow corners are good candidates for automatic detection and correspondence as well as for manual selection. We use corners produced by the BABE system. After removing corners that are inconsistent with shape and orientation constraints imposed by

the sun direction angle and estimated shadow intensity, we select sets of shadow corners in both the left and right images. Figure 3-3 shows the corners found in the left image in white. The right image corners are shown in black and are projected onto the left image using the coarse registration. Those pairs of shadow corners that are matched are shown as connected by a white line whose endpoint circles indicate the conjugate points provided to the registration process.

Building hypotheses: Control points can also be defined geometrically with respect to features or structures extracted from the imagery. Building hypotheses generated by a monocular analysis system such as BABE can be used as match features. The center of mass of these structures is defined as the corresponding control points. Compared to shadow corners, control points defined by hypothesized buildings are not always accurate, but disambiguation of buildings is easier. Properties such as shape, size, and perimeter are good criteria that are not available for point features such as shadow corners. Figure 3-4 shows the BABE boxes in the left and right images with the matched features marked in the same manner as Figure 3-3.

Other scene features: We performed experiments to obtain control points from shadow regions, edges, and segmented regions using simple histogram analysis. In each case, control points are defined as the center of mass of the structures. Shadow regions are extracted with traditional connected component extraction techniques, using an estimate of shadow intensity provided by BABE [1]. Due to variation in the shape of the shadows, shadow regions usually give poor results in complex urban scenes with very high buildings. This variation of shape is caused by occlusion of the shadow by tall structures. They can be very reliable, however, in suburban house images where buildings are separated and have simple roof profiles. Edges are another feature extracted by BABE. Only edges with significant length are used as candidates for matching. The criteria for edge matching are edge orientation, length and the intensity gradient across the edge. Figure 3-5 shows the significant lines extracted and matched in the industrial scene. Finally, unique bright points in the scene can be used to form bright blob regions. The intensity threshold for blob regions is determined by successively decreasing the intensity scale until enough regions are extracted. These features turned out to be useful for scenes with few or no man-made structures, where shadow corners, hypothesized buildings and shadow regions failed to generate enough matching candidates.

Figure 3-6 shows the superposition of BABE results using the refined registration from Figure 3-4. The offset between building hypotheses is now primarily in the column direction and can be attributed to the displacement of the building in the left and right image due to their height. In many cases we have been able to automatically reduce the row offset error to sub-pixel accuracy from an initial displacement of 15 to 20 rows in the coarse CONCEPTMAP registration.

3.2. Evaluation of automatic registration

Table 3-1 shows the local accuracy of the different scene registrations performed on the industrial scene shown in Figures 3-1 and 3-2. POLY means that actual registration is performed using a polynomial fit, whereas ISO means that the images are registered using an isometric

solution. Coarse registration is the result of CONCEPTMAP registration. Using a set of manually selected control points we are able to evaluate the accuracy of each registration in terms of row offset compared to the ideal epipolar geometry (corresponding points on the same scanlines). Polynomial approximation performs better overall than isometric approximation, but it is more sensitive to noise. Further, the isometric approximation only requires three control points. For this scene, there are enough points from any of the match features to compute a second order polynomial approximation. The resulting accuracy is comparable with that achieved using manual selection of control points.

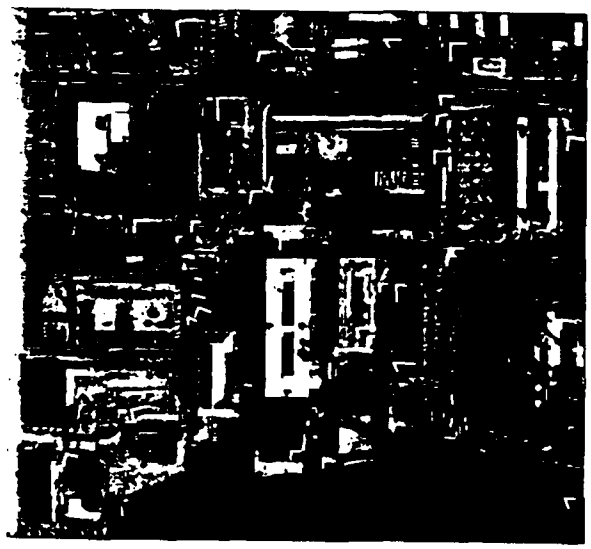

Figure 3-3: Shadow corners selected

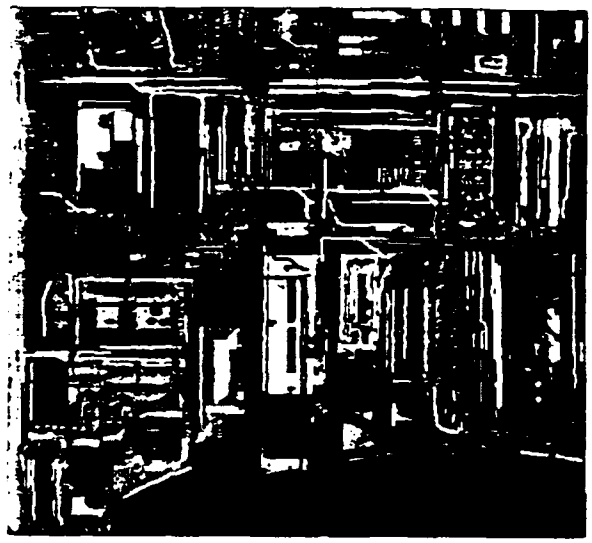

Figure 3-5: Significant lines selected

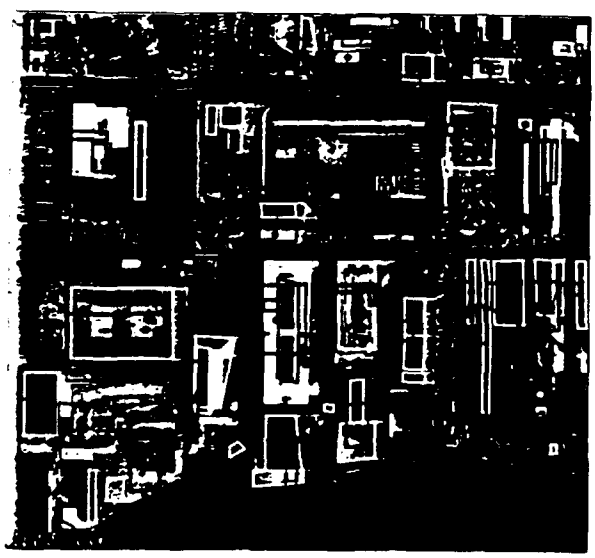

Figure 3-4: BABE building hypotheses selected

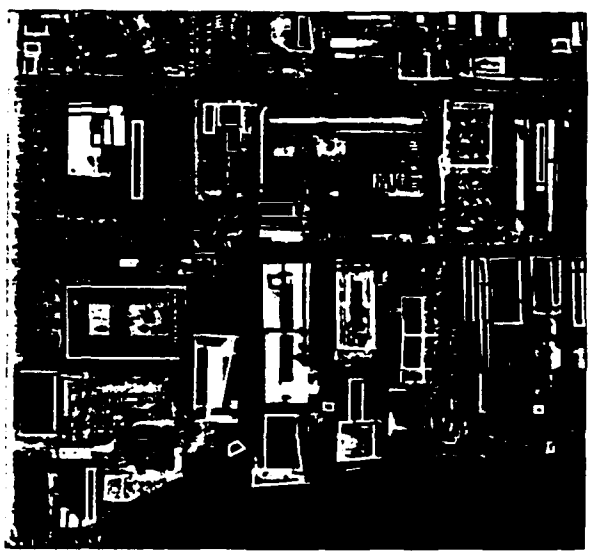

Figure 3-6: Fine registration using BABE points

In summary, scene registration is a key initial step in many tasks involving the automated interpretation of aerial images. Stereo analysis requires particular care in scene registration because of the geometric assumptions made by most stereo matching algorithms and their inability to recognize and recover from registration errors. Such registration errors usually end up reflected as gross errors in the stereo match. As a part of our goal to produce three-dimensional interpretations of complex urban scenes we have found it necessary to develop registration techniques that are accurate and robust across a variety of scene domains. We have tested our system on airport scenes, urban scenes, and suburban housing developments with varying degrees of success. Under this contract we began to investigating methods to evaluate the distribution of control points and to incorporate this evaluation into the registration system. Another area for future research is to improve our ability to recover more feature-based control point descriptions based upon other feature extraction systems, such as road detection and tracking [8].

Statistics on the quality of different registration for DC38008						
Type of Registration	Number of points	Avg. row offset	Std. row offset	Min/Max row off.	Avg. col offset	Std. col offset
Coarse	-	-20.4	1.6	-23/-16	0.4	1.2
POLY manual	11	0.1	0.3	-1/1	0.1	0.5
POLY corner	20	0.5	0.6	0/2	-0.5	1.2
POLY structure	14	-0.8	0.8	-2/2	-5.2	2.0
POLY edge	17	0.8	0.7	0/3	0.0	1.8
POLY shadow	12	-0.6	0.8	-2/1	-0.4	0.9
POLY blob	17	0.6	0.6	0/2	-0.6	1.1
ISO manual	11	-0.4	0.6	-1/1	0.6	1.4
ISO corner	20	1.0	0.5	0/3	2.7	1.3
ISO structure	14	-1.7	0.9	-3/1	-2.9	1.2
ISO edge	17	1.3	0.9	0/4	0.9	1.2
ISO shadow	12	-0.2	1.1	-2/2	3.8	1.7
ISO blob	17	0.6	1.6	-2/5	1.4	1.2

Table 3-1: Statistics for different registrations on DC38008 stereo pair

4. Stereo Analysis for Urban Scenes

Algorithms for stereo correspondence can be grouped into two major categories: area-based and feature-based matching [9]. Both classes of techniques, area-based and feature-based, have advantages and drawbacks that primarily depend on the task domain and the three-dimensional accuracy required. For complex urban scenes, feature-based techniques appear to provide more accurate information in terms of locating depth discontinuities and in estimating height. However, area-based approaches tend to be more robust in scenes containing a mix of buildings and open terrain.

We do not believe that any one technique is likely to be robust enough to perform well in the diverse set of scenes found in urban areas. For this reason we have developed two stereo matching algorithms that have complementary behaviors. In this section we describe modifications to S1, an existing area-based algorithm that uses the method of differences matching technique developed by Lucas [10, 11]. We also describe S2, a new feature-based technique that uses a scanline matching method which treats each epipolar scanline as an intensity waveform. The technique matches peaks and troughs in the left and right waveform. Both are hierarchical and use a coarse-to-fine matching approach. Each is quite general, as the only constraint imposed is the order constraint for the feature-based approach. The order constraint should generally be satisfied in our aerial imagery except in cases of hollowed structures.

4.1. Modifications to the S1 Stereo Algorithm

The S1 area-based approach uses a hierarchical set of reduced resolution images to perform coarse-to-fine matching on small windows in the two images. At each level the size of the windows for the matching process depends on the resolution of the reduced image. An initial disparity map is generated at the first level. Subsequent matching results computed at successively finer levels of detail are used to refine the disparity estimate at each level. Therefore the amount of error in the scene registration that can be tolerated by this matching algorithm depends on the size of the matching windows. However, since there is a relationship between the matching window size and the level of accuracy, simply using larger matching windows may not be desirable.

To accommodate large disparities, we modified the algorithm to use a hierarchy of different spatial resolutions. Starting with a reduced resolution dataset we compute an initial estimate of the scene disparity. With this estimate of disparity as an initial starting point, we can better refine our estimate than if we had begun matching at a coarser level. The disparity range of the scene can be used to estimate the number of different spatial resolutions, the number of levels for each resolution, and the size of the smoothing windows and scanning overlap at each level. A good estimate of the disparity range can be provided by shadow analysis, BABE box matching, or external knowledge of the terrain. We have found that good estimates of the disparity range are necessary to achieve reasonable results. This approach has been used on different images and gives better results than the standard S1 method. The results are less sensitive to registration

errors and we obtain better results on the discontinuities.

As a final step in the S1 analysis we modified the algorithm to improve the detection of the disparity discontinuities. We first compute variational left and right images using a local variation operator [12]. As an initial disparity estimate, we then use the result of the previous method and rerun the S1 procedure using just two resolution levels with the variational images to encompass errors in the previous result, and thereby locally refine the disparity estimate.

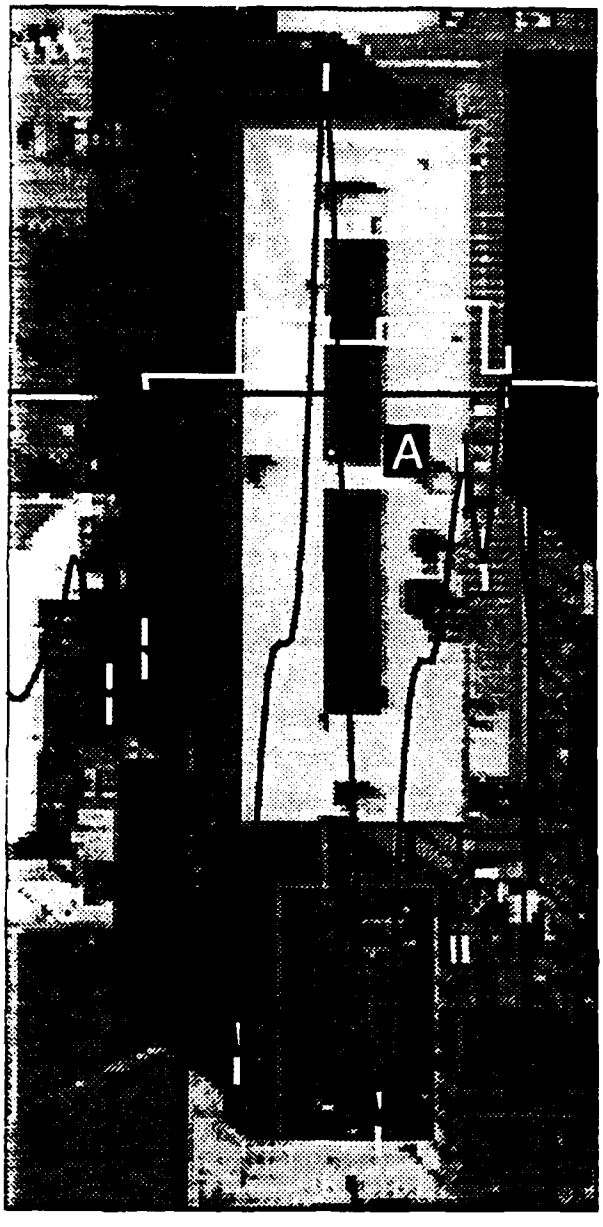

Figure 4-1: Gradient Wave Matched Points [Left]

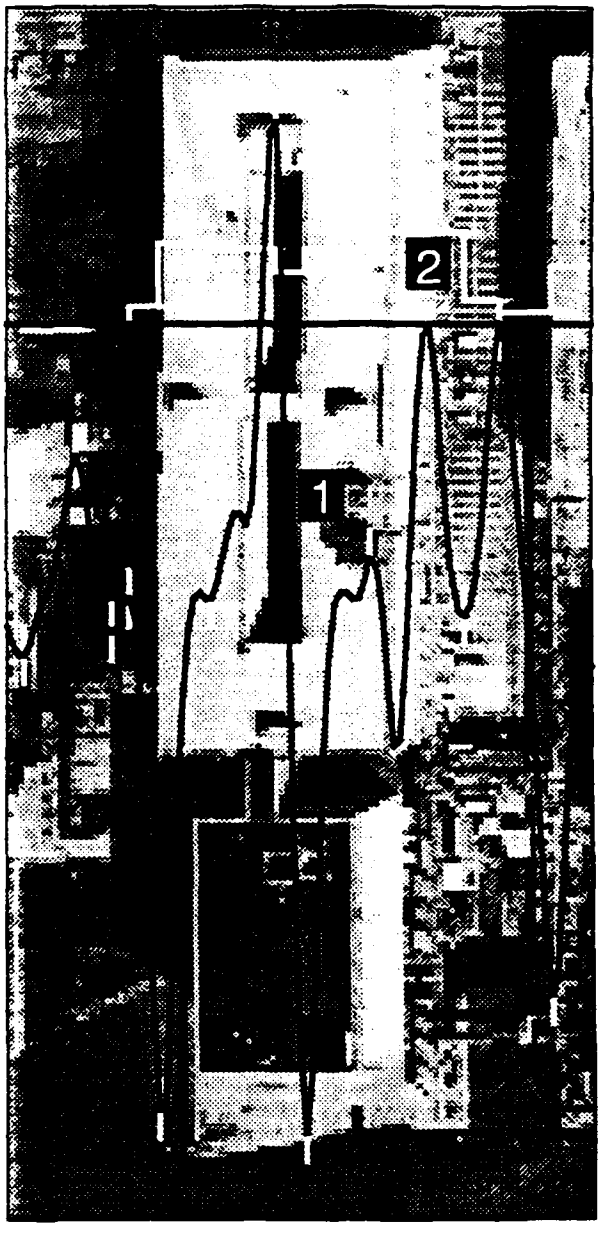

Figure 4-2: Gradient Wave Matched Points [Right]

4.2. The S2 Stereo Algorithm

S2 is a feature-based system that treats the problem of stereo matching as one-dimensional signal matching. S2 matches epipolar scanlines in the left and right image using a hierarchical approximation of the scanline intensity waveform. It matches peaks and valleys in the waveform at different levels of resolution. S2 uses intra-scanline consistency to enforce a linear ordering of matches without order reversals. It also applies an inter-scanline consistency that considers the matches in adjacent scanlines. Application of the inter-scanline constraint is used to increase the confidence of matches found to be consistent across multiple scanlines and to delete improbable matches. Since disparity discontinuity usually occurs at the intensity discontinuity, the gradient waveform is matched after the intensity matching phase to localize disparity jumps. Finally, efforts are made to detect occlusions and correct them.

The features used for matching are the intensity and gradient extremities of the scanlines. The matching criteria is simply the similarity between two extremities. Intensity extremities are easier to match than the gradient extremities, because intensity extremities vary in size and shape more so than the gradient extremities. However, intensity features may not correspond to the position of physical objects in the scene, so the gradient, the derivative of the intensity peak, is matched. Figures 4-1 and 4-2 show the left and right waveform for a single image scan-line. The horizontal black line is the scan-line being matched, the horizontal white line is the interpolated disparity profile for the scanline, and the black waveform is the gradient waveform. Minima and maxima that have been matched are marked in white.

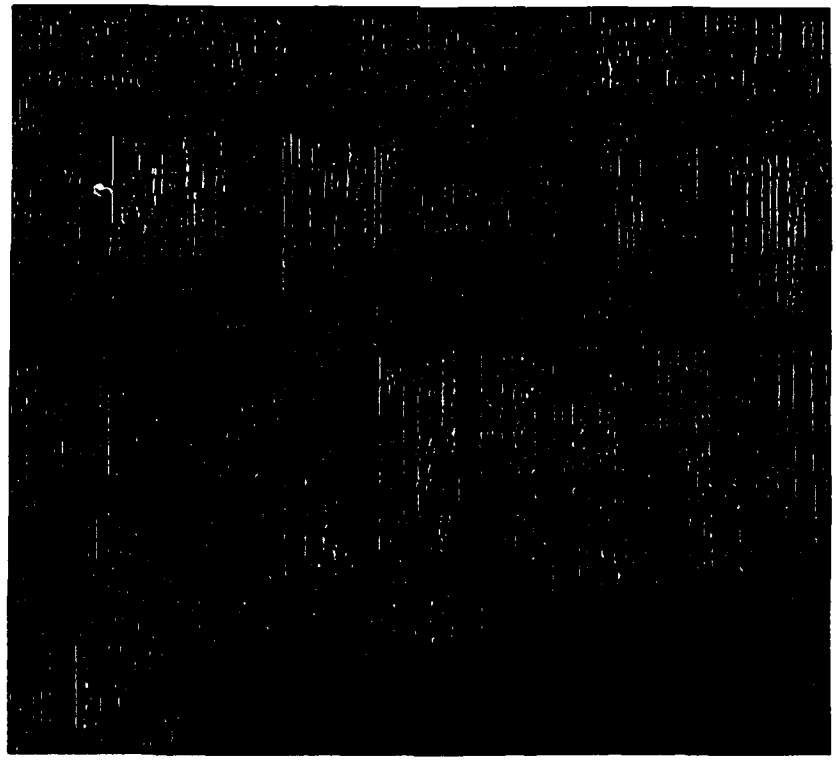

Figure 4-3: S2 sparse disparity map

Intensity features are matched hierarchically. In other words, S2 matches the most significant features first, such as points with highest and lowest intensity values. Points with succeeding values are matched later using matches at the previous coarse level as constraints. Due to the locality of matching algorithm, the optimum matches at the waveform level might not be desirable or correct from a global point of view. It is precisely for this reason that inter- and intra-scanline consistency constraints are imposed during the intensity matching phase. Inter-scanline consistency simply assumes that disparity should be nearly continuous across the scanlines. Intra-scanline assumes continuity along the scanline, unless there are strong supports for the disparity jump. The intensity waveform matches are then used to constrain allowable matches during position refinement using the gradient waveform.

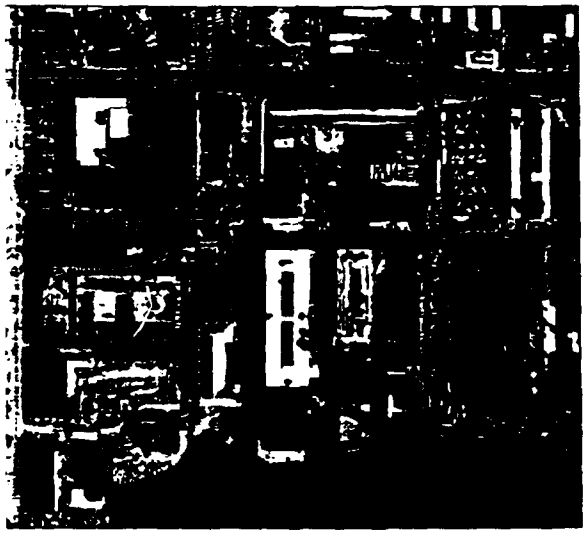

Figure 4-4: DC38008 Industrial Scene

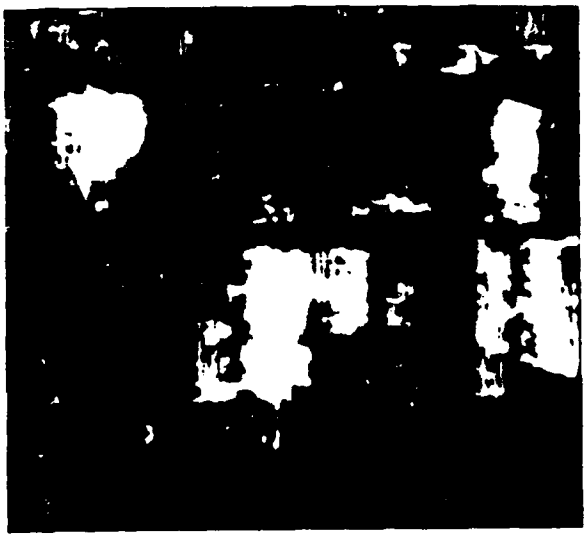

Figure 4-6: S1 Disparity Map

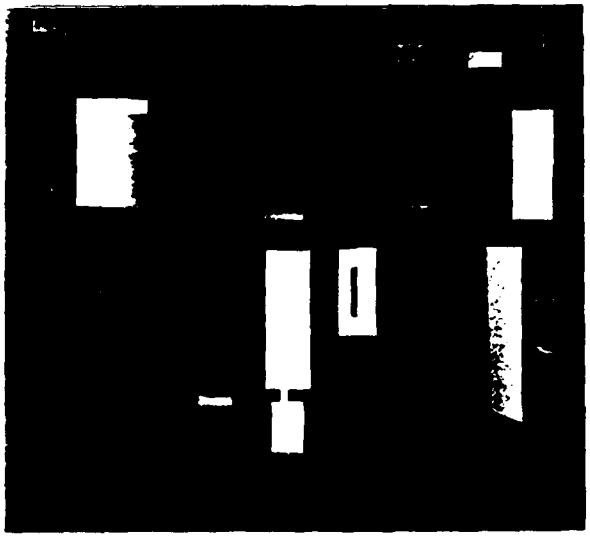

Figure 4-5: DC38008 Disparity Reference

Figure 4-7: S2 Disparity Map

S2 performs a final post processing step to explicitly deal with the problem of boundary

occlusion. We can detect an occlusion using the gradient profile when we find unmatched significant features in one profile that occur between two successive good matches where one match is a high disparity estimate and the other is a low disparity estimate. This situation is identified and corrected by allowing a two-to-one feature match. In other words, a extra feature in one profile is matched to a feature in the other profile that already has a match. At the end of this phase, we can create a sparse disparity map as shown in Figure 4-3. Points in this image represent the actual matches found by S2 and are only a small subset of the three-dimensional points in the scene. In the following section we describe the interpolation of this sparse disparity map into a dense disparity map to recover height estimates for the entire scene.

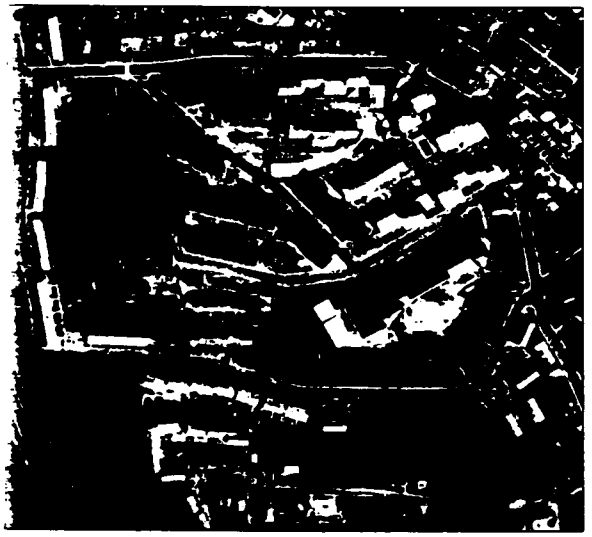

Figure 4-8: DC37405 Suburban Scene

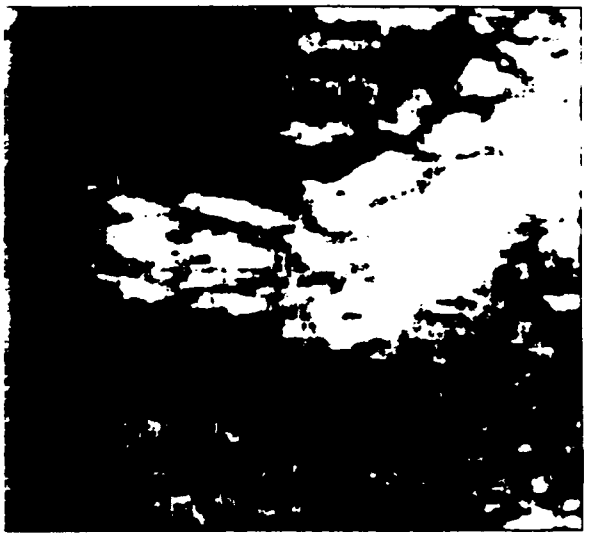

Figure 4-10: S1 Disparity Map

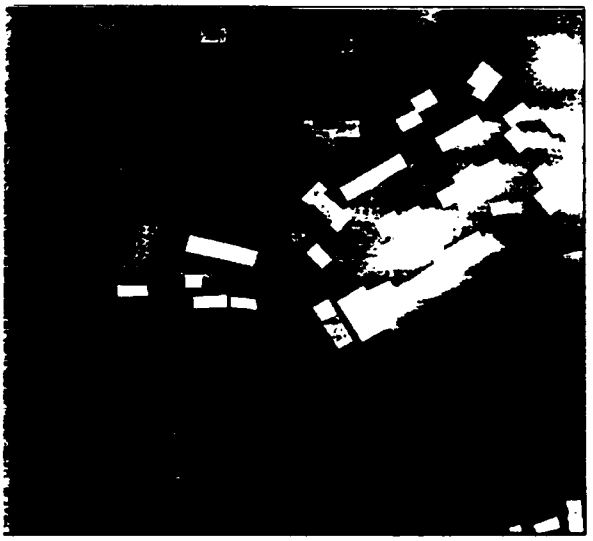

Figure 4-9: DC37405 Disparity Reference

Figure 4-11: S2 Disparity Map

One key issue in feature-based stereo matching is the interpolation process. Because we are obtaining depth estimates at sparse matching points, we must fill in depth estimates in a consistent manner in order to achieve a complete disparity estimate. There has been much work

done in surface interpolation techniques: some combined the interpolation process into normal stereo processing [13, 14], while others tried surface fitting with sparse data [15]. However, we have not found a satisfactory technique that works in both urban environments with large disparity jumps as well as in smoothly varying terrain. At present, a constant step interpolation is used because it is the most suitable method given the sharp disparity discontinuities found in urban scenes.

Figure 4-12: Denver ALV test site

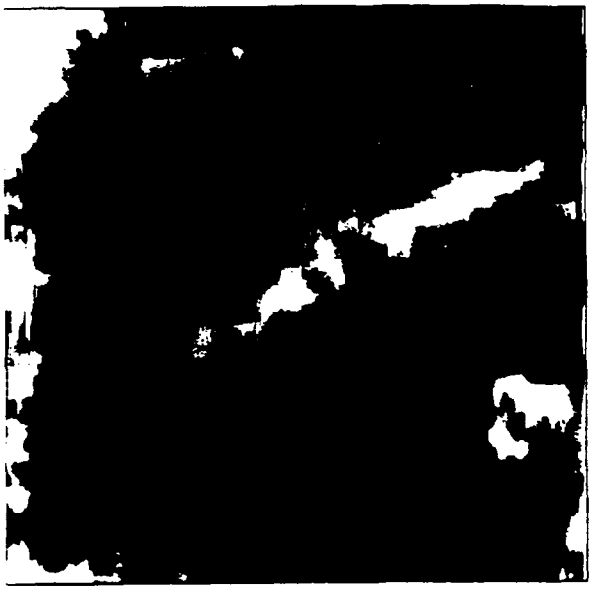

Figure 4-14: S1 disparity map for denver scene

Figure 4-13: Reference disparity map for denver scene

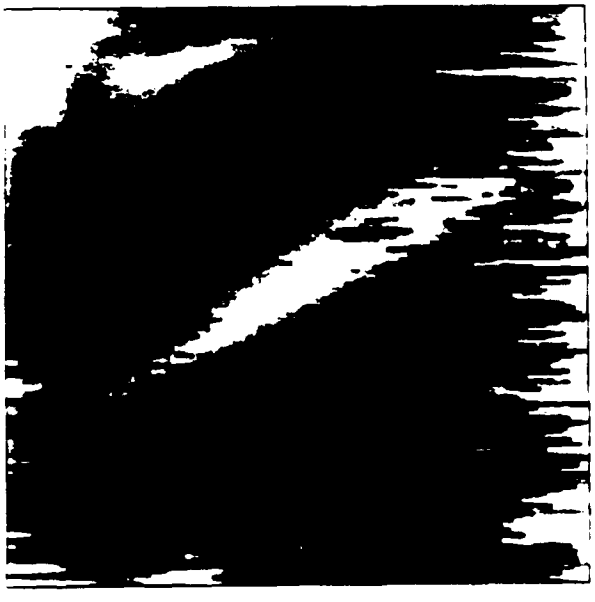

Figure 4-15: S2 disparity map for denver scene

Figure 4-3 shows the result of the S2 process in a complex industrial scene shown in Figure

4-4. White points are actual match points while black pixels correspond to points with no disparity estimate. Figure 4-7 shows the result of interpolating the sparse disparity map smoothed by a vertical median filter. Figure 4-7 shows that S2 performs well on discontinuities with most of the mismatches and errors occurring at the occlusion boundaries. In the following section we show stereo matching results for two complex urban scenes, DC38008 and DC37405, and for a scene containing rugged terrain, ALV.

4.3. Some Test Datasets

Figures 4-4 through 4-15 show current results on three test datasets. In each case we present the left intensity image of the stereo pair, a reference disparity map, and the disparity estimates calculated by the S1 and S2 matching algorithms. In all cases disparity is shown with height encoded from dark (low elevation) to light (high elevation). The ground truth disparities and the stereo disparities are scaled to the same intensity range for the purposes of visual comparison.

Figures 4-4 through 4-7 show an industrial scene containing a moderate number of complex buildings. Each building is fairly large, generally having a non-homogeneous roof texture, and have large areas of occlusion due to the building heights.

Figures 4-8 through 4-11 show a residential area having a larger variety of buildings including townhouses, apartments, and large shopping areas. It also contains rather complex terrain, where many of the townhouses are lower than the surrounding terrain.

Finally, Figures 4-12 through 4-15 show an open area with no man-made structures, the Denver ALV site. This stereo pair is included to show that the stereo matching algorithms are capable of working in highly textured areas with no depth discontinuity. The most difficult aspect of this scene was the very large disparity range, approximately 30 pixels.

4.4. Performance Evaluation

It is difficult to quantitatively evaluate the results of any stereo matching algorithm working on real, rather than synthetic, stereo image data. While random dot stereograms can provide controlled three-dimensional scene structure we do not believe they are sufficient to evaluate stereo matching algorithms in complicated imagery with natural and man-made structures. Two different evaluations are possible. We can compare a disparity result to a reference disparity map or we can compare different disparity results to one another. A true evaluation of the results, however, requires the use of a reference 'ground-truth' disparity map for comparison.

It is actually very difficult to get a good reference disparity map for an arbitrary test scene. One could imagine resorting to the use of existing digital elevation models, or paper maps with terrain contours. Unfortunately, unless one is fortunate enough to find an area with high resolution ground-truth, the accuracy of standard digital products or maps is insufficient, especially with a ground sample distance around 1 meter per pixel. We have developed a display tool to manually generate disparity maps allowing a user to select points on the registered images

and generate accurate disparity values. The user views the scene using a Tektronics 920 stereo display monitor with the imagery registered using a manual ground point selection. Once a sufficient number of points have been selected, usually a couple hundred, but depending on the complexity of the underlying terrain, we can generate a dense reference disparity map of the terrain by interpolation. Similarly, we add to the terrain disparity map, disparity regions that correspond to man-made structures. In some sense these manual disparity maps are detailed cartographic descriptions of the scene and can be much more accurate than most traditional paper-based maps. Figures 4-5, 4-9, and 4-13 show the manually produced disparity maps for the industrial, suburban house, and Denver terrain scenes.

At least three different performance measures can be calculated to evaluate a stereo disparity result. We can evaluate the general performance on a scene, the performance for all the buildings, or the performance on a building-by-building basis. The global average disparity error is computed by finding the error for each point between an estimated disparity value and the reference disparity map. This single statistic provides a quick quantitative measure of the quality of the disparity map. One can further categorize points in the reference disparity map as high gradient points, low gradient points, points with high disparity, or points with low disparity. Based upon this classification it could be interesting to evaluate the performance of various stereo matching algorithms for specific problems such as smoothing over depth discontinuities or sensitivity to disparity range.

We describe statistics on the error between the reference disparity value and the disparity result without any further classification. For our global measure we present the average error for the entire scene and the percentage of points having an estimate within +/- one pixel disparity from the reference for the entire scene. The use of +/- one pixel disparity reflects some of the accuracy limitations in the reference disparity map and is discussed further in Section 4.4.3. These simple parameters give us an idea of the magnitude of the errors in the scene, but do not give much insight into their distribution. Other error metrics such as min/max error are not very reliable since they can be caused by single point errors that may occur in either the calculated or reference disparity map.

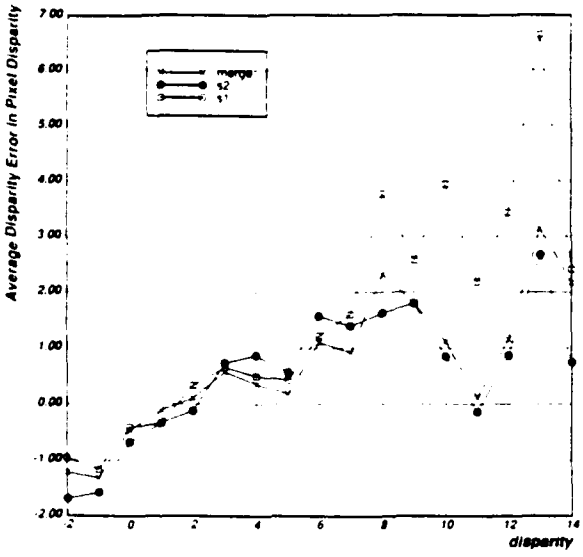

Figure 4-16: Average Error in Pixel Disparity in DC38008

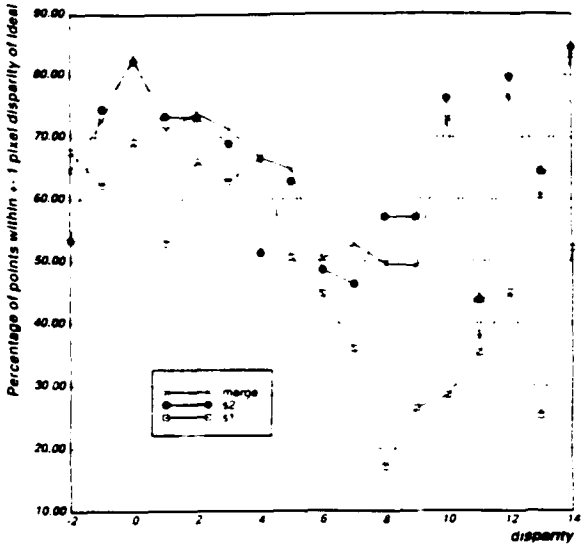

Figure 4-17: Percent points within +/- 1 Pixel of Ideal Disparity in DC38008

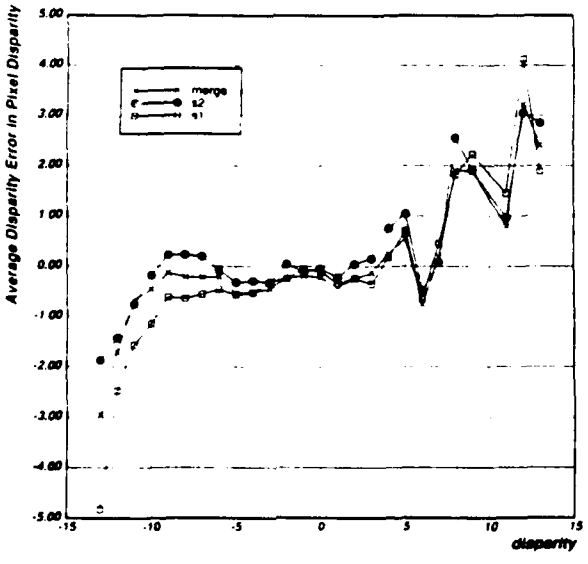

Figure 4-18: Average Error in Pixel Disparity in DC37405

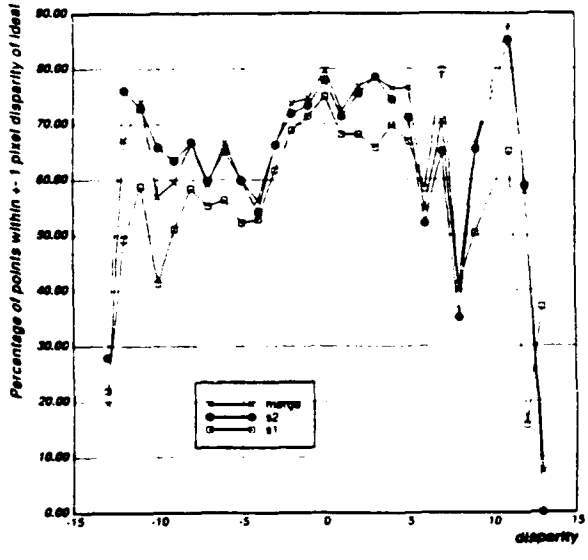

Figure 4-19: Percent points within +/- 1 Pixel of Ideal Disparity in DC37405

Tables 4-1, 4-2, and 4-3 give the global error estimates for each of the three test scenes. These global statistics show that S1, the area-based method, S2, the feature-band method and merge, the combination of S1 and S2, give very similar results across each of the three scenes. Interestingly, these measures do not seem to statistically reveal the apparent perceptual improvement achieved by merging the results of S1 and S2. We believe that this argues for a more structural analysis in addition to global scene measures.

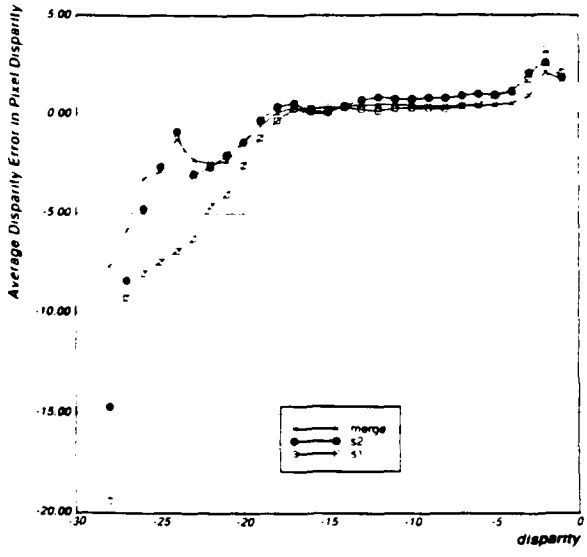

Figure 4-20: Average Error in Pixel Disparity in Denver ALV

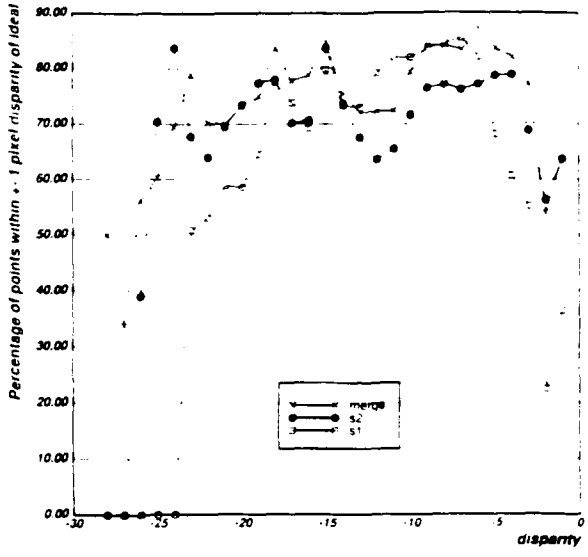

Figure 4-21: Percent points within +/- 1 Pixel of Ideal Disparity in Denver ALV

One way to address some of the issues that are hidden by global statistics is to measure the influence of the disparity value on matching accuracy for each of the methods. The graphics in Figures 4-16, 4-17, 4-18, 4-19, 4-20, and 4-21 plot error rates sorted by reference disparity. Figures 4-16, 4-18, and 4-20 show the average error in pixel disparity at each disparity level for each of the test scenes. Each contains three graphs showing the results for S1, S2, and the merged result of S1 and S2. Figures 4-17, 4-19, and 4-21 show the percentage of points within +/- one pixel of the ideal pixel disparity over each disparity range.

In general, these graphs indicate that the greater the actual disparity, the more likely the various matching algorithms will make a mistake. This is reflected in both a higher average error and a lower percentage of points within +/- one pixel of the actual disparity. These global metrics also show that in areas of low disparity, S1, S2, and their merger give similar results. For higher disparities S1 has much more of a problem in correctly estimating the disparity than does S2. Further, in most cases, the result of S1 and S2 merging produces an improved estimate causing errors to decrease.

Global Error Estimate for Stereo Matching Using Figure 4-5 as ground truth				
Stereo Method	Min/Max Disparity	Average Error % (pixel disparity)	% of points within +- 1 pixel disparity	Ground Truth Disparity Range
S1	-12/13	7%(1)	58%	-2/15
S2	-5/14	6%(1)	63%	-2/15
S1+S2	-10/14	5%(1)	59%	-2/15

Table 4-1: Statistics for different stereo matching methods on DC38008

Global Error Estimate for Stereo Matching Using Figure 4-9 as ground truth				
Stereo Method	Min/Max Disparity	Average Error % (pixel disparity)	% of points within +- 1 pixel disparity	Ground Truth Disparity Range
S1	-12/12	5%(1)	63%	-13/13
S2	-15/15	4%(1)	70%	-13/13
S1+S2	-15/15	4%(1)	70%	-13/13

Table 4-2: Statistics for different stereo matching methods on DC37405

Global Error Estimate for Stereo Matching Using Figure 4-13 as ground truth				
Stereo Method	Min/Max Disparity	Average Error % (pixel disparity)	% of points within +- 1 pixel disparity	Ground Truth Disparity Range
S1	-22/19	5%(2)	61%	-28/-1
S2	-26/1	6%(1)	70%	-28/-1
S1+S2	-25/1	6%(1)	70%	-28/-1

Table 4-3: Statistics for different stereo matching methods on Denver scene

In areas with man-made structures global accuracy statistics do not adequately convey the quality of the stereo matching system with respect to the buildings in the scene. In most cases buildings may cover only a small portion of the scene and the background terrain will statistically dominate the scene-wide estimate of disparity quality. Thus, we require a method that allows buildings to be evaluated independently or as a class of objects in the scene. Additionally, there are several metrics that can be used to evaluate both the disparity estimate and the quality of the depth jumps. We discuss these metrics in the following sections. Figures 4-22 and 4-23 are hand segmentations of the left image where we have associated a reference building IDs. Figures 4-24 and 4-25 are graphs showing the actual building heights referenced to the building IDs. We have also computed, for each building in the ground-truth, the height of

the building over its surrounding terrain. We have assigned building ID's based upon the ground-truth disparity map so that taller buildings have larger numeric ID's.

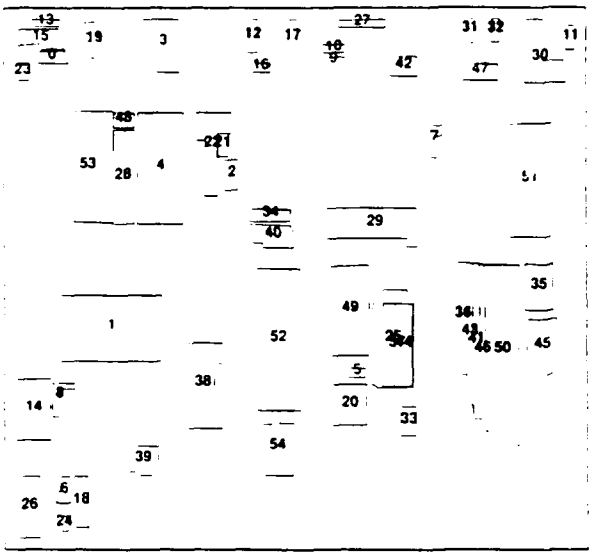

Figure 4-22: Building Index for DC38008

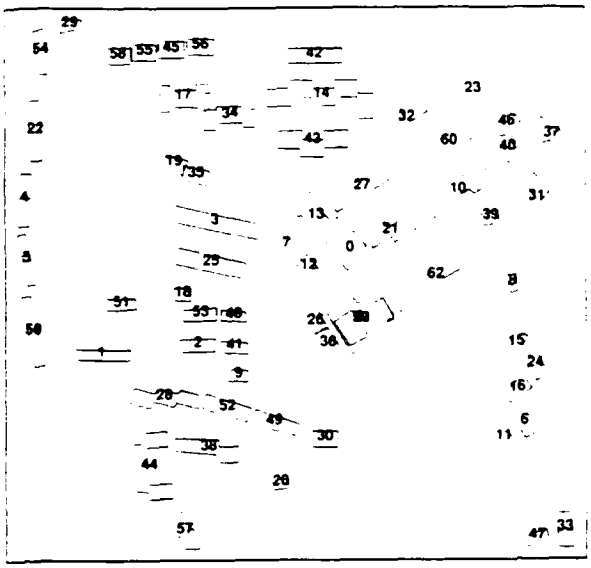

Figure 4-23: Building Index for DC37405

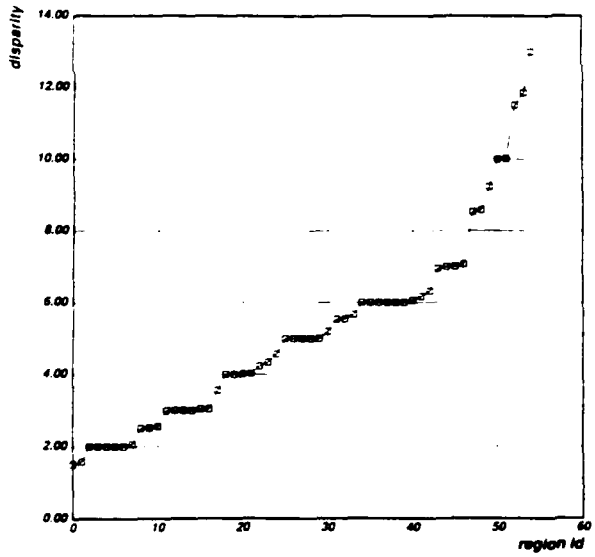

Figure 4-24: Building Heights for Figure 4-22

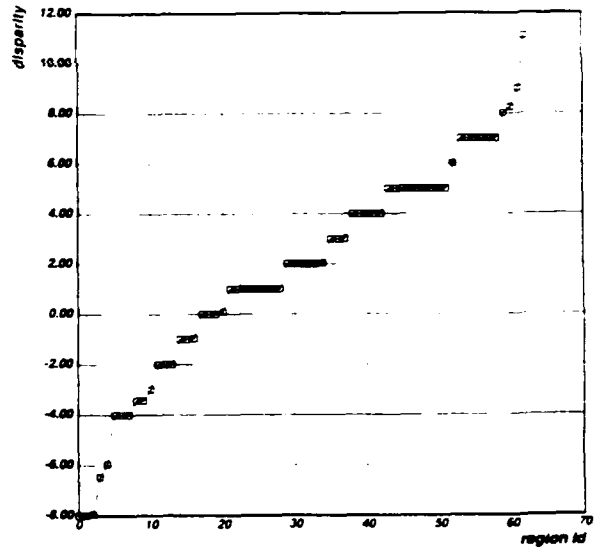

Figure 4-25: Building Heights for Figure 4-23

4.4.1. Quality of Building Disparity Estimate

In order to evaluate the performance of S1, S2 and the merged result on buildings in the scene we can gather statistics on the disparity estimate for each pixel considered to be on the roof of the building. As before, the average disparity error in pixel disparity and the percentage of points within +/- one pixel of the ground-truth estimation are good measures for performance. Figure 4-26 shows the quality of the disparity estimate for each of the buildings in the DC38008

industrial scene. The x-axis represents the ID number for each building and the y-axis shows the errors in estimated disparity for a particular building across S1, S2, and the merged result. This graphic, although a bit cluttered, shows no clear trend of performance advantage: both S1 and S2 produce a comparable result, although S2 appears to perform better, especially on buildings with greater disparity. For most buildings the error is bounded between +/- two pixels. The result of merging generally appears to improve the average error. As we have assigned building ID's sorted by disparity we can observe a trend towards increased error as we move along the x-axis.

We can also represent results using the disparity jump instead of the building ID to index the results. These graphics represent the integration of the average disparity error over all buildings with the same disparity jump. Figure 4-27 and 4-28 show the effect of disparity jump on the disparity estimate and allow us to determine whether the actual height of a building over its neighborhood (disparity jump) affects the disparity estimate produced by stereo matching. It appears that S1 is comparable with S2 for smaller buildings. This is because low buildings can satisfy the continuity constraint of the area-based method. S2 performs better on scenes with buildings having significant height because low buildings can be easily masked by random mismatches in the feature-based analysis. The merge of S1 and S2 produces results that combine the best properties of both methods.

Figures 4-29, 4-30 and 4-31 provide similar statistics for the suburban house scene, DC37405. As in DC38008 the average error for each building appears to be bounded by +/- two pixels, S2 appears to have slightly better performance than S1, and the result of the merger almost always improves the average error. Whereas S2 always appears to perform much better than S1 with respect to the percentage points (within +/-1 pixel of the correct disparity in DC38008), (Figure 4-28) this is not the case for DC37405 as shown in Figure 4-31.

These statistics allow us to pinpoint problems at a much finer grain of detail than can be accomplished with global analysis. Thus we can identify specific buildings in the scene and try to understand, at the algorithmic level, whether there are specific situations where matching could be improved. Once identified, these improvements should have an overall positive effect on the rest of the scene. The result, of course, can be subjected to the same rigorous performance analysis. Once we commit to working on complex scenes, as opposed to synthetic controlled images, the visual inspection of disparity results to discover small variations in performance becomes very unsatisfactory, except possibly at the earliest stages of experimentation. Such manual inspection greatly limits our ability to detect subtle conceptual bugs or recognize possibilities for algorithmic improvement. In some cases we can perform systematic analysis across multiple scenes. For example, in applying statistics that take into account the disparity jump for individual buildings, we can aggregate performance information for all buildings across all scenes to achieve a larger statistical sample.

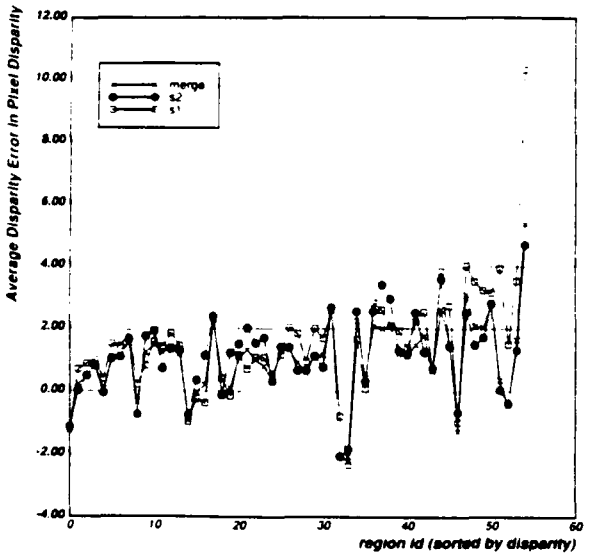

Figure 4-26: Average Error for Each Building in DC38008

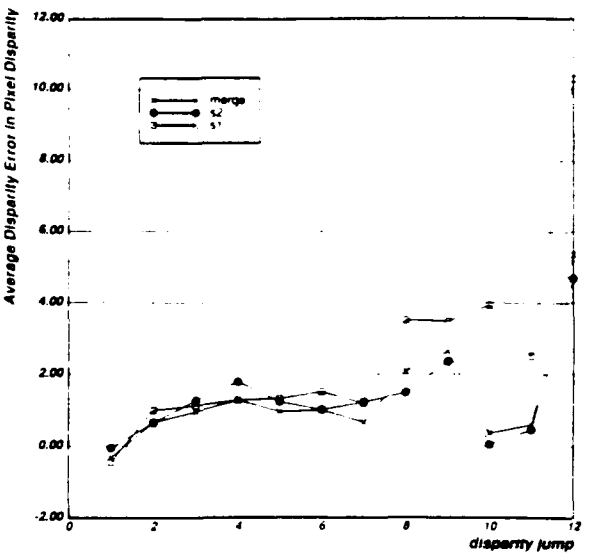

Figure 4-27: Average Error for Each Disparity Jump in DC38008

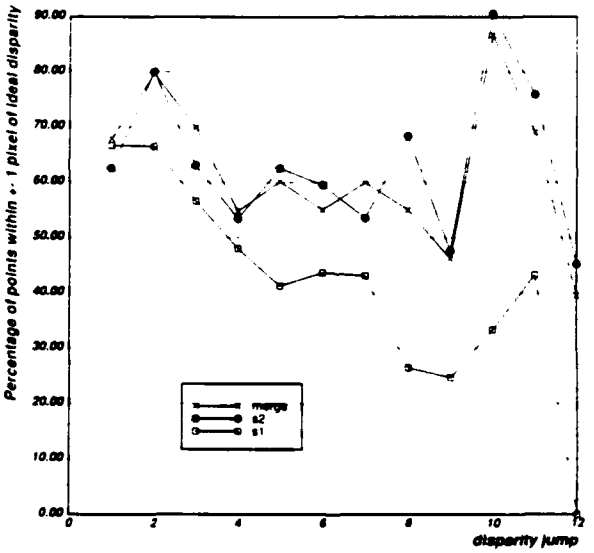

Figure 4-28: Percentage of Good Points for Each Disparity Jump in DC38008

4.4.2. Quality of Delineation Estimate

In the previous section we described techniques to measure the accuracy with which we can recover the height of buildings in the scene. For cartographic applications it is equally important that we generate an accurate delineation of the buildings with respect to their surroundings. In this section we discuss another metric which is the quality of the stereo delineation of each building in the scene. We compute *edge location* which measures the distance of the estimated disparity jump from that in the ground-truth disparity. We also measure *edge sharpness* which corresponds to the shape of the disparity jump in the estimated disparity map. Ideally, we would

expect the stereo matcher to generate a step disparity jump at the point where the actual disparity jump occurs in the reference disparity map. As before, we assume that the ground-truth disparity map accurately captures the location and the height of the building edges. In order to allow for measurement error, we tolerate some uncertainty in both the location of the edge (+/- one pixel) and the height estimate on both sides of the edge (edge sharpness). The uncertainty in edge sharpness is somewhat difficult to quantify since it depends on both the height estimate on each side of the building roof edge and on the height estimate of the neighboring ground. These estimates may be biased, since in some cases we are interpolating the ground elevation from a sparse network of points. We can alleviate this error by making sure that we select representative ground points as close to the buildings as possible.

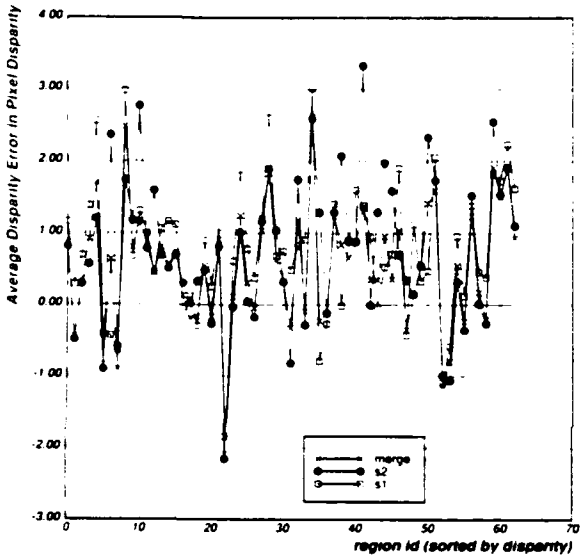

Figure 4-29: Average Error for Each Building in DC37405

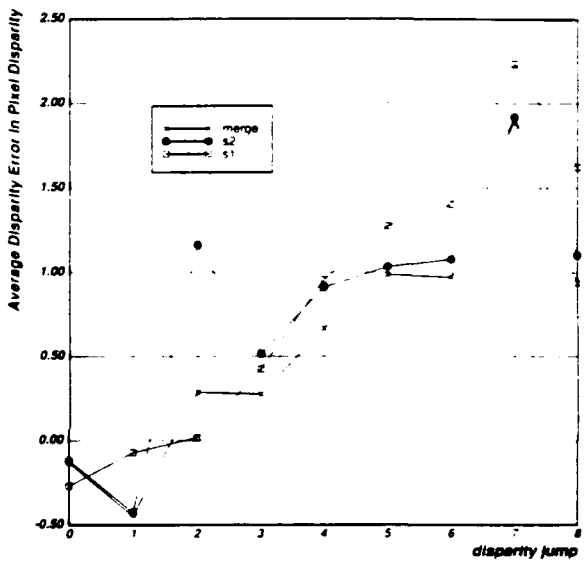

Figure 4-30: Average Error for Each Disparity Jump in DC37405

Figure 4-32 shows how we compute the edge location and sharpness for each building in the scene. The two waveforms represent the gradient of the reference disparity map and the disparity result being evaluated. The peaks in the reference disparity map gradient represent the true edge of the building in the scene. The evaluation process finds the best matching peaks in the S1, S2, or merged disparity map gradient within a neighborhood of the reference edge. The distance P corresponds to the position error of the edge in the result disparity map. The ratio Hd/Hr corresponds to the sharpness evaluation of the edge. A ratio of one is perfect. The value Hd and Hr correspond to the amplitude of the gradient related to the reference zero gradient.

Both the position error and the edge sharpness metric require that an edge point in the reference disparity map be matched with an edge point produced by the stereo matcher under evaluation. In many cases no such match is possible; that is, there is no suitable match for the reference disparity edge. In the following examples between 35% (DC37405) and 50% (DC38008) of the reference points are not matched, hence the *matchable* edges represent between 50-65% of

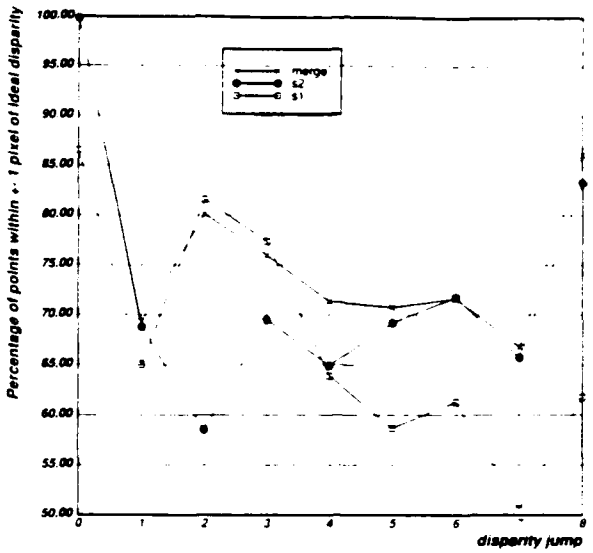

Figure 4-31: Percentage of Good Points for Each Disparity Jump in DC37405

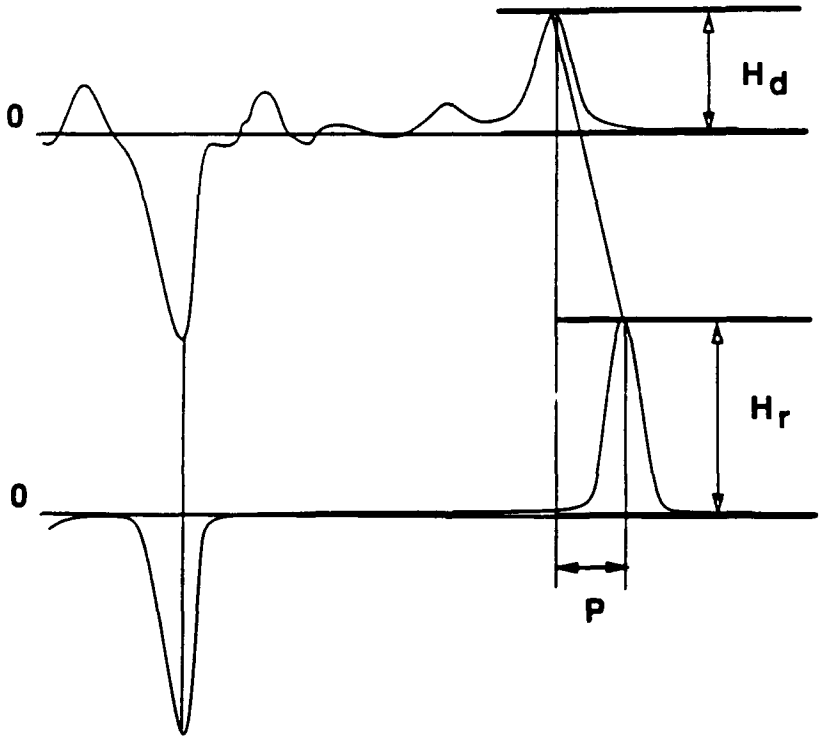

Figure 4-32: Gradient Matching for Edge Evaluation

the reference points in the scene. Figures 4-33 and 4-35 represent the average position error for the *matchable* edges across all buildings in DC38008 and DC37405, respectively.

Figures 4-34 and 4-36 shows the percentage of edges produced by the stereo matchers that are within +/- one pixel of a reference disparity map edge. These graphs are the subset of points lying in the band +/- one position error from Figures 4-33 and 4-35 respectively, plotted with

respect to all edges in the reference disparity map. In both cases the position error metric shows that the ability to accurately delineate the disparity depth jump appears to be much weaker than visual examination of the disparity maps might indicate.

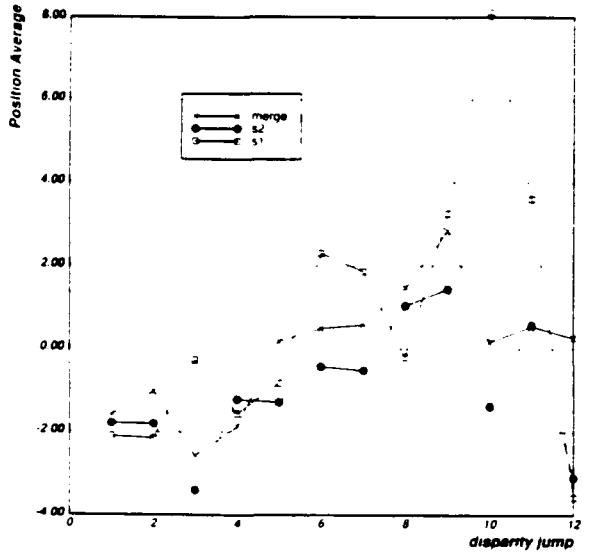

Figure 4-33: Edge Position Error for DC38008

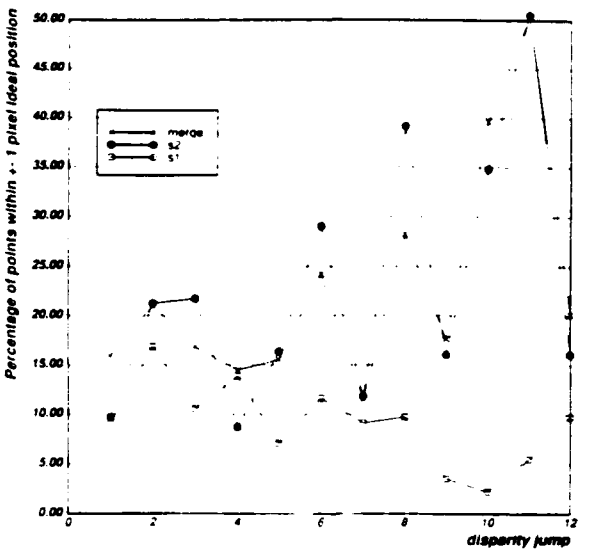

Figure 4-34: Percent Good Edgels for DC38008

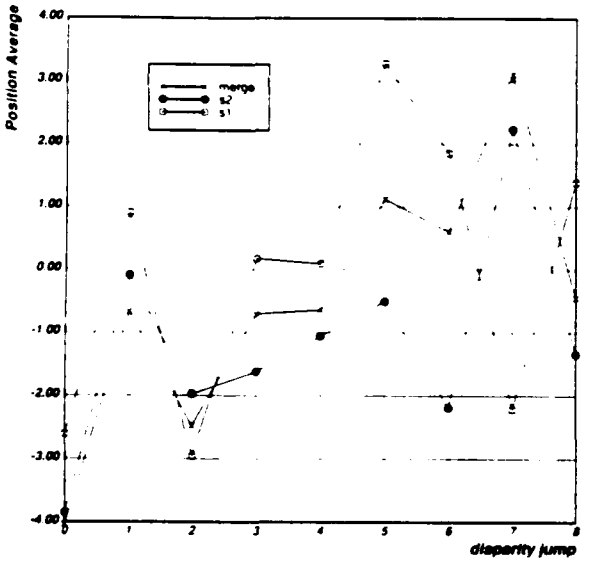

Figure 4-35: Edge Position Error for DC37405

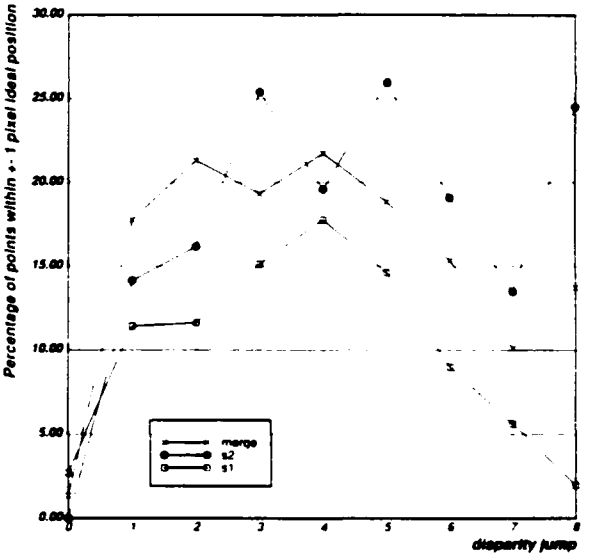

Figure 4-36: Percent Good Edgels for DC37405

For the evaluation of disparity sharpness we calculate the average edge ratio and the sharpness of edge points whose edge position is within +/- one pixel of the reference edge. Figure 4-37 represents the average edge sharpness ratio for all matchable edges across all buildings in DC38008. A ratio of one indicates a perfect step edge. Figure 4-38 shows the sharpness of edge points that are within +/- one pixel of the reference position for all buildings in DC38008. Figures 4-39 and 4-40 show the same results applied to the buildings in DC37405.

We can make several observations based upon this performance data. First, it is clear from

this analysis that S1 does n[illegible]form as well as S2 in terms of disparity delineation. Its ability to estimate the sharpness of the disparity jump (edge ratio) is likewise poorer than that of S2. However, there are some comparative advantages. S1 gives comparable results in the case of buildings with low disparity. On the DC37405 scene the S1 and S2 results are similar because the buildings in this scene do not have large disparity jumps.

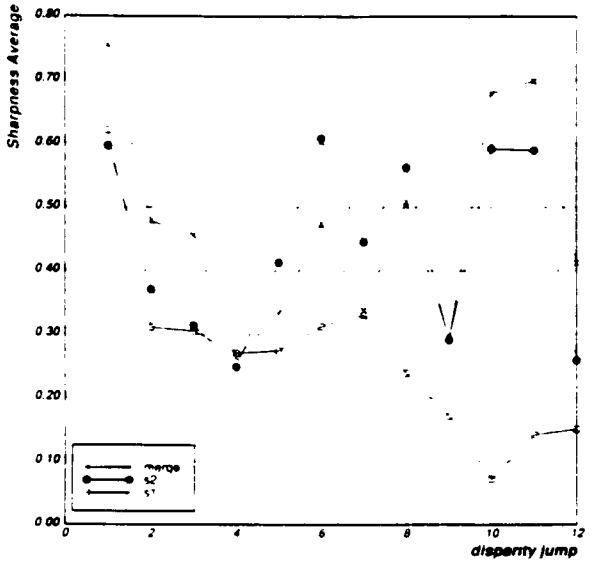

Figure 4-37: Sharpness for DC38008

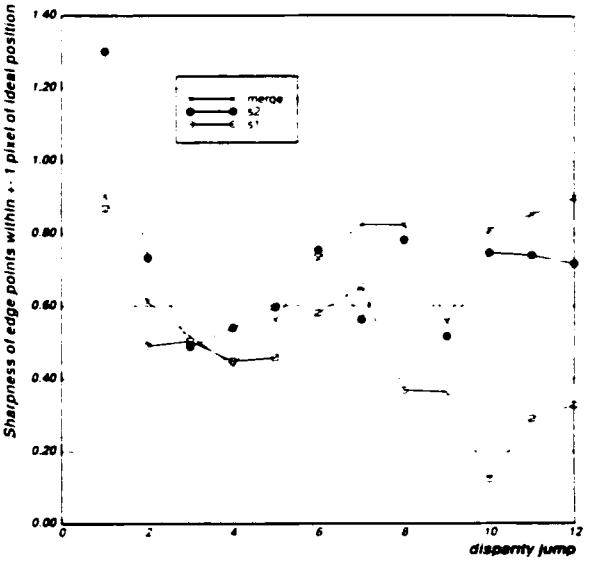

Figure 4-38: Sharpness of Good Edgels for DC38008

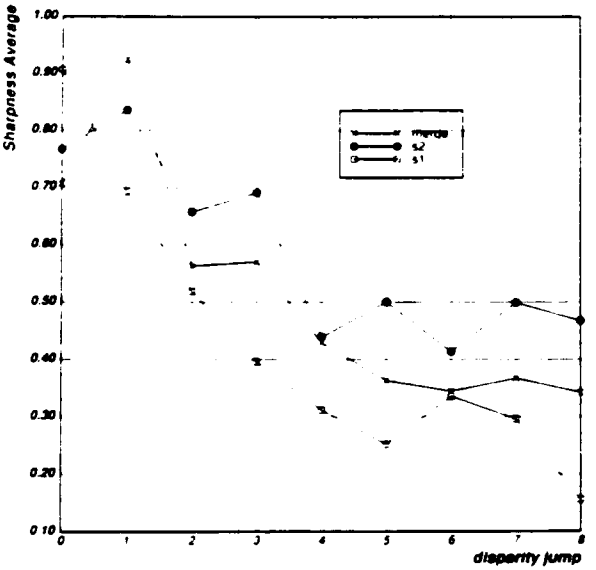

Figure 4-39: Sharpness for DC37405

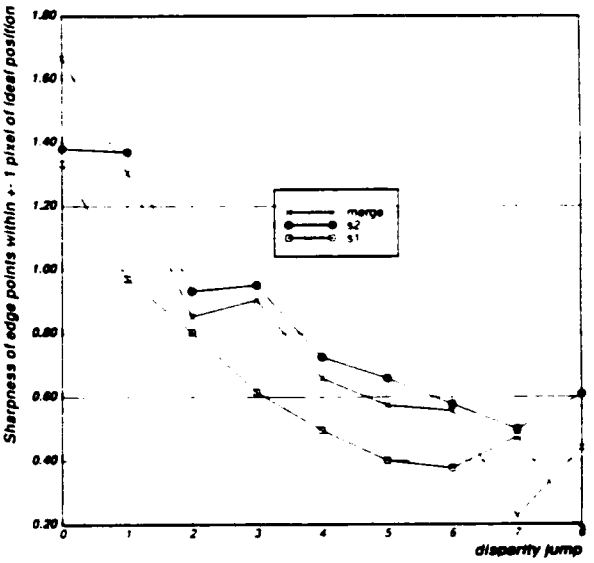

Figure 4-40: Sharpness of Good Edgels for DC37405

It is interesting to note that errors in delineation, position, and sharpness increase as the height

of the buildings increase. This is an artifact of occlusion, where higher buildings will occlude a larger area, making it more difficult to detect the exact position of the disparity jump. Edge errors seem to be comparable for both S1 and S2 for buildings with low disparity. As expected, S1 does not delineate tall buildings well and the merged result combining S1 and S2 sometimes produces a result that is an improvement over each individual method but, more often, simply decreases the maximal error.

4.4.3. Limitations of Performance Evaluation

The common theme in this section on performance evaluation is to describe a variety of quantitative measures that allow us to objectively judge how well a particular set of registration/matching techniques perform with respect to a manually compiled three-dimensional ground-truth model, and by comparison, how well they perform with respect to one another. The reference disparity map is generated using monocular and stereoscopic visualization and is a representation of the scene within a certain accuracy. In most cases the ground-truth segmentation can be constructed with enough care to provide for accurate detection of gross errors, and as a common basis for general comparison between matching methods. However, the actual accuracy of the reference disparity map has to be considered if we attempt to use it for the analysis of scene micro-structure, such as roofs with shallow pitch that are modeled as flat surfaces, small super structures such as building air conditioner units, stair well towers, and other small roof structures. These superstructures can add an error bias into the overall statistics. This bias is likely to be small; consider the fraction of error introduced in the case of a nine story building where we have not correctly modeled an air conditioner unit that rises another story over 15% of the total roof surface.

Nevertheless, we are sampling only a small subset of the actual three-dimensional points in the scene. If we count all of the building edge pixels and terrain web points manually selected for scenes such as DC38008 and DC37405, less than 3% of the scene points are used to produce the dense reference disparity map. These points are represented in a triangulated irregular network (TIN) for the terrain upon which is superimposed the building roof structures. We linearly interpolate the network in order to calculate the dense disparity map. Interestingly, S2 gives us matches for approximately 12% of the scene points which is typical for feature-based matching algorithms. As such, our performance analysis is subject to possible errors in the evaluation of S2 matching algorithm introduced due to interpolation from the sparse disparity map.

Given the lack of performance evaluation techniques in computer vision for three-dimensional scene modeling we are probably content simply to know the height of the buildings and the general shape of the underlying terrain. But we should understand that if we attempt to push performance analysis to detail the small effects of subtle algorithmic changes we may run up against fundamental limits in our ability to recover these micro-structures. Thus, in our calculations, we have added an uncertainty of +/- one pixel of disparity to the ideal ground-truth value and feel that this covers a large fraction of the inherent inaccuracies. In summary, our disparity performance evaluation has to be considered as a method to easily detect large

mismatches by the stereo analysis system: it may have some limitations in the fine evaluation of disparity values. Nevertheless, we see such techniques as the only method for effective comparison of disparity results.

5. Refinement of Disparity Estimates

One common problem for systems that interpret multiple sources of sensed data is the fusion of partial results from a variety of sources. This problem appears under many guises. For example, given a set of different scene descriptions generated from a single image using a variety of image analysis techniques, how does one intelligently combine such partial information? [16]. The introduction of additional sensor types, temporal imagery, and multiple-look imagery create dimensions along which information fusion must be performed: as such, the complexity of the problem can increase. In some cases, increased amounts of data provide improved information. This may not necessarily follow, however; complex systems having different sources of error may not reinforce correct partial interpretations nor refute incorrect ones.

In this section we describe recent research in techniques to improve the accuracy of a stereo disparity map using a segmentation of the left intensity image of a stereo pair. Thus, we are able to recover from mismatches generated during stereo matching by re-utilizing the intensity image that was originally used in the matching process. We give some experimental results on disparity refinement and describe techniques that allow for the integration of additional scene segmentations to provide for a more robust refinement process.

5.1. Disparity Refinement Procedure

In our research we utilize scene domain cues derived from monocular analysis and stereo analysis of left/right stereo image pairs. In the case of monocular analysis, one source of information is a region based segmentation of the left or right image. In the case of stereo analysis, our cues are primarily disparity maps derived from area-based and feature-based stereo matching algorithms. These image-based cues are different manifestations of man-made structures and terrain surfaces in the scene. In the case of three-dimensional reconstruction, we can make the assumption that the scene is composed of surfaces whose information content is primarily in terms of surface orientation and radiometry. Under these assumptions, we will see how estimates of three-dimensional scene structure (as encoded in a scene disparity map) can be improved by the analysis of the original monocular imagery.

We have two sources of information that can be viewed as different representations of the physical surfaces found in the scene: disparity maps resulting from different stereo matchers providing the heights of the surfaces in the scene and the initial intensity images representing the radiometric properties of the surfaces in the scene. Figures 4-4 and 4-7 show an example of "initial" data used for these data fusion experiments. Figure 4-4 is a high resolution aerial image containing a variety of buildings with complex shapes, typical of an industrial area. Figure 4-7 is a disparity map derived using a feature-based stereo matching algorithm. It is important to note that these two sources of information are "registered". That is, there is a pixel-by-pixel correspondence between points in the intensity image and points in the disparity map. In some many cases one issue complicating the use of multi-source information is the accurate registration or correspondence between the information sources themselves.

An intensity image, subject to sampling and digitization errors, poses difficulties for monocular analysis techniques such as segmentation. On the other hand, most stereo matching algorithms are fooled by different variations in the stereo pairs, which cause mismatches in the disparity maps. The mismatches in disparity maps primarily result from geometric and radiometric differences in the left and right images, rather than local digitization or sampling errors in the intensity images. Thus, it is possible to use information from the intensity images to reduce the number of mismatches introduced by stereo matching processes.

5.1.1. Region based interpretation

Our approach utilizes surface illumination information, provided by the segmentation of the monocular images into fine surface patches of nearly homogeneous intensity, to remove mismatches generated during stereo matching. First, we segment the intensity image into uniform intensity regions. These regions correspond to approximately planar surfaces in the image. We assume that the orientation and surface material are the primary factors for the radiometry of the image. Under these assumptions, uniform image radiometry is produced by a planar surface, of a certain orientation and material, in the scene.

These surfaces should have continuous linear disparity values (i.e., the disparity values of these regions are represented by continuous linear functions). Since the disparity map contains some noise, however, most of the regions segmented in the intensity image have disparity functions that are neither linear nor continuous. Ideally, we would like to approximate the actual disparity functions over the uniform intensity regions by the appropriate linear functions.

The problem of approximating a surface in three-dimensional space to a reasonable planar surface is a difficult one; we approximate such surfaces by horizontal surfaces. Then, the disparity values for each region will be the same for each pixel, and the problem is reduced to the selection of the best value for the heights of these surfaces. The general problem is that of locating of the surface which satisfies the equation

$$ax+by+cz+d=0$$

Given (x,y), we should be able to obtain

$$z = (-ax-by-d)/c$$

We assume here that $z' = -d'/c'$ only. Then the problem is to find $(-d'/c')$ that best fits the surface so that

$$ax+by+c*(-d'/c')+d\sim=0$$

or to find z' so that $z-z'$ would have a minimal value over the region (this can be the weighted mean of the z distribution or the most 'representative' value of the z distribution). In other words, we need only select a single disparity value for each region. Since we are using an over-segmentation of the image, a piecewise planar disparity map gives a good approximation of the relief in the scene. Furthermore, since we are interested in building extraction in aerial images, this approximation will be adequate.

This region-based interpretation has been developed for two different applications. We show

how this approach can support information fusion from different segmentations and well as across multiple disparity estimates based upon a local decision making evaluation. In Section 5.2.1 we describe how improved disparity maps may be obtained by correcting the mismatches produced by stereo matchers and by refining the disparity discontinuities. In Section 5.2.6 we present preliminary results in the extraction of building regions from the scene using the height information in these disparity maps.

5.1.2. Intensity Segmentation Techniques

The general scene segmentation problem is, of course, a very difficult one and has a long history in image processing and computer vision. There are no universal segmentation techniques that work well across a variety of imagery and tasks. Such low level algorithms typically differ in their approaches: they may utilize intensity-based, area-based, or edge-based techniques. Some systems combine these techniques into hybrid algorithms. We have concentrated on those segmentation methods that produce (nearly) uniform intensity regions because we wish to detect those image regions that correspond to oriented surface patches in the scene. We utilize a region segmentation algorithm based upon the histogram splitting paradigm [17] and a region growing algorithm [18] which takes into account edge strength and shape criteria [19]. Interestingly, while neither of these methods give completely satisfactory segmentation results, they provide good over-segmentations that rarely merge object/background boundaries. Both techniques will also provide different segmentations based upon modification of a small set of parameters. In our experiments we generated three scene segmentations: two by using different parameters for histogram selection, and one by using region growing. These segmentations provided the basis for our work in intensity/disparity fusion, the goal of which was to produce an improved three-dimensional scene interpretation.

Figures 5-2 - 5-4 show examples of these segmentations on the DC38008 industrial left intensity image. We ran the experiments on smoothed images (Figure 5-1) to remove intensity noise.

5.1.3. Machineseg

One of the major difficulties with region growing techniques in complex scenes is the difficulty in determining automatic stopping conditions for the merging procedure. MACHINESEG [19] is a region growing system that tries to preserve edges between regions and stops the growing procedure when certain shape or spectral criteria are not satisfied inside the region. It adds a decision procedure to evaluate the effect of the next merge operation and either allows the merge to proceed or to be rejected. In the case of disparity map refinement, we want the regions to be sufficiently uniform that they could be treated as planar (or at least "soft") surfaces. We also limited the size of the generated regions so that very small regions could not be generated, as these could be considered noise or non-representative regions. As can be seen in Figure 5-2, since we are not considering the small region, our segmentation is not a complete partition of the image; it does, however, obtain most of the representative surfaces in the image.

Figure 5-1: Nagao filtered left image for DC38008

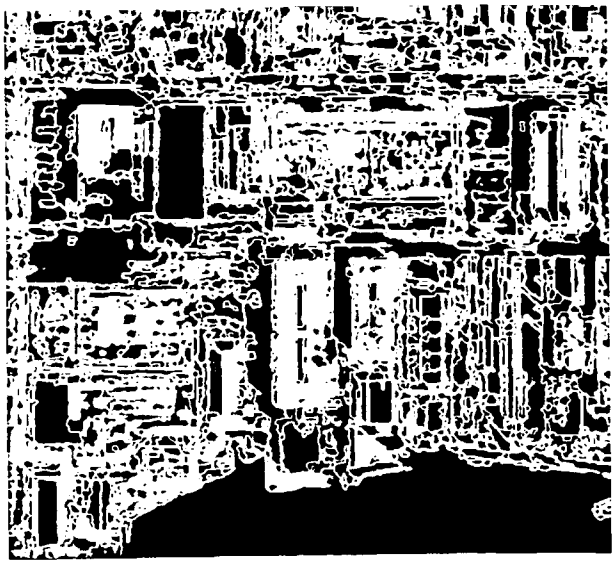

Figure 5-2: MACHINESEG segmentation on DC38008

5.1.4. Colorseg

This histogram splitting technique is based on the extraction of regions with limited intensity ranges (in other words, region of approximately uniform intensities). The technique searches for the peaks in the histogram of the image and segments the regions whose intensity values fall in windows around these peaks. The regions are then removed from the image and the process continues until all the pixels in the image have been removed. This process results in a segmentation composed of connected regions, each having an intensity range less than a certain threshold. This technique does not guarantee preservation of the edges (in particular, small edges) but it may ignore local noise with strong edges that other techniques will classify as regions. As in the previous technique, we removed very small regions (less than 20 pixels) that could be considered as noise, for further processing.

In our experiments, we generated different segmentations with different segmentation techniques. For instance, using the colorseg technique we generated two segmentations of the images, one with "uniformity" defined as a maximum of 10 intensity levels inside the region (to tolerate sensor noise and allow for imperfect planar surfaces) and another with "uniformity" defined as a maximum of 20 intensity levels (to tolerate more noise). An estimation of the noise or the average intensity range for the surfaces in the image is a delicate problem, and the use of different segmentations to estimate the intensity range inside the regions does not necessarily increase the reliability of the process. It is thus important that we obtain different segmentations of the scene that are *not consistent*, such as those in Figures 5-3 and 5-4. The fusion of these data may overcome some of the inherent problems of a single segmentation since they provide different local evaluation contexts for disparity estimates in the scene. In the following sections we show how we can merge information using different intensity segmentations.

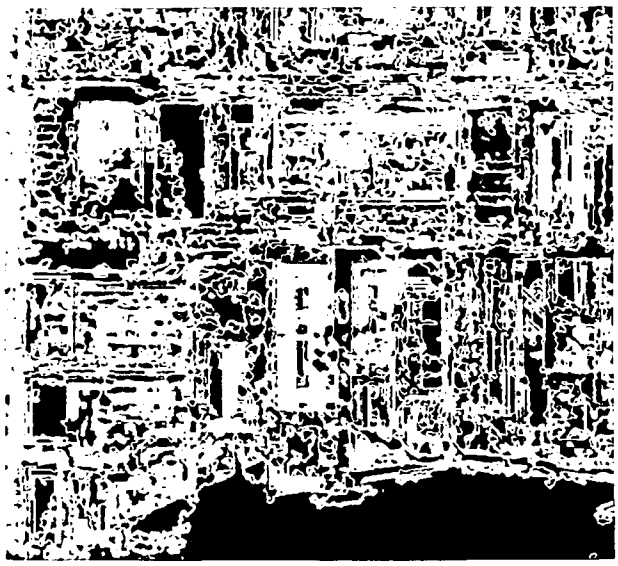

Figure 5-3: COLORSEG segmentation with 10 intensity levels sensitivity for DC38008

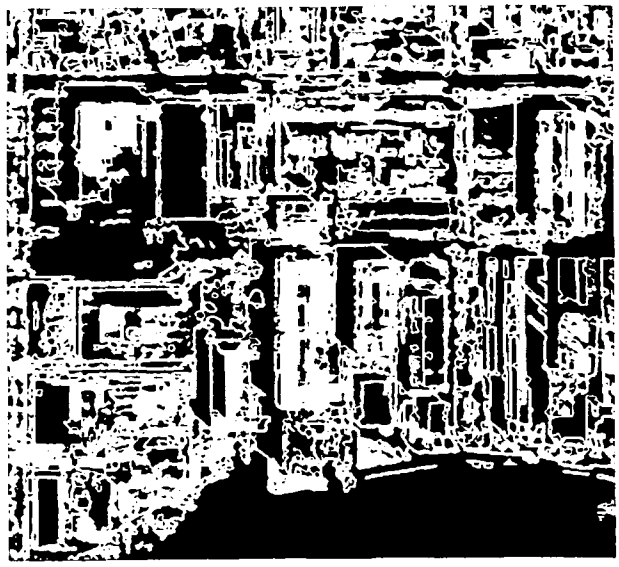

Figure 5-4: COLORSEG segmentation with 20 intensity levels sensitivity for DC38008

5.1.5. Disparity map results

Our initial height information for the industrial scene was derived using two different stereo matching algorithms. Given these sets of height information, which may or may not be reliable or unique, it becomes necessary to use a data fusion process in order to maximize the amount of useful information gained from these sets of height estimates.

We used 2 different matching techniques, one area_based (S1) and the other feature_based (S2). S1 uses the method of differences technique on neighborhoods of the image in hierarchical fashion [10, 11]. S2 performs a hierarchical matching of epipolar intensity scanlines in the left and right image [20]. The results of these stereo matching algorithms are different: S1 gives us a dense disparity map (i.e., a map containing a disparity value for each pixel in the image), while S2 gives us a sparse disparity map (i.e., a map containing a disparity value for those pixels corresponding to peaks or valleys in the intensity images).

Since we used uniform segmented regions that we assumed to be horizontal planes, a logical interpolation method for the sparse S2 disparity map is step interpolation. This produces a dense disparity map consisting of regions with uniform disparity values, which may be more easily integrated with a dense map produced by S1. Our fusion mechanism will have to correct mismatches in the S1 or S2 disparity maps and then choose the better unique disparity value for each pixel in the scene. It will have to merge very different disparity information, such as that shown in Figures 5-6 and 5-5, the two left disparity maps for the DC38008 scene.

5.2. Fusion Experiments

After different intensity segmentations and different disparity results were obtained, we applied a very simple fusion technique and developed a few experiments for the two applications under consideration. Most of the experiments have been performed for the disparity refinement process, but the results have been used for the building extraction process as well.

Figure 5-5: S1 left disparity result for DC38008

Figure 5-6: S2 left disparity result for DC38008

5.2.1. Disparity refinement

In order to refine the disparity maps (i.e., to remove mismatches, improve disparity discontinuities and obtain the best height estimate for each point in the scene), several approaches have been explored:

- Disparity refinement using one segmentation
- Disparity refinement using several segmentations
- Disparity refinement using one segmentation and several disparity maps
- Disparity refinement using several segmentations and several disparity maps

5.2.2. Simple disparity refinement

In this first approach, a histogram is constructed for each segmentation region. The values of each histogram are the disparity values in each region. The most representative value of each histogram is then selected. In our case, this value was simply that of the highest peak in the histogram. We chose this value for two reasons. The step-interpolated S2 disparity maps result in disparity histograms having only a few values, which correspond to real height values or matching noise. If the matching is reasonably robust, the noise will introduce local maxima in the histogram that will be smaller in magnitude than the best height estimate. Further, a typical region histogram for an S2 disparity map exhibits one or two large peaks and a few noise peaks that influence the average value of the histogram, making it less reliable as a representative value.

For non-horizontal regions and S1 results, the average disparity may suffice for a reasonable measure of the height of the region. A confidence score can be generated for these disparity values based on the characteristics of the histograms (and, conceivably, on the type of disparity map used as well as the nature of the region histograms). Finally, this disparity value is assigned to the entire region, under the assumption that it will be a better estimate of the height for the whole region. In most cases, this removes a large number of the mismatches, but whenever our initial assumptions about scene radiometry are not valid, our height estimates may differ from the correct height value.

We implemented this approach for each segmentation and disparity map and generated new disparity maps that were based on the initial intensity regions and disparity values. The pixels that were not considered during the segmentation were removed from these new disparity maps. Figures 5-7 and 5-8 show the results of the disparity improvement process for the different segmentations using the S2 disparity map, and Figures 5-9 and 5-10 show the results of the disparity improvement process for the S1 disparity map.

Figure 5-7: S2 left disparity result for DC38008 improved using SEG10

Figure 5-8: S2 left disparity result for DC38008 improved using SEG20

It is worth noting that a common methodology is utilized among all of the approaches described in this section. A set of attributes is computed for each region in each segmentation. Among these attributes are the statistics for the disparity values inside a region, the best disparity value, and a confidence score for this value. This allows the computation to proceed at a symbolic level on a region-by-region basis.

Figure 5-9: S1 left disparity result for Dc'sous improved using SEG1

Figure 5-10: S2 left disparity result for Dc'sous improved using SEG2

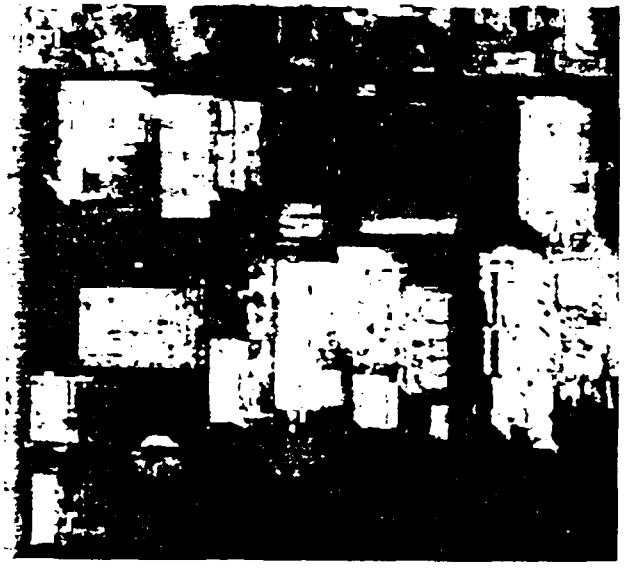

Figure 5-11: S1 left disparity result for Dc'sous improved using the merging of SEG1 and SEG2

Figure 5-12: S2 left disparity result for Dc'sous improved using the merging of SEG1 and SEG2

5.2.3. Multi-segmentation disparity refinement

In the second approach, we can merge different height estimates, given different intensity segmentations SEG1 ... SEGn and then merging the results across the different segmentations. We refine the disparity estimate for each pixel by locating the intensity region to which it belongs, for each of the image segmentations. This list of regions can then be searched to obtain

the disparity estimate attribute (computed for a given disparity map) as well as a confidence score for this estimate. The confidence score is then used to select the best disparity value, which is then assigned to the pixel. Currently a simple decision is made to select the disparity value having the highest confidence score.

An attempt is made to maximize the score for each pixel in the entire image. This is done by selecting a disparity value in all of the regions resulting from the union of the segmentations. In other words, the segmentations were merged and the best height value was selected for each of these regions, by utilizing the confidence scores computed for each region. The scoring method currently in use takes into account information about the nature of the segmentation used.

In particular, higher confidences can be assigned to sufficiently large regions in a constrained segmentation such as SEG10 than to the equivalent regions in SEG20. Information of this nature must be incorporated in the confidence function for each segmentation region.

Figures 5-12 and 5-11 show the results of merging the SEG10 and the SEG20 segmentations for the S2 and the S1 disparity maps, respectively. Depending on the confidence scores of the disparity values selected for each segmentation, we were able to obtain improved disparity estimates for some of the regions. Comparing these results to Figures 5-7 and 5-8, disparity maps obtained with the simple method, we observe some of the failings of both approaches. The initial segmentations, in some cases, are under-segmented instead of over-segmented, resulting in the grouping of regions that should have been assigned different height estimates. Another factor is the confidence evaluation function for the regions of the segmentation, which only takes simple properties of the disparity histograms of each region into account.

5.2.4. Multi-Disparity Disparity Refinement

In this approach, several different disparity maps are merged using a single segmentation, looking for consistent areas across disparity maps. This approach is similar to the simple disparity improvement approach, except that we now attempt to select the best disparity value based on a set of differing confidence scores. The score established for each disparity map at each pixel should be dependent on the stereo matching algorithm used to generate the map, and should also take into account the nature of the possible mismatches resulting from each stereo matching technique.

The major problem with all of the refinement approaches discussed in this final report is the development of a reasonable confidence evaluation function for each set of data. Currently, confidence is evaluated by a scoring function that utilizes the standard deviation and the disparity range of the histogram for each region, as well as the size of the region. Ideally, this scoring function would also take into account the nature of the disparity map. As an initial experiment, we defined a similar scoring function for each disparity map and checked for disparity consistency across segmentation regions. In Figure 5-13, the areas where disparity values differ between S1 and S2 are marked in black, as we do not use any score difference information to select the most probable height value at this stage.

5.2.5. General Disparity Refinement

For the general case we can merge the results of different disparity maps and different segmentations and look for consistency across the results. The approach is similar to the multi-segmentation method; however, we should be able to add additional height hypotheses according to the different segmentations.

Again, the processes can be decomposed into two stages. The first stage will gather the information and convert it into a common representation (i.e., region attributes). As an example, for each segmentation we should obtain a list of height estimates with scores associated with each of the different disparity maps we can use (S1 and S2). The second stage will attempt to merge this information by selecting the "correct" value from the available information, by comparing scores based on the nature and quality of the different pieces of information. If we can precisely evaluate the quality or confidence in the information, we should be able to maximize the amount of accurate data we merge from our different information sources.

There are still many experiments that have yet to be performed. In particular, experimentation needs to be done on merging the two different disparity values for the three different segmentations.

Figure 5-13: S1 left disparity
and S2 left disparity
merged using YAK

5.2.6. Building extraction

This second application of information fusion is an attempt to validate this region-based approach for scene interpretation. Using the previously described methods, we can obtain an estimate of the height of each of the composite regions in each segmentation. According to our representation of the scene, buildings are composed of a single intensity region or a group of intensity regions, and, in general, are higher than their surroundings. Therefore, regions

representing parts of a building should be higher than their neighboring regions.

For each region, a list of its neighboring regions is constructed, and the disparity values for each of these regions are obtained. Then, a weighted histogram is computed that takes into account shared boundary length and disparity information. This weighted score is then compared with the height of the region to label the region as building structure or background terrain. This building extraction process can use either the initial disparity map or the refined disparity map.

A refinement process is used to group neighboring regions with the same height in order to obtain an intermediate segmentation containing fewer (and larger) consistent regions. This grouping procedure merges connected regions having the same height to form a single region. This allows the building extraction process to use larger, and hopefully more consistent, disparity regions as a basis for the neighborhood disparity analysis. The quality of this analysis is again dependent on the accuracy of the disparity estimate, as in the previous fusion process. Figure 5-14 shows the result of such an analysis. The white regions correspond to sections of buildings. The building extraction, as done by hand, is in Figure 5-15. The building extraction process described here illustrates one facet of scene interpretation that can be performed within our refinement framework.

Figure 5-14: Building regions for DC38008 extracted using the merging of SEG10 and SEG20

Figure 5-15: Building regions for DC38008 extracted manually

6. Database Support for Spatial Databases

Automated cartographic feature extraction requires database support to store and retrieve existing map knowledge, and to incrementally update spatial databases as new information becomes available. We have begun some preliminary work in addressing such issues using realistic digital map databases. CONCEPTMAP, our spatial database system, provides a framework for storage and retrieval of many types of image data, including images and associated attributes, object boundaries (in image or map coordinates), correspondence information, et cetera. The ability to freely import and export this type of information is another important feature that CONCEPTMAP provides. To this end, we have initiated work to integrate ITD cartographic data, DMA's Interim Terrain Data format, into our spatial database management system, as well as to provide an efficient graphical user interface to the database.

The integration process includes the decoding of low level ITD symbolic and spatial data, the building of a representation for the ITD data structure specifications, and the automated generation of formated information that can be directly integrated in a known spatial database management system. The user interface supports ways of querying the data dictionaries and the data structure of the database. It also allows to display and interactively select spatial information. Finally, it provides simple means of performing semantic queries on multi-layer data.

6.1. Integration

A general decoding program has been written to read the raw ITD data format and produce human readable ascii information. This program uses a simple assembly-like language where each instruction is able to read an arbitrary binary data stream and produce an (Attribute,Value) pair text output. The processor is able to interpret numeric data on the fly to parameterize loops and decode dynamic data structures.

A set of utilities has been developed to extract the spatial information from the decoded data and produce specific data file formats for display purposes or for future database integration. Part of this process makes use of global information for absolute positioning, or segment naming conventions to provide links to the symbolic data. This is supported by a simple mechanism where the global information on a given data set is centralized in configuration files.

In order to check symbolic data consistency and produce ITD symbolic data in a format suited for the integration in a known database system, or more simply in user readable format, a data representation of the complete ITD format specifications has been built. The ITD data structure is organized in three levels: the feature level, the attribute level, and the physical structure level. These levels are described in text files so that they can easily be modified with a text editor.

The data output formats of the various steps towards integration are consistent so standard tools could be developed. All the data involved after the first stage of decoding is represented as (attribute,value) pairs stored in text format (ascii files). This format is very useful for

experimentation. It makes it easy to read and interpret the output of a process as well as edit the input of a process for testing. The text format is also justified by the fact that most of the processing involves symbolic data.

Some programs were developed to transform the ITD symbolic data into a format that can be used by a database management system. From the first two levels of the structure specification, a feature dictionary is generated. This dictionary, the feature and attribute levels, the physical structure and the global configuration files specific to the data set can then be used by an integration program (or *database compiler*) to automatically generate an arbitrary database format.

For experimental purposes, our first database compiler outputs an (attribute,value) pair text file format. The tools designed for text file parsing have been extended to support feature oriented structures (processing of paragraphs as separate features). The second part of this report shows how a simple database management system was developed with these tools to experiment on user interface issues.

The next step is to write an ITD database compiler for the CONCEPTMAP spatial database. Already, the spatial information can be integrated in CONCEPTMAP, and work is in progress to automatically generate CONCEPTMAP symbolic data dictionaries.

6.2. Interface

Once the ITD cartographic data has been organized and compiled into a usable database, the next issue is to design a user interface to this database. In the process of designing a high level graphical user interface, three levels of interfacing have been developed. The first level involves utility programs executed from the UNIX command line. This level is useful to define and develop the various functionalities needed, but is the most primitive interface. The second level is the integration of a set of functions in a common context (it is possible to have a current working data set, and help facilities). This level is still command line oriented, but is the best that can be implemented on a regular terminal. The last level is the graphical user interface implemented on a Unix workstation with the X Window System. The goal is to simplify the task of the user by guiding his steps through a session in order to optimize the use of the system by reducing the interaction time, in particular for complex tasks.

6.2.1. The database system

The first utilities that we developed provided a way to consult data and link the spatial ITD data to the symbolic data. One of the utilities allows the user to pick a segment on a graphic display and get in return the symbolic information for the feature corresponding to that segment. Another one can filter symbolic data files to extract a class of features.

An initial version of the database query system consisted of a command line interface developed using CI, the *Command Interpreter*, which integrated the various functions provided

by the consultation utilities. In this first system, the user could consult the ITD data structure specifications, display the ITD spatial data, and consult the compiled version of the ITD symbolic data. Then, a simple query system was added to search through the ITD symbolic data. This query system is based on *data filters*. The concept of a data filter in this system is analogous to UNIX *filters*, but is implemented slightly differently (no fixed length buffers) and adapted to this particular data format (more rigid and simple syntax, feature oriented format).

In this system, a feature presented at the input of a filter will generate on its output a new feature (generally itself or NULL). Therefore, when this filter is applied to a feature file it will generate a new feature file. In a database context, such a filter can be considered to be a query primitive and the filtered feature file corresponds to the result of this query. In this model, filters can be chained together to achieve the 'AND' operation or put in parallel to realize the 'OR' operation.

During a database query session, a complete trace of the different outputs obtained is kept by the system. A history is associated with each partial result so that it is possible to reconstruct the processing steps that lead to the result. It is possible to save the history along with partial results.

6.2.2. The graphical user interface

A first cut at the graphical interface for representing the feature files and filters was developed using the X Window System. The current interface is simple, though it does demonstrate the methods for graphically combining filters. It also shows how it is possible to back-propagate a query until a database match is found. This information is loaded as a text file and the various filters are applied in the right order until the query is satisfied. There are several benefits gained by using the simple representations. Because of the use of the XWindows and since our database representation is text based, the query system is easily ported from one platform to another. We currently have the system running on VAX and Sun platforms.

The current graphical interface is an adaptation of the command line interface, employing the XCI package. XCI is an adaptation of the CI command interpreter interface in the X Window environment. It replaces the command line by menus and dialog boxes for user input. Figure 6-1 shows an example of the user's interaction with the interface. For the output, separate windows are used to display symbolic and spatial data. A custom graphics package, again based on X, is used to display the spatial data. The interface packages are used only as building blocks for the database query system, so other projects within our group will be able to take advantage of the work done to produce these interfaces.

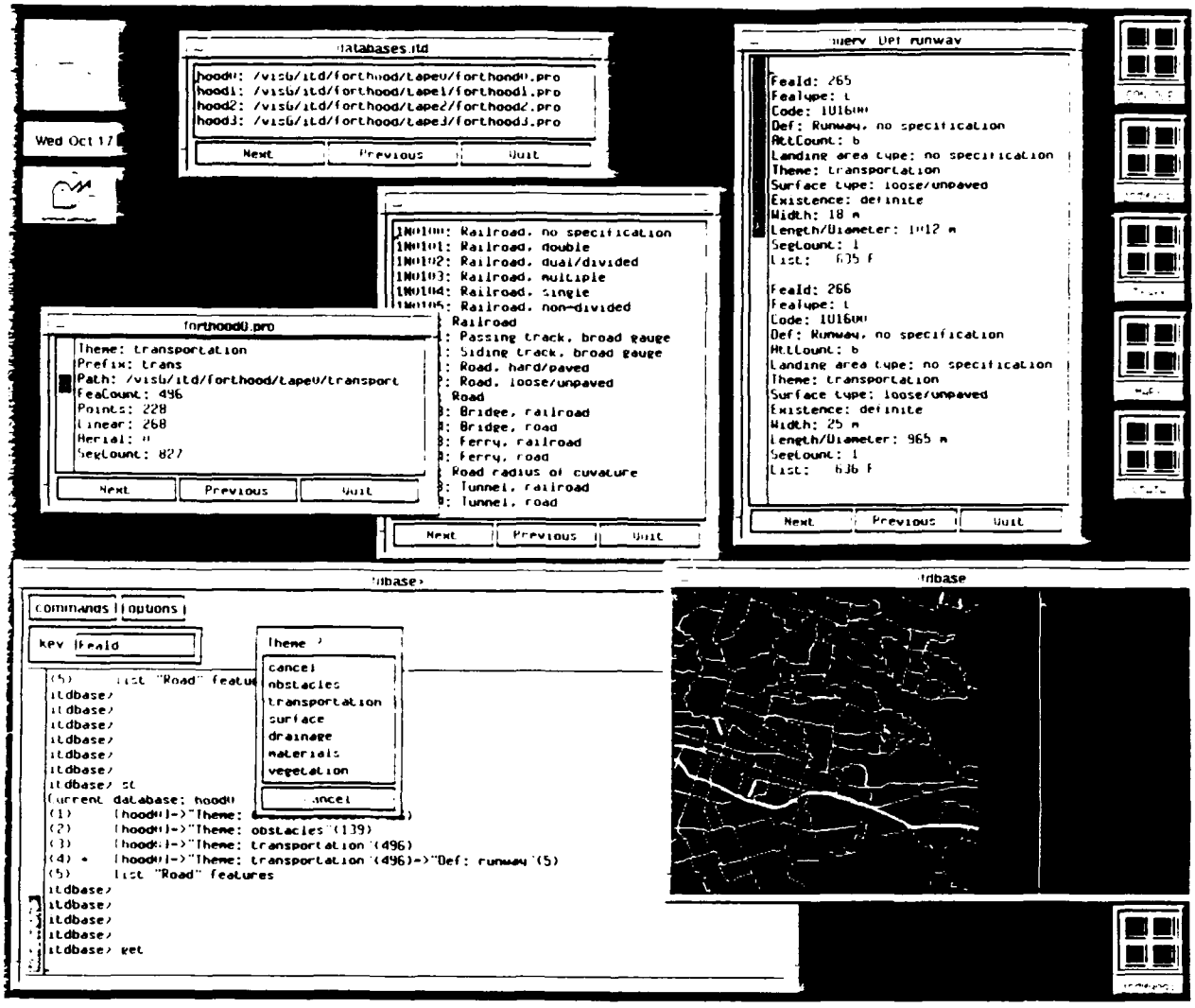

Figure 6-1: Sample Interaction with Graphical Interface

7. Conclusions

Computer vision and image understanding address difficult problems in a variety of task domains. In many cases, such as in certain industrial robotics applications, one can choose to engineer the problem domain in order to make automated sensing and manipulation tractable using current technology. In cartography, however, one is presented with two-dimensional images of the unconstrained three-dimensional world. We can not paint red squares on the corners of buildings in order to make roof detection more tractable for our computer vision techniques. Success and failure in these tasks are easily determined since we have a well understood basis for human performance in the cartographic community.

Although it is clear that humans bring a great deal of knowledge and context to bear when attempting to understand the structural and spatial relationships inherent in a scene, we are still a long way off from having such a level of expertise embodied in computer interpretation systems. The variety and complexity of man-made structures and natural terrain make the automated extraction and analysis one of the most difficult challenges for computer vision research.

In this final technical report under contract DACA 72-87-C-0001 we have described our progress toward automated cartographic feature extraction. Our research has put particular emphasis on built-up areas containing large numbers of complex man-made structures. Over the past three years we have attempted to address a fairly broad set of problems including scene registration, stereo analysis, shadow analysis, and building detection. Each of these areas addresses an important set of issues toward the development of automated tools for cartographic feature extraction.

In several cases we can see the inter-relationship between these different areas. The use of shadow cues for both registration and building detection, the use of monocular segmentations to refine disparity maps, and the fusion of various building hypothesis illustrate the need for many capable modules that can be used for a variety of purposes. Such a suite of feature extraction tools may provide the required foundation for more capable and robust systems that can reason about the structure and contents of the scene. Such a system needs to combine 'bottom-up' analysis across multiple images with *a priori* map knowledge to achieve the level of accuracy, robustness, and general performance required in order to be a useful and cost effective alternative to current manual map compilation techniques.

8. Bibliography

1. R. B. Irvin and D. M. McKeown. "Methods for exploiting the relationship between buildings and their shadows in aerial imagery", *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 19, No. 6, November 1989, pp. 1564-1575.

2. Aviad, Z., McKeown, D. M., Hsieh, Y., "The Generation of Building Hypotheses From Monocular Views", Tech. report, Carnegie-Mellon University, 1991, to appear

3. Huertas, A. and Nevatia, R., "Detecting Buildings in Aerial Images", *Computer Vision, Graphics, and Image Processing*, Vol. 41, April 1988, pp. 131-152.

4. Nicolin, B., and Gabler, R., "A Knowledge-Based System for the Analysis of Aerial Images", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. GE-25, No. 3, May 1987, pp. 317-329.

5. Aviad, Z., "Locating Corners in Noisy Curves by Delineating Imperfect Sequences", Tech. report CMU-CS-88-199, Carnegie-Mellon University, December 1988.

6. H. P. Moravec, *Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover*, PhD dissertation, Stanford University, September 1980.

7. Perlant, F. P., McKeown, D.M., "Scene Registration in Aerial Image Analysis", *Photogrammetric Engineering and Remote Sensing*, Vol. Volume 56, No. 4, April 1990, pp. 481-493.

8. McKeown, D.M. and Denlinger, J. L., "Cooperative Methods for Road Tracking in Aerial Imagery", *Proceedings IEEE Computer Vision and Pattern Recognition Conference*, June 1988, pp. 662-672.

9. Barnard, S. T. and Fischler, M. A., "Computational stereo", *Computing Surveys*, Vol. 14, No. 4, December 1982, pp. 553-572.

10. B. D. Lucas, *Generalized Image Matching By The Method of Differences*, PhD dissertation, Carnegie Mellon University, July 1984.

11. McKeown, D.M., McVay, C.A., and Lucas, B. D., "Stereo Verification In Aerial Image Analysis", *Optical Engineering*, Vol. 25, No. 3, March 1986, pp. 333-346.

12. Cochran, S.D. and Medioni, G., "Accurate surface description from binocular stereo", *Image Understanding Workshop*, DARPA, Palo Alto, CA., May 1989, pp. 857-869.

13. Hoff, W. and Ahuja, N., "Surface from stereo: integrating feature matching, disparity estimation and contour detection.", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 2, Feburary 1989, pp. 121-136.

14. Boult, T.E. and Chen, L.-H., "Synergistic smooth surface stereo", *IEEE International Conference on Computer Vision*, IEEE, Tempa, Florida, December 1988, pp. 118-122.

15. Grimson, W., *From Images to Surfaces: A Computational Study of the Human Early Visual System*, The MIT Press, Cambridge, Massachusetts, 1981.

16. Shufelt, J., McKeown, D. M., "Fusion of Monocular Cues to Detect Man-Made Structures in Aerial Imagery", *Proceedings: IAPR Workshop on Multisource Data Integration in Remote Sensing*, June 1990, Also available as CMU Technical Report CMU-CS-90-194

17. Ohlander, R. B., Price, K., and Reddy, D. R., "Picture Segmentation Using a Recursive Region Splitting Method", *Computer Graphics and Image Processing*, Vol. 8, 1978, pp. 313-333.

18. Yoram Yakimovsky, "Boundary and Object Detection in Real World Images", *Journal of the ACM*, Vol. 23, No. 4, 1976.

19. McKeown, D.M., Denlinger, J.L., "Map-Guided Feature Extraction from Aerial Imagery", *Proceedings of Second IEEE Computer Society Workshop on Computer Vision: Representation and Control*, May 1984, pp. 205-213.

20. Hsieh, Y., Perlant, F., and McKeown, D. M., "Recovering 3D Information from Complex Aerial Imagery", *Proceedings: 10th International Conference on Pattern Recognition, Atlantic City, New Jersey*, June 1990, pp. 136-146.