

SURFACE DESCRIPTION FROM BINOCULAR STEREO

Volume I

by

Steven Douglas Cochran

A Dissertation Presented to the
FACULTY OF THE GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA

In Partial Fulfillment of the
Requirements for the Degree
DOCTOR OF PHILOSOPHY
(Computer Engineering)

November 1990

Copyright 1990 Steven Douglas Cochran

Acknowledgments

I thank my advisors, Professor Gérard Medioni and Professor Ramakant Nevatia for their suggestions, advice, and direction; as well as their patience and encouragement throughout my research at USC. I thank Professor Bruce Abramson for finding time to serve on my thesis committee with very short notice.

I thank all of the people at the USC Institute for Robotics and Intelligent Systems during the last few years: Especially Andres Huertas, Professor Keith Price, and Dr. Philippe Saint-Marc for their help with problems of software and hardware. Also, Dr. Kashipati Rao and Dr. Rakesh Mohan for long talks through the night; Sylvie Menet and Shou-Ling Peng for their help with related software; and Dorothy Steele for her help and support over the years.

I thank Dr. William Hoff for his friendship and encouragement long ago which started me on this topic and his willingness to provide me with the images that he used for research.

I am very grateful to TRW which sponsored me as a Doctoral Fellow and provided the impetus to return, for a while, to academia; and to USC-IRIS for continued support as a Research Assistant.

Finally, I thank my parents Earl and Dorothy for their constant enthusiasm, encouragement and support.

Contents

Volume I

Acknowledgments	ii
List Of Figures	viii
List Of Tables	xi
Abstract	xii
Foreword	xiii
1 Introduction	1
1.1 Stereopsis in Man and Machine	1
1.2 Goals and Contributions	3
1.3 Organization of this Dissertation	4
2 Previous Work	5
2.1 Overview of Related Research	5
2.1.1 Feature Matching	7
2.2 Stereo Reasoning	11
2.2.1 Area-Based	12
2.2.2 Feature-Based	13
2.2.3 Global Optimization	15
2.2.4 Slider and Trinocular Stereo	16
2.3 Interpolation	17
2.4 Combined Methods	19
3 Problem Definition	20
3.1 Choice of Primitives and Strategies	22
4 Description of the Method	28
4.1 Overview of Methodology	28
4.2 Early Processing	33
4.2.1 Image Acquisition	34

4.2.2	Epipolar Alignment	34
4.3	Area-Based Processing	36
4.3.1	Local Variation Estimate	39
4.3.2	Correlation	40
4.3.3	Peak Extraction	41
4.3.4	Multi-level Disparity Estimate	42
4.3.5	Initial Disparity Estimate	42
4.3.6	Constraints	43
4.3.7	Interpolation	45
4.4	Feature-Based Processing	47
4.4.1	Edgel Extraction	48
4.5	Integrating Area and Edge Data	49
4.6	Visualization of Results	51
4.7	Surface Feature Extraction	55
5	Observations and Experimental Results	60
5.1	General Observations	60
5.2	Problems and Solutions	61
5.2.1	Wedding Cake (Random Dot Image)	62
5.2.2	Books	65
5.2.3	Jussieu	70
5.2.4	Blocks	75
5.3	Other Results	79
5.3.1	Montagne du Lubéron (SPOT Image)	79
5.3.2	Pentagon	83
5.3.3	Nuclear Power Plant	86
5.3.4	Fruit on a Table	90
5.3.5	Quarry Wall	94
5.3.6	Suspension Bridge	98
6	Complexity, Run Times and Error Analysis	102
6.1	Complexity Analysis	102
6.1.1	Local Variation Estimate	102
6.1.2	Estimate from Lower Level	102
6.1.3	Correlation	103
6.1.4	Peak Extraction	103
6.1.5	Disparity Estimate	108
6.1.6	Order Reversal	108
6.1.7	Multiple Viewpoint Constraint	108
6.1.8	Singleton Removal	108
6.1.9	Interpolation	108
6.1.10	Edge Extraction	114
6.1.11	Integration	114

6.1.12 Overall Complexity	114
6.2 Error Analysis	118
7 Conclusions and Further Research	119
7.1 Summary and Conclusions	119
7.2 Problems	120
7.3 Suggestions for Further Research	121
Reference List	122

Volume II

Appendix A

Introduction	130
A.1 Documentation	136

Appendix B

IRIS Library Routines	168
B.1 ImageCalc-Interface	168
B.2 Halftone	182
B.3 Image-Header	184
B.4 File-IO	204
B.5 IRIS-Cache	228

Appendix C

LINEAR	233
C.1 UTILS	235
C.1.1 IO	236
C.1.2 Utils	239
C.1.3 Windows	240
C.1.4 Display	241
C.2 General-Macros	246
C.3 Main-Modules	250
C.3.1 Get-Edges	250
C.3.2 Get-Segments	271
C.3.3 Get-Apars	297
C.4 Main-Program	303
C.4.1 linear	303
C.4.2 doc	305

Appendix D

MATCHER	312
D.1 Conv-Macros	315
D.2 Prompts	318
D.3 Geometry	324
D.4 Graphic-Macros	331
D.5 Stereo1	334
D.6 Stereo-Graphics	340
D.7 Main-Graphics	353
D.7.1 Main-Graphics	353
D.7.2 Main-Null-Graphics	358
D.7.3 Rem	361
D.8 Stereo2	362
D.9 DMAP	373
D.10 Threshold	381
D.11 Stereo3	383
D.12 Stereo4	407

Appendix E

Stereo Vision System	414
E.1 Macros	414
E.2 Variables	419
E.3 Primitives	441
E.3.1 General	441
E.3.2 Image	460
E.3.3 Graphics	481
E.4 Flavors	497

Volume III

E.5 Tools	509
E.5.1 Tools	509
E.5.2 Development	525
E.6 Image-Enhancement	528
E.6.1 Map	528
E.6.2 Enhance	536
E.7 Area-Based	541
E.7.1 Correlation	541
E.7.2 Build-Image	561
E.7.3 Merge-Views	570
E.7.4 Skeleton	582
E.8 Feature-Based	589
E.8.1 Edgel	589

E.8.2	Smooth	598
E.8.3	Match	601
E.9	Mix-Data	609
E.10	Surface Features	615
E.10.1	Surface	615
E.10.2	Folds	621
E.10.3	Features	629
E.11	Display Routines	657
E.11.1	Display	657
E.11.2	Render	673
E.12	Interface Routines	681
E.12.1	Function-Interface	681
E.12.2	Popup-Declarations	728
E.12.3	Process	843
E.12.4	IC-Interface	856

List Of Figures

2.1	Sources of Error in Area-Based Correlation	6
2.2	Range from Triangulation	8
2.3	Matched Pixels Define a Region of 3-D Space	10
3.1	Correlation Windows	23
3.2	Camera Geometry	25
4.1	Renault Part: Intensity Image Pair	29
4.2	Compressed Cross Correlation Array	29
4.3	Cross-Correlation Slice	31
4.4	Stereo Vision System Flow Diagram	32
4.5	Epipolar Adjustment	35
4.6	Renault Part — Collinear Intensity Image Pair	37
4.7	Area-Based Processing Flow Diagram	38
4.8	Renault Part: Local Variation	39
4.9	Left View of the Cross-Correlation	41
4.10	Left View of the Correlation Peaks and Edge Matches	41
4.11	Estimate Selection	42
4.12	Renault Part: Initial Disparity Estimate	43
4.13	Renault Part: Application of Constraints	44
4.14	Renault Part: Interpolation	46
4.15	Two Views Define a Smooth Surface	47
4.16	Renault Part: Edgels	48
4.17	Renault Part: Disparity Map After Incorporation of Edge Information	50
4.18	Renault Part: Labelled Points	52
4.19	Cross Section of Labelled Surface	52
4.20	Renault Part: Interpolated Estimate	53
4.21	Renault Part: Shaded View	53
4.22	Renault Part: Reconstruction	54
4.23	Renault Part: Depth Discontinuities	56
4.24	Renault Part: Orientation Discontinuities	56
4.25	Renault Part: Combined Discontinuities	56
4.26	Renault Part: Cross-Sections of Low and High Resolution	58
4.27	Renault Part: Combined Discontinuities of Low and High Resolution	59

5.1	Renault Part: Best Peaks Without Collinear Alignment	61
5.2	Wedding Cake: Original Intensity Images	62
5.3	Wedding Cake: Disparity Surface	62
5.4	Wedding Cake: Depth Discontinuities and Occluded Regions.	64
5.5	Wedding Cake: 3-D Plot of the Integrated Results	64
5.6	Books: Original Intensity Images	65
5.7	Books: Disparity Surface with only the Finest Level	66
5.8	Books: Disparity Surface Using Estimates from Coarser Resolution	66
5.9	Books: Surface Features	67
5.10	Books: 3-D Plot of the Integrated Results	68
5.11	Books: 3-D Rendered View of the Integrated Results	69
5.12	Jussieu: Original Intensity Images	70
5.13	Jussieu: Edgels	71
5.14	Jussieu: Disparity Surface	71
5.15	Jussieu: Chamferred Edges	71
5.16	Jussieu: Refinement of Depth Discontinuities	72
5.17	Jussieu: 3-D Plot of the Integrated Results	73
5.18	Jussieu: 3-D Rendered View of the Integrated Results	74
5.19	Blocks: Original Intensity Images	75
5.20	Blocks: Disparity Surface	75
5.21	Blocks: Surface Features	76
5.22	Blocks: 3-D Plot of the Integrated Results	77
5.23	Blocks: 3-D Rendered View of the Integrated Results	78
5.24	Lubéron: Original Intensity Images	79
5.25	Lubéron: Disparity Surface	79
5.26	Lubéron: Surface Features	80
5.27	Lubéron: 3-D Plot of the Integrated Results	81
5.28	Lubéron: 3-D Rendered View of the Integrated Results	82
5.29	Pentagon: Original Intensity Images	83
5.30	Pentagon: Disparity Surface	84
5.31	Pentagon: Depth Discontinuities and Occluded Regions	84
5.32	Pentagon: 3-D Plot of the Integrated Results	85
5.33	Pentagon: 3-D Rendered View of the Integrated Results	85
5.34	Power Plant: Original Intensity Images	86
5.35	Power Plant: Edgels	87
5.36	Power Plant: Disparity Surface	87
5.37	Power Plant: Chamferred Edges	87
5.38	Power Plant: Refinement of Depth Discontinuities	88
5.39	Power Plant: 3-D Plot of the Integrated Results	88
5.40	Power Plant: 3-D Rendered View of the Integrated Results	89
5.41	Fruit Scene: Original Intensity Images	90
5.42	Fruit Scene: Disparity Surface	90

5.43	Fruit Scene: Surface Features	91
5.44	Fruit Scene: 3-D Plot of the Integrated Results	92
5.45	Fruit Scene: 3-D Rendered View of the Integrated Results	93
5.46	Quarry Wall: Original Intensity Images	94
5.47	Quarry Wall: Disparity Surface	94
5.48	Quarry Wall: Surface Features	95
5.49	Quarry Wall: 3-D Plot of the Integrated Results	96
5.50	Quarry Wall: 3-D Rendered View of the Integrated Results	97
5.51	Suspension Bridge: Original Intensity Images	98
5.52	Suspension Bridge: Disparity Surface	98
5.53	Suspension Bridge: Surface Features	99
5.54	Suspension Bridge: 3-D Plot of the Integrated Results	100
5.55	Suspension Bridge: 3-D Rendered View of the Integrated Results	101

List Of Tables

3.1	Comparison of Matching Methods	21
5.1	Wedding Cake: Matching Statistics	63
6.1	Breakdown of Overall Timing	103
6.2	Local Variation Estimate Times	104
6.3	Pyramid Estimate Times	105
6.4	Correlation Times	106
6.5	Peak Extraction Times	107
6.6	Initial Disparity Times	109
6.7	Order Reversal Times	110
6.8	Viewpoint Constraint Times	111
6.9	Singleton Removal Times	112
6.10	Area Interpolation Times	113
6.11	Edge Extraction Times	115
6.12	Area/Feature Integration Times	116
6.13	Overall Run Times	117
6.14	Comparison of Matching Results	118

Abstract

We describe a Stereo Vision System which attempts to achieve robustness with respect to scene characteristics, from textured outdoors scenes to environments composed of highly regular man-made objects. Unlike most stereo approaches, it integrates “Area-based” and “Feature-based” primitives. This allows it to take advantage of the unique attributes of each of these techniques. The area-based processing provides a dense disparity map and the feature-based processing an accurate location of discontinuities. The integrated disparity map is accurate enough to detect depth and orientation discontinuities, within the limits of the available resolution.

As a result of this integration, edgels parallel to the epipolar lines, discarded in most feature-based systems, play an important function (along with all edgels) as the locations of potential discontinuities. In real-world scenes, our system can locate the depth discontinuities and occluded regions; in areas where surface matches cannot be determined, it labels the points as unknown rather than providing an incorrect guess. Resulting surface patches are very accurate, so the system also provides an estimate of convex and concave orientation discontinuities.

A detailed description of the Stereo Vision System is accompanied by a step-by-step illustration using a stereo pair of a real-world scene. Ten further examples follow, which point out specific problems and solutions, as well as demonstrate the wide domain of applicability. We have obtained very good results on complex scenes in different domains and have been able to locate visible surfaces, occluded areas, depth discontinuities and, in most cases, orientation discontinuities in the images.

Foreword

It is always a problem to present stereo research results, since much of the excitement is missing without the stereo effect. In the book, “Foundations of Cyclopean Perception” [40], Bela Julesz provides an enclosed set of lenses to aid in viewing stereo pairs. There also exist lenses which serve the same purpose for viewing side-by-side stereo pairs or for viewing anaglyphic images. Other papers have used “cross-eyed stereo” in which the left and right views are swapped so that they may be viewed with the left eye looking at the right image and the right eye looking at the left image.

In this document, we have decided that the most widely usable method is to show the stereo images as side-by-side pairs for viewing with standard stereo lenses which are widely available and which some persons are able to view unaided. An example of this sort of image is shown below. Those stereo pairs which are especially important to the understanding of the paper are printed at a higher resolution to make them clearer.



Chapter 1

Introduction

1.1 Stereopsis in Man and Machine

Humans are able to surmise depth in 2-D “monocular” images and to perceive it through the stereoscopic fusion of a pair of images. The process used by the human visual system to achieve this, however, is not well understood. How, then, can we produce autonomous systems which are able to extract depth information about, and interact with, their environments? One way is to attempt to emulate some of those processes, which seem to be present in the human visual system, and learn what does and does not work. This approach allows us to attempt automatic visual recognition of objects, and, perhaps in some cases, to gain insight about the operation of the human visual system.

The perception of depth from a stereo pair provides both a qualitative and a quantitative measure in depth determination that is not available from a monocular view [13] and produces the ability to detect objects which are camouflaged. This ability is made possible by the overlap of the visual field of the eyes which produces a small but perceptible lateral displacement of the image of an object projected onto the retinas. This lateral displacement, or disparity, is inversely related to relative distance from the point of fixation which is in turn defined by the **vergence angle**, the angle at which the eyes tilt inwardly toward the nose.

The subject of this thesis is the process of stereoscopic fusion. We have developed a system which allows the detection of the relative disparity between points, using features extracted from the scene. Depth or range information extracted from stereo scenes has been used in several areas including: Robot Navigation [5, 6, 30, 33, 59, 72, 78], Industrial Parts Recognition [4, 18, 31, 38, 39, 44, 54, 56, 61, 63], and Stereo Mapping [7, 11, 34, 35, 38, 44, 60, 63, 64, 67].

The recovery of the 3-D characteristics of a scene from multiple images taken from different points of view, may be viewed as consisting of the following steps (from Barnard and Fischler [10]):

1. Image Acquisition
2. Camera Modeling
3. Feature Acquisition
4. Image Matching
5. Depth Determination
6. Interpolation

Of these steps, image matching is widely considered to be the most difficult to solve, and is clearly dependent on the choice of feature primitives. Given two views of a scene, a correspondence must be established between those points, or features, which are visible in both scenes. When the matched features are low-level and dense, such as the intensity at each pixel, we call the matching strategy **area-based**, while for sparse, more abstract features, such as edge-segments (*e.g.* the boundary of a shadow) we use the term **feature-based**. Some systems, like ours, use a hybrid approach with multiple features both sparse and dense.

The problem of matching the selected features is made difficult by several problems:

Photometric Variation The light which is reflected from the scene and recorded by the camera depends on the position of that camera relative to the scene, as well as noise and nonlinearities in the camera itself. Thus, when a camera is moved to a new position, or when two cameras view a scene from two viewpoints, the intensity at the corresponding points may be different.

Occlusion Occlusion is due to the occurrence of a depth discontinuity which causes an obstructed view of part of the scene, behind the occluding edge, and thus is “observed” by only one of the cameras. The edge of the image can also act as a sort of occluding edge in this sense.

Repetitive Texture When the texture is repeated, such as the bricks in a brick wall, multiple possible correspondences exist. This problem is exacerbated when coupled with photometric variation and can lead to wrong matches which, due to these distortions, appear to be better matches within a local area.

Lack of texture In real world scenes, most objects and surfaces are textured. While texture often does not give rise to useful, matchable abstract features, such as segments, it forms a basis for the real strength of an area-based match by providing a statistically matchable pattern. However, part of the scene may be without texture and, in that area of the scene, no area-based or feature-based match will be possible.

1.2 Goals and Contributions

Stereoscopic matching, or the generation of a dense depth map, should not, however, be considered to be the principal goal of this process. Rather, it should be viewed as one of the first steps toward the understanding of a scene. Therefore, it is crucial for the important elements of the depth map, such as depth and orientation discontinuities, to be detected and localized. The detection of the discontinuities is obtained from the area-based process, but, because this process tends to blur the results, these discontinuities cannot be accurately localized. Fortunately, this is where the feature-based processing works best, since the area-based process gives the same values at the features, it is really the existence of the edgels (edge elements) that is important. Since discontinuities tend to give rise to changes in intensity, a subset of the edge-features corresponds the locations of those discontinuities. Thus, the edge-features provide the exact, non-blurred, location for the discontinuities. This also provides a use of those edgels parallel to the epipolar lines which most feature-based methods throw out as unmatchable.

The main difficulty with feature-based methods is that, once the edges are matched, we still have the problem of deciding which side of an edge is the foreground and which is the background, or whether both are on the face of an object as in the case of a surface-marking. The area-based process provides such an estimate, thus a combination of the area-based and the feature-based results can provide better results than either process alone.

Once depth-discontinuities are detected in the area-based estimate, they may be accurately localized by finding the corresponding edge feature. Then the geometry of the scene allows the estimation of occluded regions. Once this information is known, the disparity surfaces may be smoothed (away from the depth discontinuities) to remove local irregularities. From the smoothed disparity surfaces the orientation discontinuities can be estimated as those areas of the greatest positive and negative curvature. These discontinuities are also associated with the edge-features and may be repositioned locally.

The major contributions presented in this thesis are:

1. A use for the “horizontal” edge-features,
2. The integration of area-based and feature-based results,
3. The location of the occluded areas in the scene, and
4. The extraction of depth and orientation discontinuities.

1.3 Organization of this Dissertation

We start in Chapter 2 by giving an introductory to stereo processing and a summary of related research in this field, and in Chapter 3 provide a more detailed description of the specific problem area and the goals of this research. Next, in Chapter 4, we present our approach to stereo correspondence in detail and illustrate the step-by-step processing on a stereo image pair of a Renault part, an image pair used by researchers in stereo vision. In addition, we present reconstructions of the original scene. In Chapter 5 we show the results of processing on several other images and present a discussion of the problems encountered and the tradeoffs involved with our approach. Chapter 6 presents the complexity and error analysis of our system and its results. We discuss the contributions of this research to the field of computer stereo vision in Chapter 7, and highlight those areas where these ideas may be applied. In addition, suggestions for further research are offered here.

Chapter 2

Previous Work

2.1 Overview of Related Research

Various methods have been used in the attempt to determine a description of objects contained in a photometric image. In a single image, object contours and vertices have served as monocular cues, while multiple images of a scene taken from slightly different spatial locations allow the range from the cameras to the scene to be determined by triangulation. This “stereo” approach is generally divided into two groups: area-based and feature-based. These groups represent the level of abstraction of the entities being matched. The area-based techniques attempt to find matches of features present throughout the image, such as the intensity at each pixel or the correlation of a local window of intensities; while the feature-based techniques attempt to find matches of sparser, more abstract, features such as edges or segments.

Area-based methods have been applied successfully to the analysis of aerial terrain images, where the surface can be assumed to vary smoothly and continuously. However, these methods have difficulty with scenes that contain abrupt changes in depth due to large surface gradients or to the presence of depth discontinuities. This is because the correlation windows that cross such regions in the image cannot usually be correctly matched, as is illustrated in figure 2.1. The first case, can be managed, by locally changing the width of the correlation window [60, 68]. However, in the case of a depth discontinuity, there is no easy solution. Using a smaller window can reduce the amount of error, and therefore yield a better 3-D position, but it also reduces the statistical significance of the cross-correlation, and increases the sensitivity to noise.

Since features are abstractions of the underlying intensity changes rather than the actual intensities, they are not as subject to the mismatch across a depth discontinuity. Only when this mismatch causes the transition across the edge to be inverted

Figure 2.1: Sources of Error in Area-Based Correlation.

(dark-to-light as opposed to light-to-dark) is matching a problem, and even then, the position of the edge is well localized. Thus, feature-based analysis provides more precisely positioned features in the individual images and can attain correspondingly higher accuracy for the correspondences in 3-D. Arnold [1,2] indicates that edge-based techniques offer an order of magnitude improvement in accuracy over area-based correlation methods. There are, however, two major problems with edge-based stereo. The first is that it provides only a sparse depth map, and the second is that when real data is used, there are still errors due to noise and occlusion which must be accounted for. Also, there are difficulties in detecting and linking the edgels, which can cause edges to be broken or missing.

Most of the stereo research related to feature based matching falls into two areas: feature matching and surface reconstruction. There are several different approaches to finding a set of stereo matches and to reconstructing a surface from the matched information. In addition there are tradeoffs between them. For instance, it is easier to match abstract features, such as line segments, than for less abstract ones, such as an area-based correlation window. But then it becomes more difficult to reconstruct the underlying surface from the matching information. Some researchers have attempted to combine these two processes into one in order to provide a **global** consistency constraint into the matching process. In this section we review those approaches which have most strongly influenced our work, and refer the reader interested in the state-of-the-art in computational stereo to the excellent reviews by Barnard and Fischler [10] and, more recently, Dhond and Aggarwal [25].

2.1.1 Feature Matching

Before examining these methods we first explain how finding the matching points gives us the range to these points and how the surface interpolation interacts with the matching. Then some history of the main techniques in stereo processing is presented.

Geometry

The basis of stereo matching is that given a point, P_w (see figure 2.2), whose projections P_l and P_r onto the image planes of a pair of cameras (designated, here, as “left” and “right”); we are able to calculate its position in space — especially its distance from the cameras. If we designate the Cartesian coordinates the projections to be (X_l, Y_l) , and (X_r, Y_r) in the respective camera coordinate systems, then to solve for the location, (X_w, Y_w, Z_w) , of P_w in the world coordinate system, we assume that the cameras may be modeled by a “pinhole approximation”¹ with a focal length of f , and that we know the relationship between the three coordinate systems. With no loss of generality, we let the coordinate systems of the world and the left camera be the same, then to simplify the problem, we also assume that the camera axes are parallel and that the coordinate system of the right camera be similar but translated a distance B along the X-axis. This distance is called the “Stereo Baseline.” Then by similar triangles we have:

$$\begin{cases} \frac{x_l}{f} = \frac{X_w}{Z_w} \\ \frac{x_r}{f} = \frac{X_w - B}{Z_w} \end{cases} \quad (2.1)$$

solving for X_w in (2.1) gives:

$$\begin{cases} X_w = \frac{x_l Z_w}{f} \\ X_w = \frac{x_r Z_w}{f} + B \end{cases} \quad (2.2)$$

¹We place the focal plane in front of the hole rather than behind to simplify the drawing.

Figure 2.2: Range from Triangulation.

Equating the left sides in (2.2) and solving for Z_w gives the relation between the range and the disparity, $D = (x_l - x_r)$, in figure 2.2.

$$Z_w = \frac{Bf}{x_l - x_r} = \frac{Bf}{D} \quad (2.3)$$

substituting back into (2.2) we obtain

$$X_w = \frac{x_l B}{D} \quad (2.4)$$

From similar triangles and (2.3) we can also specify:

$$Y_w = \frac{y_l B}{D} \quad (2.5)$$

Range and disparity, then, are related and inversely proportional to each other. Disparity is the amount that the projection of a point appears to shift from one viewpoint to another. Now it is clear that once we know that projection P_l matches P_r in the two images planes, we can find where that point is in 3-D space with the additional knowledge of the separation of the images (the stereo baseline, B) and the focal length (f) of the two cameras.

Note also, that when we match pixels on the image plane, we are working with small, approximately square, matched regions which define an oddly shaped hexahedral region of space as shown in figure 2.3. Because of this discrete nature of the digitized images, the coordinates of the actual matched point can be anywhere within this volume, and therefore have an error of up to $\pm\frac{1}{2}$ pixel in the X and Y location and up to ± 1 pixel in the disparity. The actual (relative) range error is directly proportional to the image sampling interval and inversely proportional to the stereo baseline and the focal length [69].

Interpolation

We wish to describe surfaces, rather than individual points, and the point matches presented in the prior section generate a surface only if they are dense. If we begin with a sparse feature (for feature-based matching), then the best that we may hope for is a more or less sparse set of matches. The algorithms which attempt to extend these sparse matches to form surfaces are called Interpolation Methods. They may be grouped into three methods:

Figure 2.3: Matched Pixels Define a Region of 3-D Space.

Relaxation: in which a measure of the total energy of the surface is reduced relative to constraints provided by the matched points and those which determine the surface energy,

Analytic: in which some best-fit is made between the matched points and an analytic surface, and

Heuristic: in which unknown points on the surface are determined based on local rules and surrounding points.

The first is most often used for the sparse, feature-based, matches; the last works best for the dense, area-based, matches; and the second method is used with both approaches.

2.2 Stereo Reasoning

The modern interest in the correspondence problem may be traced to the introductory work of Julesz [40] who realized the usefulness of stereopsis in breaking camouflage. He also introduced the random dot stereogram as a research tool which has been used to establish that the human visual system is able to extract disparity information even in the absence of monocular cues. In addition, he pointed out the need for some mechanism to select the “correct” matches from a set of possible ones. This imposes more constraints on the possible matches, which is useful since the matching problem is under-constrained and there exist several reasonable solutions, only one of which represents the real-world. These additional constraints impose a local and global consistency to resolve this ambiguity and to select a solution.

The process of stereopsis has been regarded by Mayhew [52] as consisting of two subproblems:

- (i) the correspondence problem which is the problem of measuring the disparity of each point in the two eyes’ projections, and,
- (ii) the interpretation problem which is the use of disparity information to recover the orientation and distance of the surfaces in the scene.

Mayhew and Frisby [53] describe psychophysical studies which suggest that when ambiguity exists in the local matching, those matches which preserve the “figural continuity” are to be preferred. In addition, correspondences between spatial frequency channel² matches may be used to help disambiguate the within-channel fusions.

²A band-pass channel that allows only a limited spatial-frequency range, see Marr[48, pp. 61–63].

2.2.1 Area-Based

Mori *et al.* [60], Hannah [32–34], Nevatia [61], Moravec [59], and Panton *et al.* [64] have attempted to find the corresponding points using area-based correlations. These methods make the assumption that the texture of the region near a point in the scene is:

1. The same from both points of view, and
2. Detectably different from other points in the scene.

While methods exist to reduce the possible search region, the area-based systems still fail when either of the above two assumptions are false within the search region. Surface discontinuities, as well as very steep surfaces, can cause the first assumption to be false, as well as surface reflectivity and lighting. The second is violated when the scene has either a lack of texture or when the texture is repetitive.

Area-based methods have been applied successfully to the analysis of aerial terrain images, where the surface varies smoothly and continuously. They offer the advantage of directly generating a dense disparity map, but are sensitive to noise and breakdown where there is a lack of texture or where depth-discontinuities occur.

Mori *et al.* [60] used an iterative prediction and correction method to improve their area-based processing of aerial photographs by varying the size of the window and they then used edgels to verify their prediction. Their program was only demonstrated for aerial imagery and they had no method of handling occluded areas. This adaptive cross-correlation was re-introduced as part of a coarse-to-fine hierarchical control structure by Quam [68] with the same limitations.

Hannah [32, 33] also showed improvement in area-based cross-correlation by using dense features abstracted from the intensity data and introduced heuristics for inferring the distinctions between occlusions, correspondence errors, and off-the-edge overlaps. She has implemented a complete system [34] for stereo processing with little or no operator intervention. First a set of well-scattered, reliable matches are obtained by locating interest points based on variance and edge strength (using a modified Moravec operator), and then utilizing an unconstrained hierarchical match algorithm. Next, a camera calibration is performed and an epipolar-constrained hierarchical matching algorithm is used to match the interest points. Those points which are evaluated (using autocorrelation) to have the most reliable matches are used as “anchors” for a final matching of all of the interest points. The matches are checked at the finest level by reversing the role of the left and right images. An interpolation may also be performed to construct a regular grid of points if desired.

2.2.2 Feature-Based

The currently preferred approach in the vision research community is to match more abstract features, rather than texture regions in the two images, since such features are less sensitive to noise and reduce the search space. Feature-based analysis provides more precise positioning (for the feature) in the individual images and it can attain correspondingly higher accuracy for its correspondences in 3-D (Arnold [1, 2]). The most commonly used features are points along the edges of intensity discontinuities. These points, or edgels, are useful because they represent those locations at which most of the scene information is available. Thus the advantages of feature based stereo are that these methods are faster, since fewer points must be processed, and potentially more accurate, because the edgels may be located with sub-pixel precision. In addition, since edgels represent an abstraction from the intensity image, they are less sensitive to photometric variations such as absolute intensity, contrast, or illumination. Systems that use edgels as the matching primitives for stereo correspondence include Barnard and Thompson [11], Grimson [31], Milenkovic and Kanade [56], Hoff and Ahuja [36], Tsuji *et al.* [78], and Boulton and Chen [18].

Other researchers, including Arnold [1, 2], Baker [7], Ayache and Faverjon [4, 5], Lim and Binford [44], Medioni and Nevatia [54], Ohta and Kanade [63], and Ayache and Lustman [6] use connected sets of edgels, either curves or linear segments, for the matching primitives. While Lim and Binford [44], Price [67], and Mohan [58] matched segmented regions of the image.

One disadvantage of the feature-based methods is that since they provide only sparse matches they require interpolation as well as some method for modeling occlusion. In addition, the feature-based process may be confused by a large local change in disparity, and it is very difficult to incorporate the smoothness assumption into the matching strategy since it is most likely to be violated at edges.

- Barnard and Thompson [11] used a relaxation labelling approach to determine the most likely pairing of candidate points from two images. They use figural continuity to guide the relaxation, but generate an extremely sparse set of matches.
- Marr and Poggio [49] proposed a computational model of human stereo vision, using zero-crossings in the Laplacian of the Gaussian of the intensity as a matching feature. They suggest that three constraints should be satisfied in choosing global correspondence: compatibility, uniqueness, and continuity. The latter is similar to the figural continuity constraint proposed by Mayhew and Frisby [53]. Grimson [31] implemented an improved version of this model, which gives good results when there is a sufficiently dense set of features. It has

difficulties near the occluding edges since it uses local continuity to verify the correct matches and therefore tends to fail along these edges. In addition, this method requires a fairly dense set of features which typically is not associated with feature-based methods.

- Arnold [2], and Baker [7] used various forms of the Viterbi dynamic programming algorithm to match edges, and Ohta and Kanade [63] extended Baker’s inter-scanline search, again using dynamic programming, to find an optimal matching surface (however their three-dimensional search is very expensive).
- Medioni and Nevatia [54] matched linear edge segments which cut through the epipolar lines and thus automatically insure inter-scanline continuity. Ayache and Faverjon [5] extended this approach by using segment continuity to propagate the matching and a grid of “buckets” to speed the processing of the local neighborhood. Mohan *et al.* [57] also used propagation of matching information along the segments, as a separate process from the actual stereo matching, to provide error detection and correction.
- Lim and Binford’s system [43, 45] extended Arnold’s work beyond curve matching by attempting to identify and to use many types of features: Edgels, junctions, curves, surfaces and bodies; as “quasi-invariants.” They then perform the stereo matching on all of the features using the hierarchical correspondence to support or reject potential matches.
- Mohan [58] matches “collated features,” which are features such as curves and junctions which have been grouped into “perceived” arrangements: linear features, parallels, U-contours, rectangles. Features are grouped based on proximity, continuation, symmetry, closure and familiarity (this latter is restricted to recognizing rectangles). Like Lim and Binford, these high-level features serve as quasi-invariants for matching. Their system has given impressive results on a number of block world scenes and on aerial photographs of buildings. However, they produce only sparse matching at a high (abstract) level.
- Wildes [79] took a different approach by treating the disparity as a vector field relating the two images and used analytic techniques from classical field theory to effect the recovery of surface depth, orientation, and discontinuities. He notes that his approach is not at odds with studies on human subjects which indicate that the humans may base their judgments of surface discontinuity on depth and first-order surface geometry only, without using second-order or

higher characteristics. His algorithm, however, provides information only in the region near discontinuities.

2.2.3 Global Optimization

Global optimization strategies are commonly provided in two ways: First by some form of interaction between spatial frequency channels. This may be a simple coarse-to-fine strategy, such as that presented by Marr and Poggio [49] or involve the parallel interaction of information in different channels as discussed by Mayhew and Frisby [53]. The coarser or lower-frequency channels have a larger “local” area while the finer or high-frequency channels provide more exact features.

The other approach, discussed by Barnard [9] and by Zhou and Chellappa [82, 83], is to define some form of global energy to the matches and search for a state which minimizes the total energy. This energy function attempts to characterize the extra constraints which may be imposed on objects in the real-world in order to maximize the number of matches while preserving, as much as possible, the preference for the local change in disparity being small and continuous. It also allows for the integration of multiple sources of information.

Marr and Poggio’s original cooperative stereo algorithm [49] provides for global correspondence by generating the edgel matches by means of a cooperative process incorporating three rules:

- (1) Compatibility. “Black dots can match only black dots.” More generally, the zero crossings can only match others with the same sign and rough orientation.
- (2) Uniqueness. Almost always, a black dot from one image can match no more than one black dot from the other image.
- (3) Continuity. The disparity of the matches varies smoothly almost everywhere over the image. This rule is also characterized by the phrase: “Matter is Cohesive.”

In their later work [50] they incorporate a coarse-to-fine strategy wherein the coarse match provides a disparity estimate which is used to locally align the images for the finer-level processing.

Mayhew and Frisby [53] provided some improvements to this by stressing figural continuity and correspondences between channel outputs to achieve the global optimization. The advantages of propagating information between spatial channels rather than only in a coarse-to-fine direction is that an initial mistake at one level is not automatically propagated to all other levels, but can be cooperatively

corrected. In addition, the requirement of consistency of the disparity along the contours matches the real world better than the more local agreement on constant disparity of Marr and Poggio.

Barnard [9] used a modified version of the Metropolis algorithm to achieve a dense matching based on two competing constraints. First that the matched points should have a similar intensity, and second, that the disparity surface should be smooth. The Metropolis algorithm implements a simulated annealing, which is a stochastic optimization technique which avoids local minimum of the state energy by slowly lowering the “temperature” which constrains the probability of a random change increasing the system state. As the temperature is lowered, the final “cold” state is one that represents an optimal or near-optimal solution in which the two constraints are approximately satisfied. This approach works fairly well but is very slow and does not model occluded regions in the scene.

Zhou and Chellappa [82, 83] have used a neural network of binary “neurons” for matching, where each neuron represents a discrete point in the disparity space and is mutually interconnected to the others. In this model, global optimization is achieved through stochastic relaxation under a set of constraints between the neurons: Only one neuron firing for a given pixel and the network attempts to minimize the total energy of the system which is defined to be a function of the absolute difference in intensity gradient and a local flatness constraint.

Drumheller and Poggio [26] have attempted to combine several current algorithms to extend the Marr and Poggio scheme. They have implemented a parallel stereo algorithm for the Connection Machine. The most interesting aspect of their implementation is their active use of a “forbidden zone,” (where no matches can occur) if a given match is considered valid. This constraint enforces opacity and uniqueness, but the continuity forces a flat rather than smooth surface.

2.2.4 Slider and Trinocular Stereo

Some researchers have used more than two views to supply additional constraints to the stereo matching. These may come from a collinear array of cameras, or a linearly moving camera, which provide a larger baseline (more accuracy) along the entire array while retaining a small change in position between adjacent views (better matching), as used by Nevatia [61], Moravec [59], Bolles *et al.* [16], and Matthies *et al.* [51]. Or the views may form a non-collinear pattern (usually of three cameras), in order to gain extra sets of epipolar alignments, as demonstrated by Milenkovic and Kanade [56], and Ayache and Lustman [6]. This allows for selection of the correct match from among the potential binocular matches — but yields only a sparse set of trinocular matches.

Nevatia [61] utilized a series of closely spaced views, which has an advantage of reliability over the use of only two images. In this work, the similarity of binocular stereo and motion stereo is shown, one dealing with an instant in time and the other allowing a limited degree of motion.

Moravec [59] used a series of images taken at precise intervals by a computer controlled cart. The resulting images are matched pairwise by a correlation operation which “frequently matches features incorrectly.” The correct matches are derived by comparing the results from multiple pairs of images. Bolles *et al.* [16] extended this idea by extracting features which have an extent through time, giving a “spatial-temporal solid” of data which may be sliced along the time axis for each epipolar line of the image giving an “epipolar-plane” in which motion-stereo can be analyzed. Matthies *et al.* [51] used a slider stereo acquisition of images to drive a Kalman-filtering based model of motion/position stereo. This has the advantage over the “epipolar-plane” approach of allowing the images to be processed by incrementally refining the depth model.

Yachida *et al.* [80] used an edge-based matcher based on a trinocular arrangement of cameras in which the cameras are placed in a non-collinear arrangement so that the matching along the epipolars is reinforced by having three sets of pairs, which must agree to generate a match. This approach is interesting since it presents one solution to the “horizontal” problem, however, there are fewer points matched. Milenkovic and Kanade’s [56] also used this trinocular arrangement and defined a set of confidence measurements for each aspect of the match: position, orientation, and photometric similarity. Ayache and Lustman [6] extended the feature-based method of Ayache and Faverjon [5] to three cameras using the same prediction and verification approach to select and match candidate edges.

2.3 Interpolation

The other aspect of the recovery of 3-D characteristics of a scene that we plan to explore is the interpretation of the match data, especially to allow an interpolation of data between the sparse points provided by the feature-based stereo matching.

In both area and feature based stereo correspondence methods, it is often necessary to interpolate values for those regions for which no disparity can be found. Some methods such as those by Hoff and Ahuja [38] and Boulton and Chen [18] have included the interpolation into the normal processing. In addition to interpolating, these methods are also used to smooth surfaces and to isolate depth discontinuities.

Grimson [31] interpolated surfaces from sparse depth data using variational methods, while Terzopoulos [76] attempted to locate discontinuities by locating significant

inflection points on the resultant surface. Grimson [31] and Terzopoulos [74, 75] provide mechanisms to interpolate between the sparse data to provide a surface representation which represents the surface with the minimum energy that passes through the constraints using a variational calculus technique, the latter using in addition a multilevel approach developed by Brandt [19, 20]. But this method has the problem noted by Blake [14] and by Cochran and Medioni [23] that, if the density of constraints (matched edgels) is insufficient in the image, then the surface tends to “sag” in the direction of the zero-energy plane. The reason for this is that the accurate modeling of a thin plate (which is used as a surface model since it bends without creasing) is a rather intractable, nonlinear problem.³ Both Grimson and Terzopoulos make some simplifying assumptions in order to approximate the surface energy by quadratic variation. These assumptions are:

1. The plate is thin compared with its extent.
2. The displacements of the plate from its equilibrium position (the zero-energy plane) are substantially in the direction of the viewer and transverse displacement is negligible.
3. The deflection of the plate is everywhere small compared with its extent.
4. The deflection of the plate is everywhere small compared with its thickness.

The above assumptions assume a *flat* rather than a *smooth* surface which is grossly violated in almost any image.

Blake and Zisserman [15] introduced their “Graduated Non-Convexity” algorithms, which allow the direct search for depth discontinuities and orientation discontinuities respectively. This algorithm uses a series of related convex energy functions (*i.e.* functions whose local minimum is the global solution) which successively make better approximations to the global best-solution which is represented by a non-convex energy function due to constraints added to allow the discontinuities.

Another method was used by Hoff who produced an interpolation of his data points (matched zero-crossings) by growing planar patches which he broke at discontinuities. Hoff used a heuristic similar to those proposed by Terzopoulos [75], and Langridge [42] for locating discontinuities during or after the interpolation process. But, like the analytical methods, Hoff’s approach still requires dense features in order to return reasonable results.

³Without the simplifying assumptions we get von Karmann’s *large deflection theory* of the thin plate which is considerably more complicated and leads to Euler-Lagrange equations in the form of two coupled, nonlinear, partial differential equations of the fourth order. Von Karmann’s equations are described in detail in [47, 77].

Saint-Marc *et al.* [70, 71] presented a method to smooth the surface while preserving discontinuities to facilitate the detection of discontinuities, and Sinha and Schunck [73] have recently introduced another process for preserving discontinuities while performing the surface reconstruction which uses a weighted bicubic spline which is adapted across discontinuities.

Boult and Kender [17] have analyzed other class and norm pairs than those used by Grimson and Terzopoulos, and can provide solution methods which exhibit less overshoot and undershoot at discontinuities. However all of these methods make the assumption that the scene is smooth with no discontinuities, or that the discontinuities are given.

2.4 Combined Methods

Baker [7] used a combined approach based on the Viterbi Dynamic Programming algorithm. Instead of matching the edges as most approaches do, he matched the half-edges: each edge generates a left and a right half to be matched. Now the problem of whether the edge represents a discontinuity becomes dissociated from the matching process since the edge is broken into two half-edges, each terminating a different interval. Once the half-edges are matched, the intervals between them are analyzed and the pixels making up the intervals are matched, again using the Viterbi algorithm, this time matching the intensity.

Hoff and Ahuja [36–38] attempt to combine the feature matching, contour detection, and surface interpolation into one process. Their results are very impressive, but they fail when their matching features (zero-crossings) are too sparse, also they cannot accurately locate the surface discontinuities. Their method begins by extracting edgels using a Laplacian of the Gaussian operator and the matching proceeds using both a left-to-right and a right-to-left pass. The possible surfaces (first planar and then quadratic) are considered using a Hough transform for points within a circular patch about each point on a grid across the image. The objects in the scene are found growing smaller planar patches at known surface points until they meet and merge or break at discontinuities (which are found by fitting a bipartite patch at various orientations). Finally, a piecewise smooth surface is derived.

Chapter 3

Problem Definition

The goal of this research is to provide an alternative to active methods of recovering dense range images (such as laser or acoustic rangefinders) with sufficient accuracy to allow the extraction of surface discontinuities and the labelling of occluded regions. There are several reasons for this: passive methods may be applied after the fact to images gathered for other purposes and there exist applications where the use of active methods is unwise. But, the most important reason is that humans, and other animals, are able to perceive passive stereo scenes and therefore some method must exist.

At USC, a parts identification system was developed by T.-J. Fan *et al.* [27] which uses a laser rangefinder to obtain data and generates a feature description of a scene which is matched to a parts database. A reasonable application of this research, then, would be to replace the current active ranging for the initial acquisition, matching, and surface analysis. To this end, analysis was first done on feature data, but there was a problem in separating the foreground from the background.

This led to the observation that all stereo matchers are really feature matchers, the difference is in how the matching operator is applied and that the low-level “features” that the area-based matchers use are matched on a statistical basis and the matching is performed on a regular grid throughout the image, rather than the sparse, one-to-one paring for the more abstract, high-level, feature-based primitives (*e.g.* an edgel or segment). In providing the clean, higher-level feature, some data must be thrown away. Thus, there is a tradeoff between the advantages of these two extremes in terms of density, and quality of information extracted from the scene. Some of these are shown in table 3.1.

In addition, the main difficulty with current stereo correspondence algorithms such as [2, 4, 11, 31, 54, 56, 72, 78], is that they often fail when dealing with added or missing features due to noise, surface discontinuities, and occlusion present in real images. Also, it is often difficult to determine in detail the correct disparity of the matched features in feature-based methods since the same generalization that

Area-Based Methods	Feature-Based Methods
Dense	Sparse
Bad near discontinuities	Good at edges
Slow	Fast
Needs some interpolation	Needs massive interpolation
Less accurate	Very accurate
Exact registration required	Approximate registration OK

Table 3.1: Comparison of Matching Methods.

allows a feature to be represented also introduces some ambiguity in how to provide a point-by-point pairing of the feature in one image with its corresponding feature in a second image.

What then, are the advantages and disadvantages of these two methods? Can they be combined or are they just different ways of using the same information? Table 3.1 shows some of different aspects of these two approaches. Certainly if we use both an area-based and a feature-based process, we should be able to get a dense disparity map and should be able to accurately determine the location of the edges, and thus, have the best of both worlds (if we can combine this information). Speed is a problem, but only on a serial machine — where the feature-based methods have a big advantage of limiting both the number of features and the search-space for the match. But in a parallel system, with the parallelism across the image, this difference may be much reduced or removed, since it is easier to map the area-based methods onto a parallel architecture, and the solution does not require the massive interpolation needed by feature-based approaches.

To provide a context in which our algorithm works, we assume that the input consists of a pair of passively-acquired collinear (or approximately collinear) images, with a fairly narrow stereo angle (from about 5° to about 30°) and the output is a dense disparity map with the surface features (visible/occluded/unknown) and the discontinuities (depth/convex/concave) labelled. In addition, we desire that the system should not guess in regions which are without texture and for those regions in which no stereo matching should be made, such as the cut-off edges of the scene and those areas that are partially occluded. It is important, therefore, to know the location of such areas and to mark them as being unmatchable locally (as opposed to forcing a random match). A list of unresolved hypothesis/problem entries should also be generated to alert higher-level processes of problems with the interpretation which may need the application of more scene specific knowledge such as assumptions of object type to be resolved. In addition, feature abstraction, labelling and matching

information is available so that it is possible to feed back information to the system or for it to serve as additional input to the high-level processing.

We view this system as being an important link between existing low and medium level feature extraction systems and higher level systems such as **ACRONYM** [21] or as an additional channel for a system like **MOSAIC** [35].

3.1 Choice of Primitives and Strategies

As discussed above, the problem that we wish to solve is that of finding an accurate representation for the visible surfaces of a set of objects in a scene, by the analysis of a pair of stereo intensity images of that scene taken by two cameras. For the area-based matching we need a feature which is as invariant as possible to changes in viewpoint from two different angles and for which the relative rather than the absolute values may be used. Since we desire a dense match for the initial disparity estimate, we also need a dense feature. We have chosen a normalized cross-correlation of intensity combined with a confidence measure based on the local variation in intensity. In addition, we require that the correspondences chosen be in agreement for the two views. That is, if the two views do not agree on the location of the point then either one of them is wrong, or the point is visible only from one camera (and therefore most likely also wrong, since there should exist no possible stereo correspondences). The few points for which the two views agree on at an incorrect disparity are discarded through the use of the smoothness assumption, and by imposing an order constraint on the final match.

This **correlation window**, then, is the fundamental area-based primitive. Each voxel, or volume element, in an array defined by the width, height and disparity is assigned the result of a comparison of the regions near the associated central pixels of the two windows as shown in figure 3.1.

The feature-based primitives that we use are **edgels** and **segments**. We choose these features because they are indicative of intensity discontinuities which, typically, occur only at object boundaries, at surface discontinuities and across surface markings (*e.g.* in textured regions).

Edgels are defined as the location and orientation of a step edge in the intensity, located at pixel or sub-pixel accuracy. The orientation is perpendicular to the direction of the maximum change in intensity such that the lighter side is to the right. These serve as the source of our most accurate information about the feature position.

Figure 3.1: Correlation Windows.

Segments are linear approximations of connected sets of edgels which are oriented in the same direction and form lines or curves. Such segments represent the projections of object boundaries, surface markings and other sources of discontinuities in the intensity image. These segments may be used to obtain the feature matches, but usually provide nearly identical information (along the edges) as the area-based matches.

However, we desire to accept and work with real world scenes and therefore we do *not* assume that our data is perfect, rather, we expect missing and extra edges and junctions due to noise and conditions that we assume do not exist in our analysis but which do occasionally occur and produce errors during the low-level and preprocessing steps. We feel that a realistic system must be prepared to detect and correct such errors while at the same time choosing heuristics which serve to quickly and accurately determine the components of the scene.

To make this problem more tractable we make some assumptions and define some constraints which vastly simplify the problem with little loss in generality for real-world scenes.

Assumptions:

Solidity. Any edge bounds two surfaces or is an interior curve (surface marking) on the surface. In addition, sufficient resolution is available so that objects do not degenerate into laminae or wires in the image.

Opacity. Surfaces are opaque: objects located behind, or occluded by a surface, cannot be seen. This allows us to assign a unique surface to a region (or equivalently the range at each point) in the image.

Smoothness. The range, and hence disparity, varies smoothly almost everywhere, thus the edges of surfaces and surface markings vary continuously in depth along their length and this is preserved as the primary geometric invariance between the two images. This allows us to apply a *figural continuity* constraint to verify the matching and to indicate the presence of discontinuities.

Limited Fusion Interval. The fusion interval, the region front of and behind zero-disparity in the image, is limited. This allows us to limit the affects of any aliasing and speed the processing at the expense of the possibility of not being able to match some very close or distant objects.

Small Vergence Angle. The angle between the viewpoints is small. Therefore the changes in geometry from one view to the other is minimal.

Constraints:

Epipolarity. This is derived directly from the geometry of stereoscopic projection from which it follows that a plane passing through the optical centers of the cameras projects to a straight line in each camera's image. Epipolar geometry defines the constraint on the relative positions of a pair of corresponding points in a stereo image pair. Typically this constraint is applied where the stereo image pair is obtained from two cameras whose spatial positions are related by a simple horizontal displacement. The plane defined by a point in space and the center of the two cameras is called the *epipolar plane*, while the pair of lines cut by this plane in the pair of images are called the *epipolar lines* (see Figure 3.2, from Barnard and Fischler [10, page 558]).

Clearly these *epipolar lines* represent the only possible point projections in one image can be matched by a single point in the other image. Each pair of points representing a different distance along the *epipolar rays* extending from the camera centers and through the respective point projections. When the camera image-planes

Figure 3.2: Camera Geometry.

are coplanar, then the epipolar lines are parallel, and when the parallel epipolar lines are aligned, then the configuration is called a “Collinear Epipolar Geometry.”⁴

Order Reversal. Piecewise continuous sections of the scene as seen from one view will have the same left-to-right ordering in the other view. This constraint is useful in removing some noise in the scene and follows from the Solidity and the Limited Fusion Interval assumptions. While order reversals can occur in rare instances, it is reasonable that such reversed matches be thrown out and only the match with the larger context or which is inside the fusion area be considered as a valid match.

Agreement Between Views. By evaluating the best matching both from left-to-right, then again from right-to-left and requiring that only points that support an agreement from each view are retained, we can obtain a much better estimate. In the special case where there are no features within the correlation window, as indicated by a very low value in the local variance measure, a value of zero is assigned. This is strongly supported by the Small Angle assumption. Other systems have used left-to-right processing followed by right-to-left processing to provide a check on the match. However we desire to use it as a constraint at a fundamental level prior to the actual matching. The advantage of using two views along with the enforcement of opacity and uniqueness assumptions has been discussed before by Drumheller and Poggio [26], and Yuille and Poggio [81]. However unlike these approaches, we allow multiple matches along the line-of-sight by accepting a weaker level of agreement between matches. This allows us to consider surface smoothness rather than surface flatness since we may have a one-to-many relationship between pixels when the surface slopes sharply away from the camera.

Minimum Patch Size. Small isolated patches are removed. If a match has very little support, then it is not even considered. This follows partly from the Smoothness assumption, and is a necessary step in removing noise caused by various factors throughout the system.

Unique Match at Each Point. The Opacity assumption directly implies that each point will have only one match. This constraint

⁴See Baker *et al.* [8, pages 330–333] for a detailed discussion of epipolar geometry.

is very useful because it simplifies the processing and the associated data structures, however we often desire that a many-to-one relationship exists so that a steep gradient may be cleanly defined. We can use this constraint and allow a many-to-one relationship by maintaining both a left and a right view.

Chapter 4

Description of the Method

The following sections present a detailed explanation of the processing performed on each image pair. To clarify this processing, the results at each step are illustrated using the Renault part shown in figure 4.1.

4.1 Overview of Methodology

The methodology used in this research can be understood from the analysis of a slice through the disparity space formed by the cross-correlation in which each pixel along a row of the left image is compared with all of the pixels of the right image, within the band of disparities that are defined by the fusion interval. Figure 4.2 depicts such a cross-correlation. The slice through the left and right images form the “axes” of a square image with the correlation data distributed as narrow band along a diagonal. Here each grid point represents the result of correlating the values in a window centered at a pixel from the left slice with a similar group of values from a window centered at a pixel from the right slice, according to the correlation formula:

$$C_{x,y,d} = \frac{1}{1 - \frac{\left(\sum \text{Left}_{x,y} \text{Right}_{x+d,y}\right)^2}{\sum \text{Left}_{x,y}^2 \sum \text{Right}_{x+d,y}^2}}$$

But this representation is difficult to work with for two reasons: First, as a data structure, there is a lot of wasted space since the disparity is usually an order-of-magnitude smaller than the x and y dimensions of the images. The second reason is conceptual, since the actual geometry of the viewpoints (cameras) is aligned at 10° – 15° to each other it is easier to visualize the matching when the data structure also has a narrow angle between the viewpoints. The cross-correlation represents the viewpoints at right angles. The compressed cross-correlation format that we use

Figure 4.2: Compressed Cross Correlation Array.

is a compromise which provides a closer-to-normal 45° between the viewpoints and wastes very little storage space. This is accomplished by sliding slices from the filled portion of the array such that the zero-disparity (dashed) line is horizontal.

Figure 4.3(a) shows such a compressed view of the cross-correlation of a pair of corresponding rows of the Renault Part image (see figure 4.1). The peaks of the cross-correlation (bright areas) represent the best matches. Figure 4.3(b) shows the extracted peaks which have been overlaid with the matched (solid) and unmatched (dashed) edgels (see figure 4.16).⁵

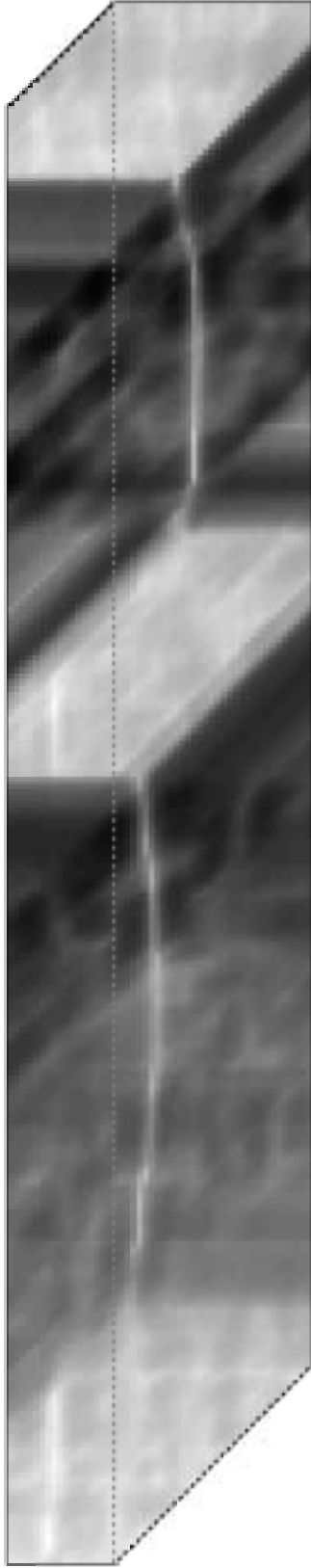
Figures 4.3(a) and (b) show the line-of-sight from the two images such that the left image line-of-sight, or view, is oriented vertically, while the right image view crosses it at a 45° angle. The horizontal dashed line through the center corresponds to zero disparity (that is, the point at which the crossing lines of view are from the same column of each image) and the points are plotted with negative disparity (points closer to the camera) below the zero line and positive disparity (points farther from the camera) above the zero disparity line.

By looking at figures 4.3(a) and (b) it should be clear that, for most points, the correct disparity can be found by simply extracting the peaks in the array. Once this is done, we need to focus our attention on the problem areas only. The first occurs in areas without measurable texture. For these, it is important *not* to generate matches from the cross-correlation information. Where multiple matches (peaks) occur, we prefer the highest peak (best correlation) on which both views can agree. It is possible (although unlikely) for the correct peak not to be selected, so we also must impose a smoothness constraint on the surfaces, that is, the peaks should form a piecewise continuous “ridge-line.” Also, we assume that order reversals due to the spatial shift between viewpoints cannot occur within the relatively narrow fusion interval. An example of such a reversal is the two small ridges in the middle of Figure 4.3(b). The remaining incorrect areas represent points that are either occluded in one of the views or “visible” points on which the two views cannot agree. These areas may be removed by using the results from multiple resolutions to help to select the longer ridges in preference of locally higher ones. Finally, there is the “blurring” beyond the actual edge, which can be seen in figure 4.3(b) where the peak-ridges extend beyond the matched edgels (the solid vertical and 45° lines) that represent the object boundaries.

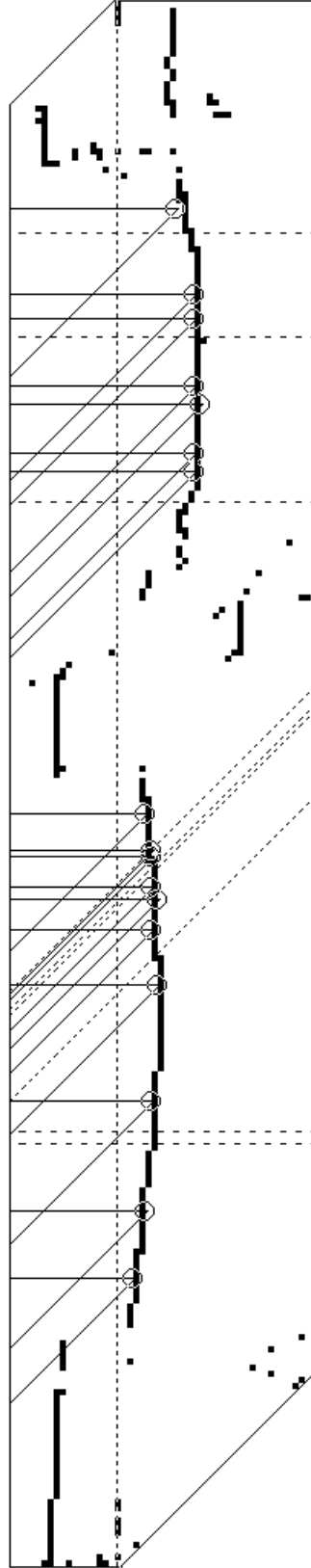
Figure 4.4 shows the overall control flow used in our Stereo Vision System.

- First, the original images are adjusted so that their corresponding epipolar lines lie along corresponding rasters.

⁵The edges were matched using Medioni’s segment matcher [54].



(a) Left View of the Cross-Correlation.



(b) Correlation Peaks and Edge Matches.

Figure 4.3: Cross-Correlation Slice (Row 188 of the Renault Pair in figure 4.1). The corresponding edge matches from the same row in figure 4.16 are shown in (b).

Figure 4.4: Stereo Vision System Flow Diagram.

- Next the resulting images are reduced by a convolution with a Gaussian and subsampled to form a pyramid of (usually three) image pairs. Each pair of slices from this stereo pyramid is separately processed starting with the coarsest (most reduced) pair.
- The initial feature-based and area-based processes proceed independently to produce a set of edge-features and a dense disparity estimate.
- These are then combined to produce a dense disparity estimate with less blurring. This corrected estimate is used to improve the matching in the successively larger slices and is used to extract the surface features: specifically labelling all of the points as being visible or not, and marking the depth discontinuity contours. The discrete disparity surface can then be smoothed, and unknown areas may be supplied by interpolation, in order to extract the orientation discontinuities.

The entire processing at a given resolution scale, which we illustrate on the Renault part of figure 4.1, is therefore as follows: First we generate the cross-correlation volume for the entire image. Then we extract peaks, subject to agreement from both views. This provides us with the correct matching at most textured points that are visible from both views. We then apply a smoothness assumption and the ordering constraint and remove the conflicting “matches.” Next we attempt to fill the gaps in the image from the lesser peaks in the cross-correlation and repeat our checks until we cannot improve the matching. Finally we extract the edgels from the original intensity image and use them as (monocular) cues as to the location of possible depth discontinuities during an adaptive smoothing process which trims surfaces that have overrun beyond the object boundary. This gives us a very good set of matches, which we can further improve by labelling all of the points as being visible from both views or not, and marking the depth discontinuity contours. The discrete disparity surface can then be smoothed in order to extract the orientation discontinuities. The following subsections present a detailed account of these steps.

4.2 Early Processing

The early processing produces the organization of the data that is expected from the low-level vision routines. It includes the initial acquisition of the stereo image pair, the processing to adjust the epipolar alignment to provide the remainder of the program with a collinear epipolar geometry (which greatly simplifies the processing) and this initial viewpoint is cropped to center the relative disparity about the chosen center depth of the image. The last step provides the vergence of the two views about

which the stereo fusion interval is set. This initial preparation of the data is partly automatic and partly performed by hand.

4.2.1 Image Acquisition

We can use any stereo image pair generated with approximate collinear epipolar geometry or to which the camera transform has been applied to effect this alignment. Within our lab, we use a $480 \times 512 \times 8$ bit image. We use one camera to minimize the error due to differences between cameras and shift that camera between viewpoints using a NEAT Linear Table. This positions the camera to the two viewpoints as close as possible so that the epipolar lines are aligned with the scanlines. This minimizes the distortion caused by the registration process which aligns the epipolar lines with the scan lines to effect the collinear epipolar geometry. Next, we acquire the images with a Sony Model XC-38 CCD Video Camera in conjunction with a Matrox MVP-AT Video Digitizer installed on an IBM PC/AT.

4.2.2 Epipolar Alignment

To greatly simplify the processing of the data, and with no loss of generality, we transform the images so that their epipolar lines lie along the image rasters to form a collinear epipolar geometry [7]. This can be done using the camera transform which in turn can be derived from just a few matched points in the image [30], or when the alignment is nearly correct a small linear correction may be applied. We chose the latter approach because it is a fast and simple linear process that produces good results. This transform uses four matched pairs of points $(A_1 : A_2 \cdots D_1 : D_2)$ near the corners of the images as shown in figure 4.5(a) and (b). Each image point (X, Y) in one of the images may be interpolated to match its corresponding point in the other image by shifting up or down each column according to the following equation, where the points P_1 and P_2 are the intersection of the X -th column with the lines $\overline{A_1 B_1}$ and $\overline{C_1 D_1}$ in the image to be warped (figure 4.5(a)) and the points P_3 and P_4 are the intersection of the X -th column with similar lines in the new image (figure 4.5(c)) where the points A' , B' , C' , and D' have the same X -th coordinate as the A_1 , B_1 , C_1 , and D_1 and the Y coordinate is taken from the matching points A_2 , B_2 , C_2 , and D_2 .

Figure 4.5: Epipolar Adjustment.

$$\begin{aligned}
X' &= X \\
Y' &= (Y - P_{1y}) \frac{P_{4y} - P_{3y}}{P_{2y} - P_{1y}} + P_{3y} \\
&= \left(Y - \frac{(D_{1y} - C_{1y})(X - C_{1x})}{D_{1x} - C_{1x}} - C_{1y} \right) \\
&\quad \times \frac{\left(\frac{(B_{2y} - A_{2y})(X - A_{1x})}{B_{1x} - A_{1x}} + A_{2y} \right) - \left(\frac{(D_{2y} - C_{2y})(X - C_{1x})}{D_{1x} - C_{1x}} + C_{2y} \right)}{\left(\frac{(B_{1y} - A_{1y})(X - A_{1x})}{B_{1x} - A_{1x}} + A_{1y} \right) - \left(\frac{(D_{1y} - C_{1y})(X - C_{1x})}{D_{1x} - C_{1x}} + C_{1y} \right)} \\
&\quad + \frac{(D_{2y} - C_{2y})(X - C_{1x})}{D_{1x} - C_{1x}} + C_{2y}
\end{aligned}$$

One of the images may be adjusted to match the other or each image may be adjusted halfway, using

$$\begin{aligned}
A' &= \left(A_{1x}, \frac{A_{2y} - A_{1y}}{2} + A_{1y} \right) \\
B' &= \left(B_{1x}, \frac{B_{2y} - B_{1y}}{2} + B_{1y} \right) \\
C' &= \left(C_{1x}, \frac{C_{2y} - C_{1y}}{2} + C_{1y} \right) \\
D' &= \left(D_{1x}, \frac{D_{2y} - D_{1y}}{2} + D_{1y} \right)
\end{aligned}$$

in figure 4.5(c) so as to equally distribute any error introduced by this transform. First one, say the left image, is warped halfway to match the right, and then the right image is warped halfway to match the the original left.

Figure 4.6 shows the Renault part stereo pair after the epipolar adjustment. The image is $251 \times 256 \times 8$ and the disparity ranges from -30 (near) to 15 (far) pixels. The left and right images have been adjusted to bring the corresponding image points into alignment on the same scan lines. The images were then cropped to remove a few lines at the top and bottom for which some image points were not available.

4.3 Area-Based Processing

The area-based processing matches a measure of the local texture in the image pair to produce a dense disparity map. The input to the area-based process is a stereo



Figure 4.6: Renault Part: Collinear Intensity Image Pair.

pair of intensity images and an optional pair of disparity estimates from a prior pass. The overall process is shown in figure 4.7. First a measurement, which we call the local variation, is made of the image texture. This is similar to an inverted auto-correlation and returns a small value when there is little or no matchable texture. It is used during the estimate phase to inhibit the initial stereo matching. Next a normalized cross-correlation of the two intensity images is calculated, which yields a maximum value at the best matches. The peaks are then extracted from this cross-correlation.

When there is an estimate of the disparity from the processing of a prior level of the image pyramid, then that estimate must be re-scaled to the current level by doubling the image size and the disparity values. Now, the initial disparity estimate for the current level can be made. Here the strongest peak (disparity) is selected at each location from both the left and right viewpoint; subject to there being enough local texture as indicated by the local variation, and optionally guided by a prior estimate.

The last part of the area-based processing is a loop which applies a set of constraints to the matched points and performs a surface interpolation guided by the cross-correlation peaks.

Figure 4.7: Area-Based Processing Flow Diagram.

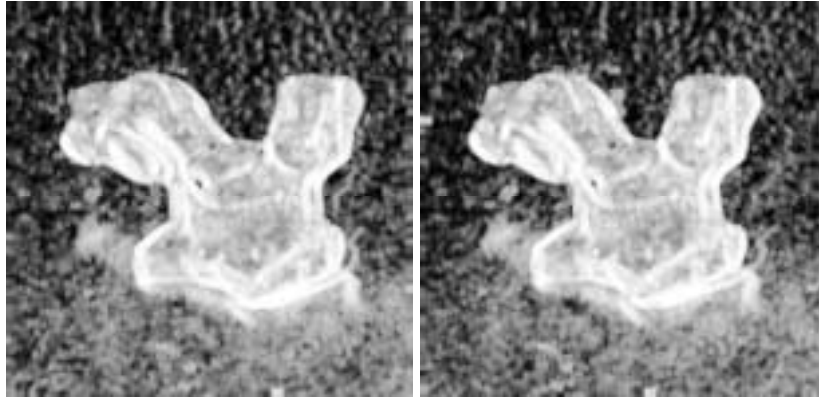


Figure 4.8: Renault Part — Local Variation.

4.3.1 Local Variation Estimate

The first step is to determine where there is significant texture and where there is not. This is done by calculating a measure of the matchable texture which we call the “local variation.” This measure is given by the formula:

$$V_{x,y} = \sum_{j=-\frac{m}{2}}^{\frac{m}{2}} 1 - \frac{\bar{x}_j}{\sqrt{(\frac{n-1}{n})\sigma_j^2 + \bar{x}_j^2}}$$

where \bar{x}_j and σ_j are calculated along each row of an $n \times m$ window. $V_{x,y}$ becomes a minimum of zero when all of the values in each row of the window are the same, and reaches a maximum when $\sum_j \sigma_j$ is large — especially when the intensity is low. The reason that the value is summed over the raster rather than within the window is that a horizontal stripped pattern is not matchable and must be treated as a textureless region. Figure 4.8 shows an enhanced image of this function applied to the Renault Part intensity images. Note that the abrupt changes in intensity cause high values and that the apparently textureless background does contain matchable texture as can be seen by the similarities in the variation images. When this measure drops below 10^{-6} for a full resolution image, the response is that the region is considered unmatchable (*e.g.* unknown). If the area is very small (up to about six pixels), the interpolation will fill in a reasonable value, otherwise we consider a result of “unknown” to be the correct value for these textureless regions as opposed to “guessing” a value (which is a job for higher-level processing).

4.3.2 Correlation

The correlation volume is generated using a modified normalized cross-correlation bounded by manually supplied minimum and maximum disparities. In the human visual system, this value is fixed and is called Panum’s fusional area. In a parallel architecture with a separate set of processors at each disparity level, this would also be fixed. However, for a serial machine, it speeds the processing if this value is settable by the user close to the minimum necessary size. As we will later show, if the correct match lies outside the supplied fusion area, then the points that cannot be matched are marked as “unknown.” In order to add the out-of-fusion-interval parts of the image, the vergence must be changed, and a world-model built up with multiple re-evaluations of an image. This is an aspect of a more complete vision system [3, 24], and not pertinent to this study.

The correlation is given by the formula:

$$C_{x,y,d} = \frac{1}{1 - \frac{\left(\sum L_{x,y} R_{x+d,y}\right)^2}{\sum L_{x,y}^2 \sum R_{x+d,y}^2}}$$

where the summations are over a local window about the left (L) and right (R) points indicated and the images are padded with a reflection of their values near the edges. We use a 9×9 correlation window which seems to be a good compromise between a larger window, for a unique match, and a smaller one to minimize blurring. This correlation yields a peak (maximum) at the best matches. If an exact match occurs, the denominator becomes zero and the result is set to the largest single float value, which indicates a very good match.

Figure 4.9 shows a slice through this cross-correlation volume at row 188 from the bottom which cuts through the two lobes of the Renault Part as highlighted in figure 4.1.

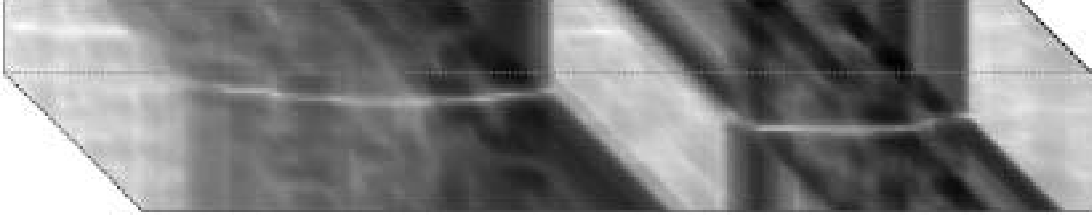


Figure 4.9: Left View of the Cross-Correlation (Row 188 of the Renault Pair in figure 4.1).

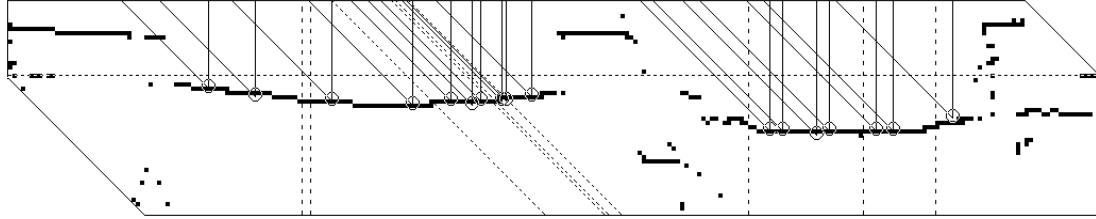


Figure 4.10: Left View of the Correlation Peaks and Edge Matches.

4.3.3 Peak Extraction

Figure 4.10 shows the peaks selected from the cross-correlation slice Figure 4.9. The peaks are defined to be the set of points $P_{x,y,d}$ such that:

$$P_{x,y,d} \left. \begin{array}{l} \geq P_{x,y,d+1} \\ \geq P_{x,y,d-1} \\ \geq P_{x-1,y,d+1} \\ \geq P_{x+1,y,d-1} \\ \geq \frac{\max_i P_{x,y,d+i}}{2} \\ \geq \frac{\max_i P_{x-i,y,d+i}}{2} \end{array} \right\} \begin{array}{l} \geq \text{its 4-connected neighbors in a} \\ \text{row slice through the uncompressed} \\ \text{correlation volume.} \\ \\ \geq \text{one-half the strongest peak in} \\ \text{each direction along the views} \\ \text{through } (x, y). \end{array}$$

The first four inequalities are the real definition of a peak. In this case a peak means that the correlation centered at a given pair of pixels must be greater or equal to that of the values centered at a pixel to either side (along the epipolar row) of the given one in either of the two images. The reason that equality is allowed in this definition of “peak” is that the peak may extend for several pixels and no one pixel would meet the strict definition. The last two inequalities limit the search space by removing any peaks which are relatively weak, which we have arbitrarily defined as one-half the value of the strongest peak. This restriction is applied along each view.

Figure 4.11: Estimate Selection.

4.3.4 Multi-level Disparity Estimate

When there is a disparity estimate generated from a prior pass at a coarser resolution in the image pyramid, then this estimate is expanded by zooming all three axis by a factor of two. This estimate is used to guide the selection of the initial disparity and during the interpolation in order to provide a preference for a more global smoothness over the local strength of the correlation. Internally there is a small complication, in order to keep the disparity values as unsigned integers an offset is added. The zero disparity value is given by some value, say offset_i . Since each level of the pyramid may have disparity ranges that are not quite double that of the prior level, the new offset may not be twice that of the estimate being re-scaled. In order to assure that the correct values are assigned, the old offset is first removed, the disparity value doubled, and the new offset (offset_{i+1}) is added.

$$d_{i+1} = (d_i - \text{offset}_i) \times 2 + \text{offset}_{i+1}$$

Some clipping may be performed if the estimated values are outside the fusion interval of the current level of the pyramid. Any values that lie outside of this region are set to the null value.

4.3.5 Initial Disparity Estimate

The initial estimate of the disparity at the current level incorporates the selected peaks along with the local variation and any prior disparity surface estimate. It is generated by selecting the maximum peak for each pixel of both the left and right images, subject to the following restrictions: First, if there is an estimated disparity value from a lower-resolution pass over these images, then the maximum peak within a range about the estimate is selected. If there is a nearby discontinuity in the estimate then the maximum is chosen from multiple ranges as shown by the shaded areas in figure 4.11. The lines show the lines-of-sight for two pixels in each

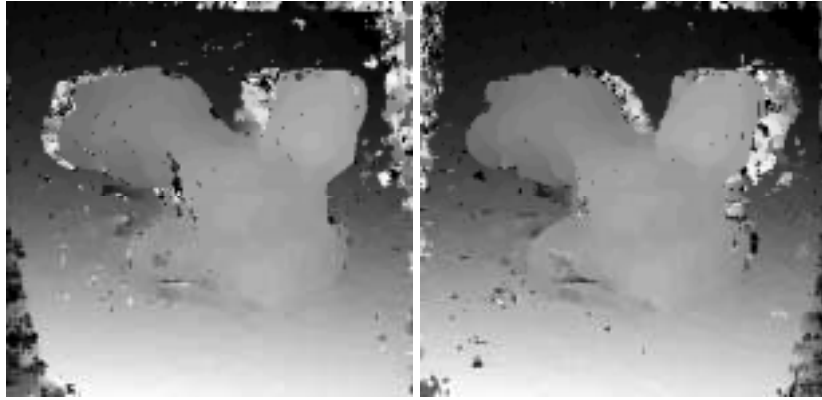


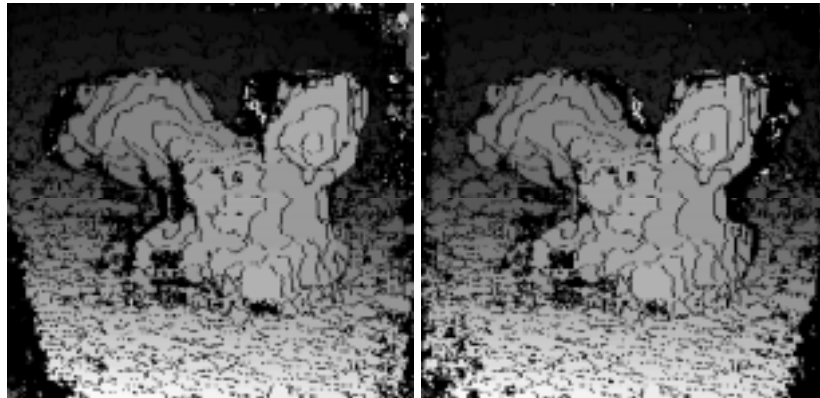
Figure 4.12: Initial Disparity Estimate Derived by Selecting the Best Peaks of the Cross-Correlation of the Images in Figure 4.1.

view. Second, no disparity estimate is made at those points which have a low local variation.

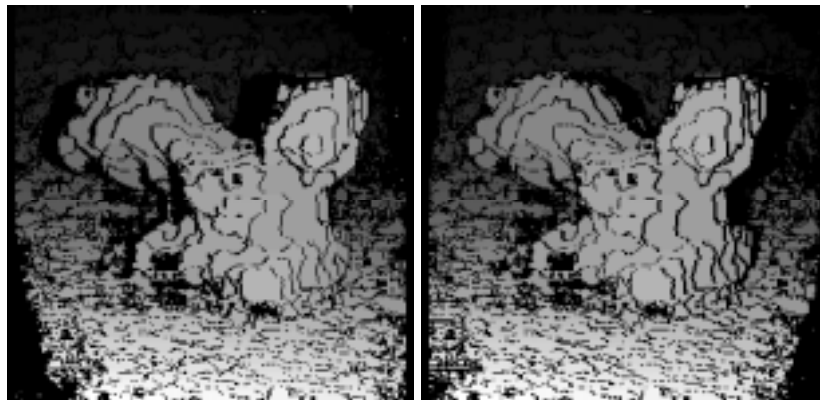
This raw estimate contains a number of incorrect matches, which are apparent in figure 4.12. However, most of these errors occur at the occlusions and the off-the-edge positions where no correct match is possible. The remainder of the incorrect matches are due to noise, photometric variations, and geometric distortions. The next portions of the area-based processing attempt to remove the incorrect matches and substitute the correct ones.

4.3.6 Constraints

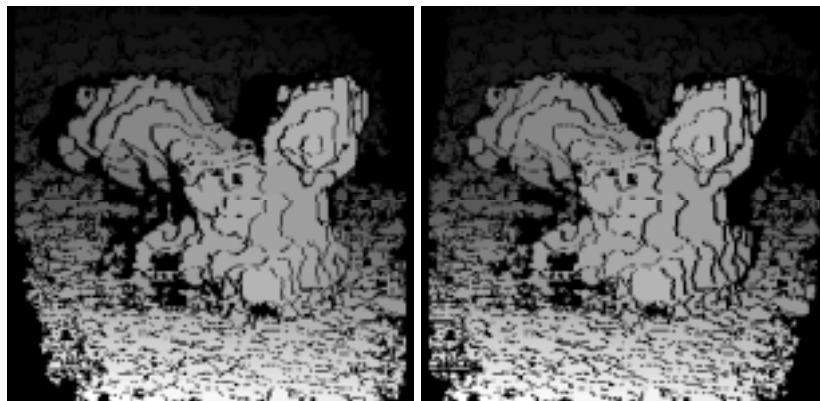
Most of the errors can be removed by the application of three constraints. The effect of this is shown in figure 4.13. The first is to make use of the fact that we have two views of the same scene. If the initial estimates do not agree, then we can safely assume that either one or both estimates are incorrect. At first this agreement on matched points must be exact, but during the later interpolation, the constraint is relaxed to allow a disagreement of from 2 to 4 pixels so that the disparity surface is allowed to slope and is not forced to be flat and perpendicular to the camera axis. Figure 4.13(a) shows the disparity points for which the estimates from each view are in agreement. The second constraint forces the removal of small patches whose order from left-to-right is reversed from view to the other. This is reasonable since we assume that there is a limited fusion interval about the selected vergence, and it is very unlikely for small wire-like objects to be strung in front of the scene. The



(a) After Checking for Agreement of Both Views.



(b) After Removing Order Reversals.



(c) After Removing Isolated Pixels.

Figure 4.13: Application of Constraints to the raw estimate in Figure 4.12.

effect of this constraint is shown in figure 4.13(b). The last step in removing noise is that all isolated pixels are deleted.

Isolated pixels are those pixels that have a disparity which differs more than 2.5 pixels from the average value in the surrounding 5×5 neighborhood. Pixels with fewer than 6 (of its 24) neighbors are also removed.

Now the estimate, shown in figure 4.13(c), is reduced to those points which have a high probability of being correct. Any erroneous matches that remain at this point have sufficient local support to be retained (see figure 5.7 for an example), so it is worth while to remove a few good points in this process since they are usually replaced by the interpolation in the next step. Even if they are not, the ratio of incorrect to correct matches is increased during this phase of the processing.

4.3.7 Interpolation

The last step of the area-based processing is to interpolate through the unknown areas. This is done by starting with the remaining estimate points and, using them as “seeds,” expanding the disparity estimate along the disparity surface formed by adjacent peaks. A cross section of these surface patches is visible as the lines of peaks in figure 4.10. The interpolation proceeds in three steps and after each of these steps, the constraints are re-applied. Correlation peaks that are adjacent to the existing “seed” matches are added to the surfaces according to the following rules. If more than one adjacent match is possible, then the strongest is used, if two or three are equal, then they are selected according to the ordering: same disparity, closer, farther. During the first interpolation pass, only exact matches of peaks between the two views are allowed. That is, only peaks which exist in both views with the same disparity. After 6 cycles, the resulting surface is shown in figure 4.14(a).

During the second pass, a small difference in disparity is allowed so that surfaces that are changing rapidly may be represented by a one-to-many match between the two views. This means that the selected matching points from the two views may differ by some small amount. This threshold represents the amount of gradient that is allowed for a smooth surface (rather than restricting the matches to form only flat surface patches perpendicular to the camera axis) as is illustrated in Figure 4.15. Where the left view’s visible points are labelled with letters, the right with numbers, and the matches are listed below the labels. This surface cannot be represented by either single view with only one disparity value for each cell, but two views can adequately model it. We always use a delta of ± 2 , for the coarsest level of the pyramid and allow a bit more for each finer level. Figure 4.14(b) shows the disparity surface at the finest level for the Renault part after this step is performed for 6 cycles with a delta of ± 4 .



(a) After the First, Strict Agreement, Pass.



(b) After the Second, Weak Agreement, Pass.



(c) After the Final, Median Filter, Pass.

Figure 4.14: Application of Interpolation Passes to the “seed” estimate in Figure 4.13(c).

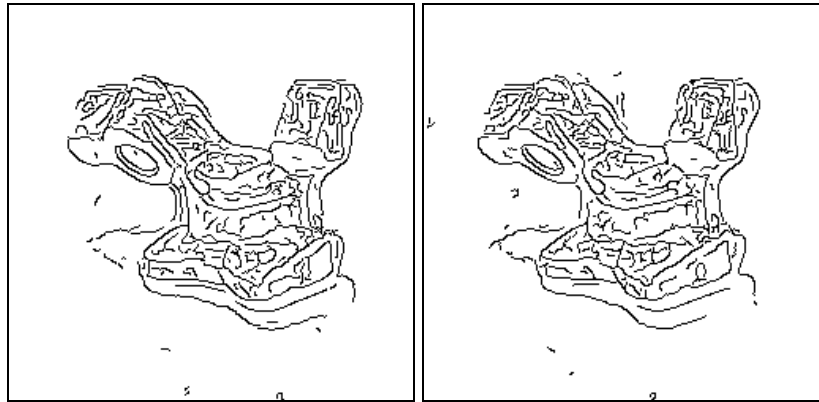
Figure 4.15: Two Views Define a Smooth Surface.

Finally, small holes of up to 6 pixels are filled with a median value. This is because the definition of peaks is quite restrictive and, therefore, holes exist in the disparity surfaces formed by the peaks. A median filter works quite well for this, although a better approach may be to re-evaluate the local peaks in the cross-correlation volume. Figure 4.14(c) shows the final disparity surface.

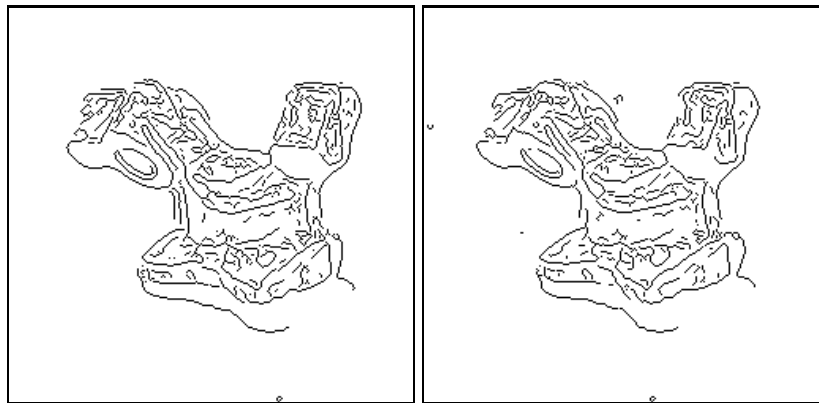
The remaining unknown points are considered to be either occluded or to have insufficient texture to be matched, depending on the associated value of the local-variation image. Figure 4.14(c) represents an impressive result for an area-based process, but does not have the accuracy that we desire due to the tendency to “blur” beyond the object contours (the more strongly textured surface “leaks” into the less textured region).

4.4 Feature-Based Processing

Before we attempt to correct the blurring in the area-based result, we must first extract the edge features. The feature-based processing is performed separately from, and potentially in parallel with, the area-based processing.



(a) LINEAR Edgels.



(b) Canny Edgels.

Figure 4.16: Edgels from Figure 4.6.

4.4.1 Edgel Extraction

We can use any estimate of contour edges. The Stereo Vision System allows a choice of either Canny's Operator [22] or LINEAR [62] edges. We prefer LINEAR because it allows us to extract a reasonable set of edges from a wide domain of images without changing the default parameters. In LINEAR, the edge detection is performed by convolving a given set of stereo images with six 5×5 masks corresponding to ideal step edges in several directions. The maximum of the convolved output at each pixel gives the magnitude of the edge at that pixel, and the direction of the masking giving the maximum output determines the edge direction. Binary edges are obtained by thinning and thresholding the edge magnitudes. These edges are then linked to their neighbors based on proximity and similarity of orientation. Finally the linked boundary segments are approximated by piecewise linear segments and the chains of edgels and their linear approximation are returned.

Figure 4.16(a) shows the result of applying LINEAR to the Renault Part images with a strength threshold of 10.0 (which we use for all of the images), while Figure 4.16(b) shows the result of Canny’s edge finder, using the default thresholds 200 and 400.

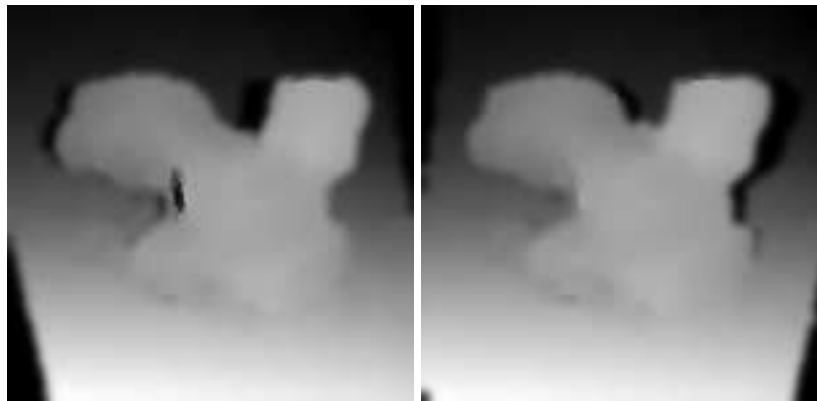
4.5 Integrating Area and Edge Data

The area-based disparity maps define a dense surface which is reasonably accurate, except for a tendency to “blur” beyond the object contours (the more textured surface “leaks” into the less textured region). Since the foreground objects tend to be more textured than the background, the effect of most blurring is to enlarge the foreground.

The area-based match can be refined using the edge information because the edges were located with an oriented mask suited to accurately locating the intensity discontinuities and thus the accurate localization of the discontinuities. We want to demonstrate the importance of the use of the edgels to refine the area-based match; Our approach is to first smooth the disparity map, keeping the disparity at the edgels fixed. All points whose disparity is shifted by more than a constant amount (we use 1.0 pixels) are discarded, removing the blurred fringe around the actual contour edges.

The smoothing is implemented using the adaptive smoothing formalism developed by Saint-Marc *et al.* [71]. Adaptive smoothing is a process which smooths a signal while preserving discontinuities and which can be used to facilitate the detection of discontinuities. This method uses a cascade of convolutions with a small, adaptable kernel which is adjusted at known or detected discontinuities so as to inhibit the propagation of surface information across the discontinuity. We supply the process with “discontinuities” that are really the feature edges. This does not allow any surface information to cross the edges during the smoothing. By locating where the surface is unstable (has little or no local support), we can detect the blurred portion of the area-based match at the expense of some loss of correctly matched points, which are usually replaced during the subsequent interpolation and labeling.

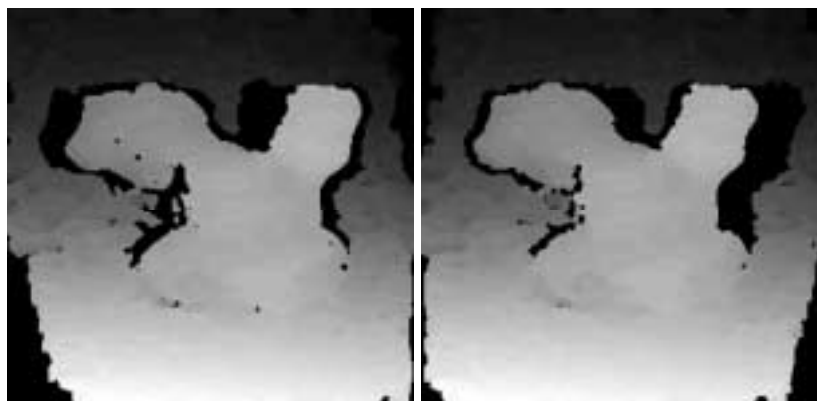
Note that we are using the edgels as monocular cues, so that even the edges which have a similar orientation to the epipolar line directions play an active role. Such edgels are usually discarded in feature-based approaches. Figure 4.17 illustrates the result of this process on the results of the refined disparity map in figure 4.14(c). Figure 4.17(a) shows the smoothed view after 20 iterations of the small kernel with the edges from figure 4.16(a) used as the supplied discontinuities. Figure 4.17(b) shows the absolute difference of the before and after disparity maps. The darker



(a) After Smoothing.



(b) Difference of Non-Smoothed and Smoothed Surfaces.



(c) Disparity Surface with "Unsupported" Regions Removed.

Figure 4.17: Disparity Map After Incorporation of Edge Information.

regions represent those areas with the greatest displacement, and hence, without local support. These areas are removed when they exceed an absolute difference of 1 pixel, yielding the cleaned-up disparity map shown in figure 4.17(c) after two cycles. Other approaches are discussed in chapter 5.

4.6 Visualization of Results

A depth estimate, such as that of figure 4.17 may be used to improve the next finer resolution of the pyramid and, finally, at the end of the chain of pyramid levels, to label points and features in the images.

In the context of a full vision system, the estimated disparity and pixel labeling would be supplied to the next level of processing. Instead, we will use this information to display the results in a several ways for comparison with existing work.

So far we have two types of points, those with a disparity estimate (“known” points) and those without (“unknown” ones). When we have a foreground surface terminated at a feature edge, we can further discriminate between the “unknown” points by examining the geometry of the cameras and the disparity estimate. If we think of the *other* camera as a light, we can imagine of the occluded area as its shadow. Figure 4.18 shows the labeling for the Renault Part. Figure 4.19 shows a cross section depicting the blurred regions that were removed, the visible (but unmatched) surfaces now uncovered and the occluded regions. Those unknown points which are left over may be considered points which should be visible, but were unmatched for various reasons. In figure 4.18, the intensity values mean the following:

1. Known disparity values are shown in white.
2. Points that should be visible, but were not matched, are shown in light gray.
3. Points visible only from this view because of image clipping in the other view (or initial masking) are shown in dark gray.
4. Penumbral points, visible only from this view, are shown in black.

To get a shaded representation of the original figure, we first interpolate along those regions labelled in figure 4.18 as being visible. This interpolated disparity image is shown in figures 4.20. Next, we assign the smoothed surface a simple reflectance function, select a pair of viewpoints and assign a position for a single source of “light.” Figures 4.21 show the shaded images generated for the Renault Part. Most of the error is due to digitization noise and some flattening at limb edges.

Figure 4.22(a) is a 3-D plot of the left-image-disparity data from figure 4.20(b) sampled at every third pixel. The disparity ranges from -15 (far) to +30 (near) and

Figure 4.19: Cross Section of Labelled Surface.



Figure 4.20: Disparity with Interpolated Values for the Unknown Visible Points.

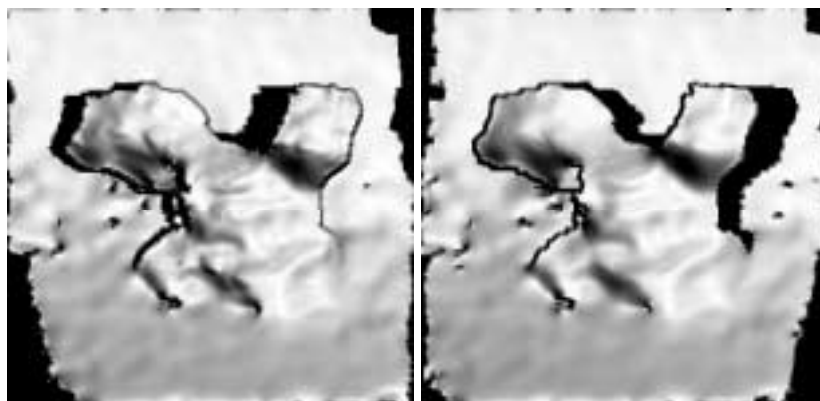
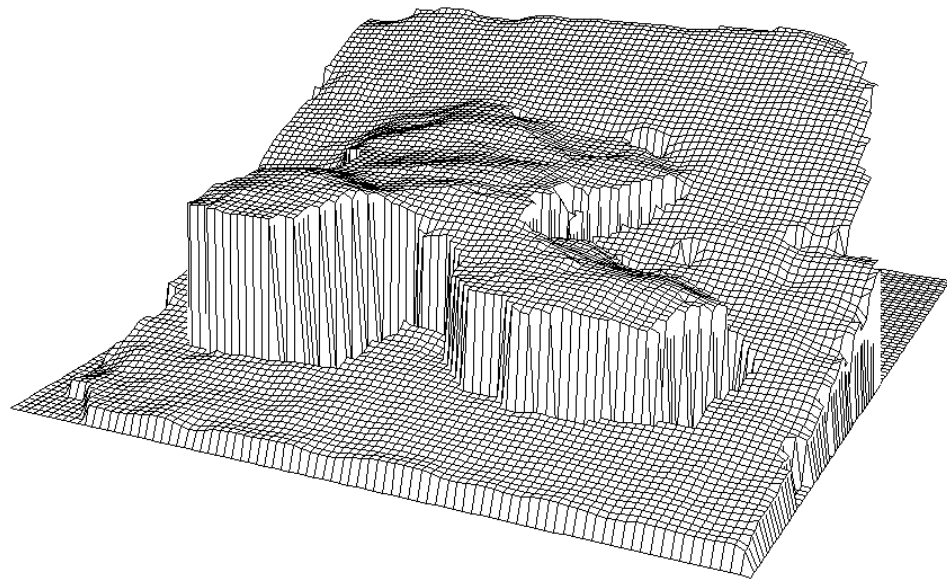


Figure 4.21: Shaded Representation of the Interpolated Estimate.



(a) 3-D Plot of the Final Result.



(b) 3-D Rendering of the Final Result.

Figure 4.22: Reconstruction.

the zero-plane of the plot is offset 17 pixels back from the zero-plane of the image. A rendered view is also shown in figure 4.22 in which the original intensity values (figure 4.6, left image) were projected onto the surface rotated 45° about the Y-axis and scaled by 4 in the original Z-axis.

4.7 Surface Feature Extraction

To provide a qualitative measure of the accuracy of our surface estimate, we will attempt to extract the surface features: The depth discontinuities and both the concave and convex orientation discontinuities. Our goal, then, is not simply to clean up the disparity estimate, but to produce a description of the scene. We would like to be able to segment the scene into surface patches corresponding to the faces of the objects composing the scene, and eventually to group these individual patches into objects in the scene. In order to do this, we must locate the borders or edges of these patches. These borders occur where either the surface breaks (depth discontinuities), or where it creases (orientation discontinuities).

From the prior processing we know the location of the penumbral, visible and off-the-edge areas and we know the resolution of the disparity surface: We can locate the depth-discontinuities by looking for any jumps which are greater than the resolution step-size and which occur between points which are visible or between borders of the penumbral and the visible areas. These discontinuities are shown in figure 4.23.

Locating the creases in the disparity surface is more difficult since the estimated curvature is extremely sensitive to quantization noise [28], which requires us to smooth the local surface as much as possible. Smoothing, however, also removes or attenuates the desired features. We need a process which smoothes noise due to the quantization, and also preserves the surface orientation discontinuities. Since we know that we have the surface data to approximately one pixel resolution, the relative error of any pixel is within ± 1 pixel. Therefore our approach is to first repeatedly smooth the surface with a binomial (or other approximate-Gaussian kernel), but attenuate sharply any change over one-half of the step-size from the original discrete surface. We do this by first smoothing, then masking those regions that were smoothed too much and replacing their values. Finally we smooth some more to remove the artifacts generated by the masking. This gives us a smooth surface with some ripple while preserving the creases where there is sufficient resolution. To finish the process we apply the adaptive smoothing approach proposed by Saint-Marc *et al.* [70] to the derivative of the partially smoothed disparity data. This process preserves the stronger creases while removing the weaker ones. Finally, we calculate the local curvature at each point, and threshold and thin the maximum positive and

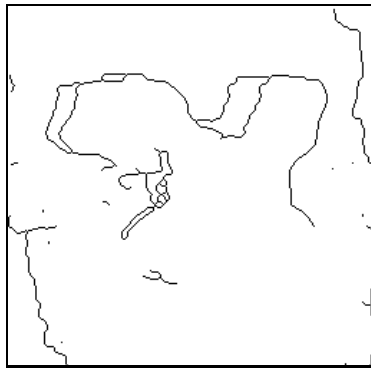


Figure 4.23: Depth Discontinuities from the left view of Figure 4.20.

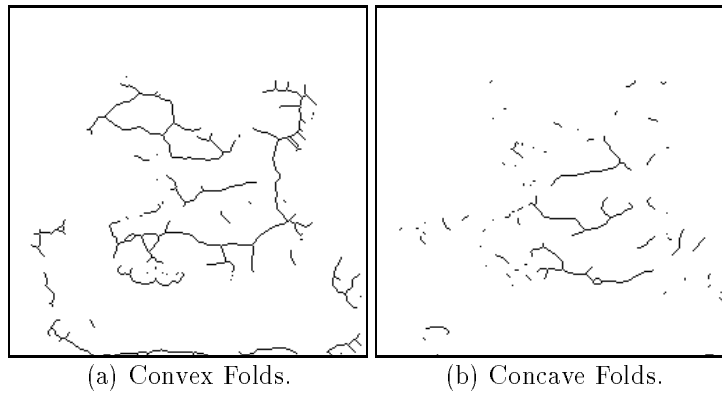


Figure 4.24: Orientation Discontinuities from the left view of Figure 4.20.

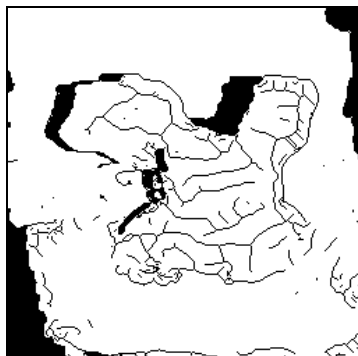
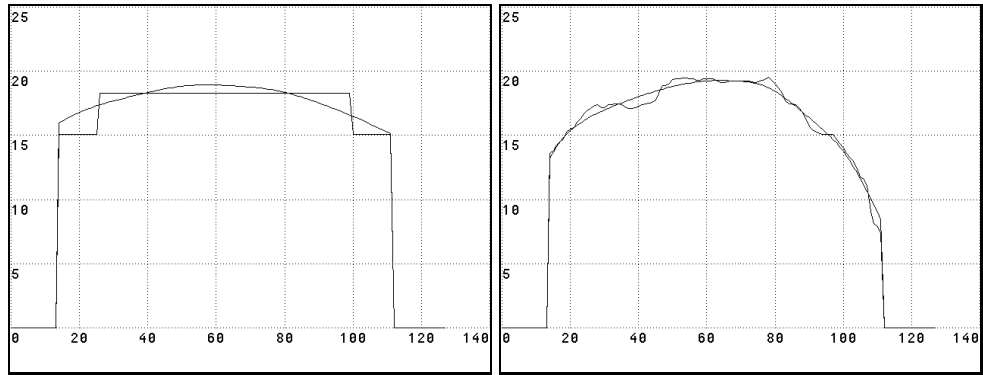


Figure 4.25: Combined Discontinuities from the left view of Figure 4.20.

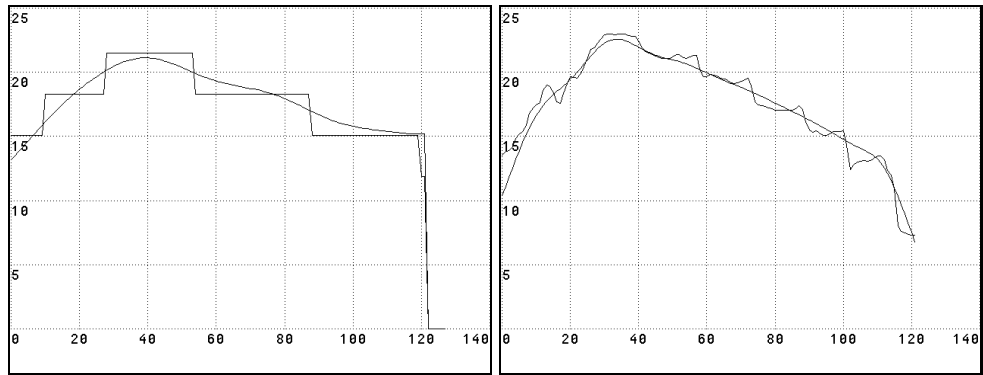
negative curvatures. Figure 4.24 shows the estimated negative (convex) and positive (concave) creases, while figure 4.25 shows the combined depth and orientation discontinuities along with the occluded and off-the-edge regions marked in black.

The poor localization, noise of the creases and, to a lesser extent, the depth discontinuities are due, in large part, to the lack of available resolution. Figures 4.26 (a) and (c) show cross sections through a small section of the Renault image, both before and after the smoothing described above. Figures 4.26 (b) and (d) show the same cross sections through a similar section at finer resolution (this section was cropped from an image at twice the resolution, and then processed separately). Figure 4.27 shows the extracted discontinuities from each image. The low and high resolution images have been adjusted for comparison. The disparity change across the face in the low-resolution image is about 3 pixels, while for the high-resolution the change is around 7 pixels. While the sharp horizontal crease across the bottom of the images is clear at both resolutions, the more rounded vertical crease, to the right of center, was completely distorted at lower resolution.



(a) Horizontal Slice, Row 66 (Low).

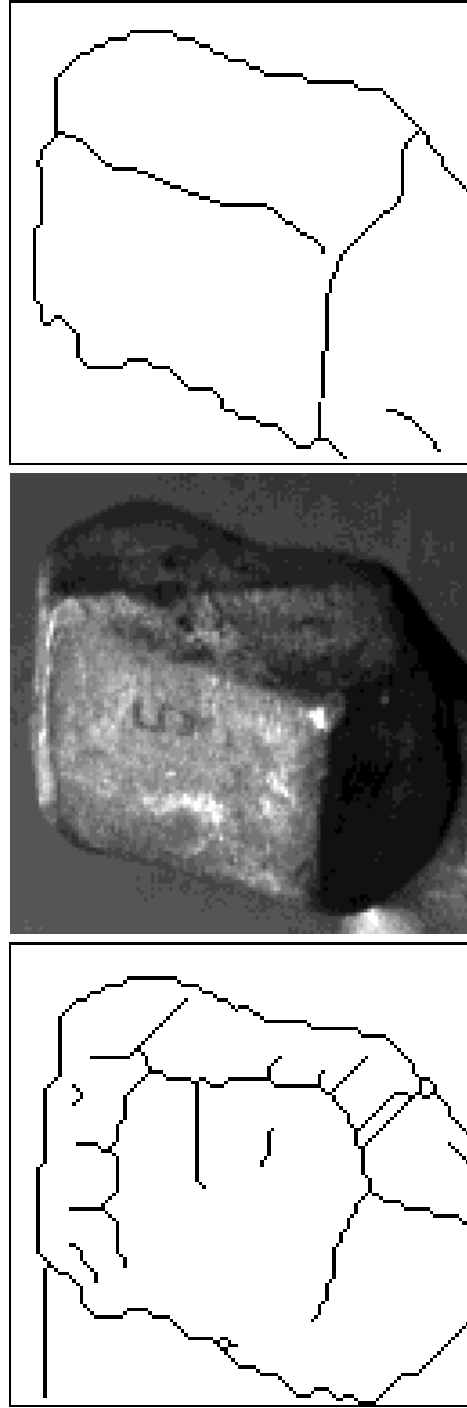
(b) Horizontal Slice, Row 66 (High).



(c) Vertical Slice, Column 60 (Low).

(d) Vertical Slice, Column 60 (High).

Figure 4.26: Slices Through Low and High Resolution Sections from the Renault Image.



(a) Low Resolution. (b) Cropped Intensity Image. (c) High Resolution.
Figure 4.27: Combined Discontinuities from Low and High Resolution Images.

Chapter 5

Observations and Experimental Results

5.1 General Observations

What have we learned from this work? What aspects are important and where should new work be directed?

First of all, we have looked at only one instance of integrating data from two abstractions from a set of data — the correlation window and the edgels. The most important lesson is not to throw out information: Whenever we make an abstraction, we also lose some information. Thus we must use caution when using abstractions and consider what is being lost. Edgels are often used as matchable abstractions since they limit the search space and can provide accurately matched points — however, edges are sparse and when they are parallel to the epipolar lines, they cannot be matched. Such horizontal edgels are typically thrown out, but as we have shown this report, they can be used; if not to match, then to show the location of a discontinuity. The correlation window usually provides a very dense feature but loses the exact location in space provided by the edgel-based methods.

The second thing that we have learned is that the impact of our assumptions should be considered. For instance, we have assumed that the images are in a collinear epipolar geometry — others have also made this assumption. However, if the images are off by only one to four pixels, when using a 9×9 correlation window the resulting correlation can degrade very quickly, as shown in figure 5.1. If we make such an assumption, we must insure that our processing is robust relative to that assumption, or that the assumption is not violated. In the case of the collinear epipolar geometry, we have found that the images need to be very exactly aligned or the matching tends to fail. This adjustment cannot be done by simple raster realignment, since the epipolar lines cross the raster lines at an angle.

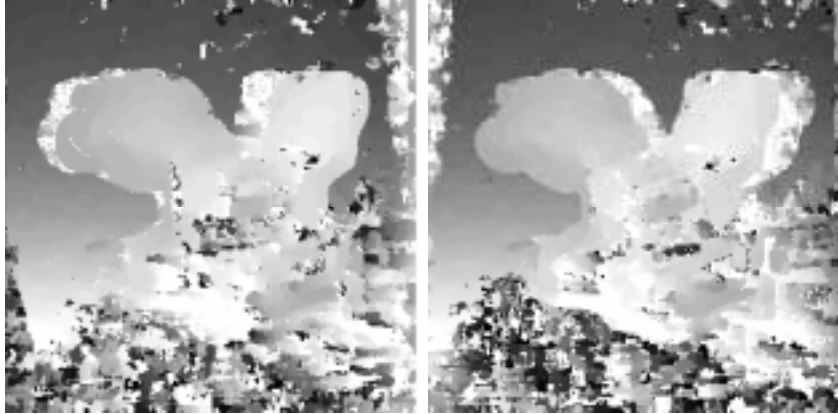


Figure 5.1: Renault Part — Best Peaks Without Collinear Alignment.

The next observation is that although we have used intensity edges to limit the search space for discontinuities, the real features that we should use are the texture boundaries. The best texture edges currently produced, such as those produced by Malik and Perona [46] and Perry and Lowe [66] are not sufficient to guide the boundary localization. For this reason we have simply used the intensity edges as the principal feature.

The most important observation is that there is a real value in combining multiple sources and multiple abstractions of information.

5.2 Problems and Solutions

The following four examples illustrate various aspects of our algorithm including some of the problems that we have encountered and their solution. They include a random-dot pair, an aerial scene and two indoor image pairs.

5.2.1 Wedding Cake (Random Dot Image)

The first example to be shown is an extreme case, a random-dot stereogram “wedding cake” with a disparity change of 8 pixels between planes. The test image, shown in figure 5.2 is $128 \times 128 \times 1$ bits and has an approximately equal number of white and black pixels. This stereo pair shows the area-based processing under the best conditions — where each correlation window has a large amount of information and there is no noise. The disparity range at the finest level is 21, which is ± 8 pixels about zero plus an extra 2 pixels at the extremes. The errors come from two sources: First the corners are rounded, because, with no edge to guide the processing, the cross-correlation smoothes over corners. Second, where there is an occluded region, chance clustering yields mismatches which give rise to the rough vertical boundaries

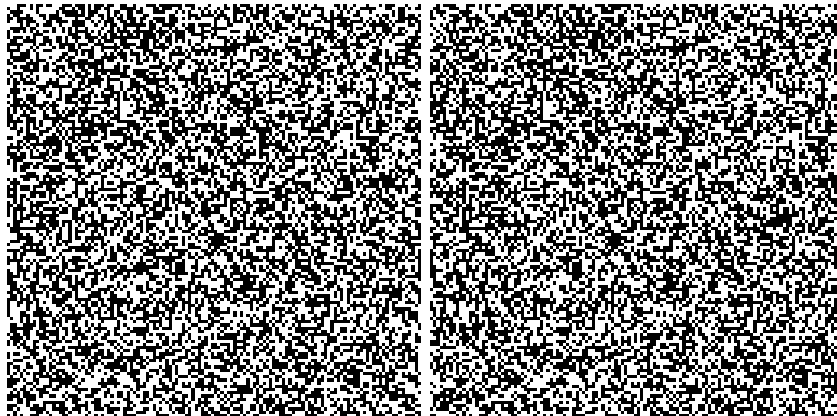


Figure 5.2: Wedding Cake — Original Intensity Images.

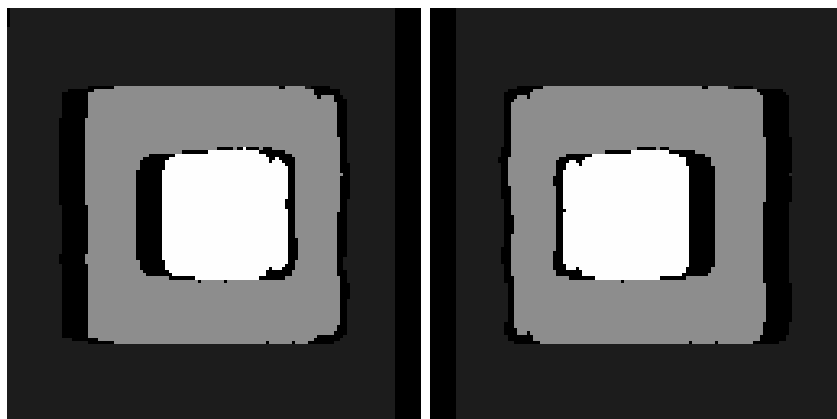


Figure 5.3: Wedding Cake — Disparity Surface.

View	Good Match/Label		Bad Match/Label	
	Binocular	Monocular	Binocular	Monocular
Left	84.64%	11.49%	0.16%	0.62%
Right	84.84%	11.57%	0.19%	0.54%
Total	96.27%		0.76%	

Table 5.1: Wedding Cake — Matching Statistics.

visible in figure 5.3. The most important thing to be noted here is the conservative result which yields only 0.76% of the matched pixels in error and marks the occluded pixels as unknown rather than simply guessing when insufficient information is present.

Table 5.1 shows the correct and incorrect matching statistics at the finest level of the matching process. The first column gives the viewpoint, either “Left” or “Right.” The second column shows the correct matches: The **Binocular** values are for those points which were correctly matched, while the **Monocular** ones are those which were correctly left unmatched. The third column gives the incorrect matches, also broken down into incorrectly matched points which are visible in both views (binocular) and those occluded points which were mistakenly matched with some point (monocular). The total does not add to 100%. That is because the program did not give a match for some points that were visible in both views, which represents an acceptable result, that is neither a correct nor an erroneous value. Therefore, the final results shown here have 96.27% correctly matched points with 0.76% in error and 2.97% unknown.

Figure 5.4 shows the depth discontinuities and the occluded regions of this stereo pair, and figure 5.5 shows a 3-D Plot of the image reconstructed from the matched pixels.

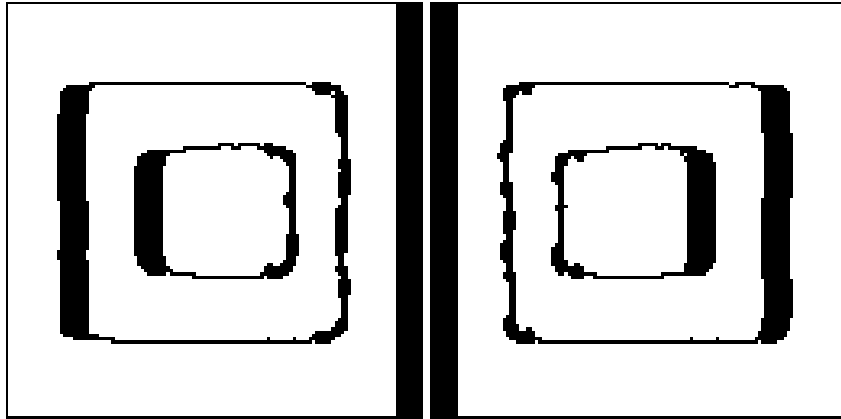


Figure 5.4: Wedding Cake — Depth Discontinuities and Occluded Regions.

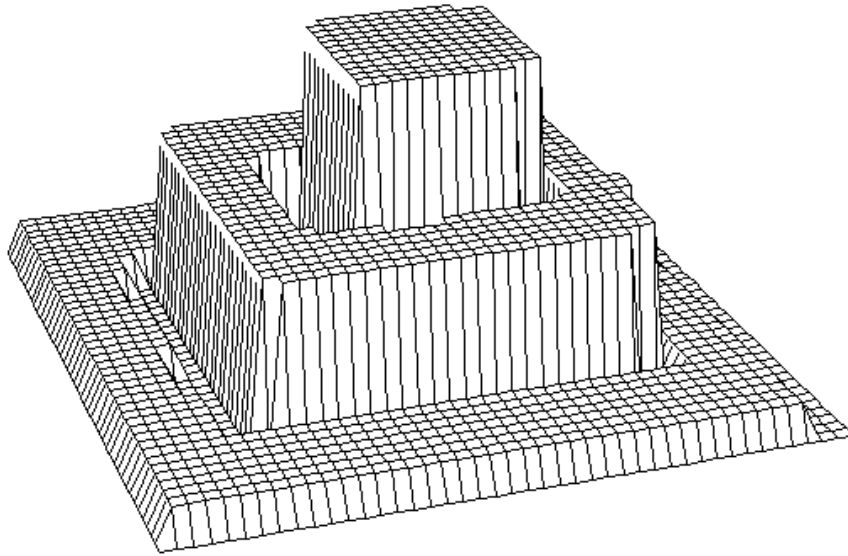


Figure 5.5: Wedding Cake — 3-D Plot of the Integrated Results.

5.2.2 Books

The bottom pair of figure 5.6 shows a $488 \times 488 \times 8$ stereo pair of a stack of books obtained from the University of Illinois, courtesy of Dr. W. Hoff, now with Martin Marietta. The disparity of the stereo pair ranges from 50 pixels (far) to -18 pixels (near). This is the first scene presented in this report for which the multi-level approach yields a different result from the single-level pass at the finest resolution. Without a more global scheme, the local matching has a very strong preference for the wrong match as can be seen in the resulting disparity surface using only the finest intensity level (figure 5.7). The cause for this is geometric and photometric distortion which makes the incorrect match a *better* one. To overcome this we use the multi-level pyramid as shown in figures 5.6 to get a better approximation to the correct surface.

Figure 5.8 shows the final disparity views, and Figure 5.9 shows the surface features. Finally, Figures 5.10 and 5.11 show the 3-D views of the disparity and the rendered scene respectively.

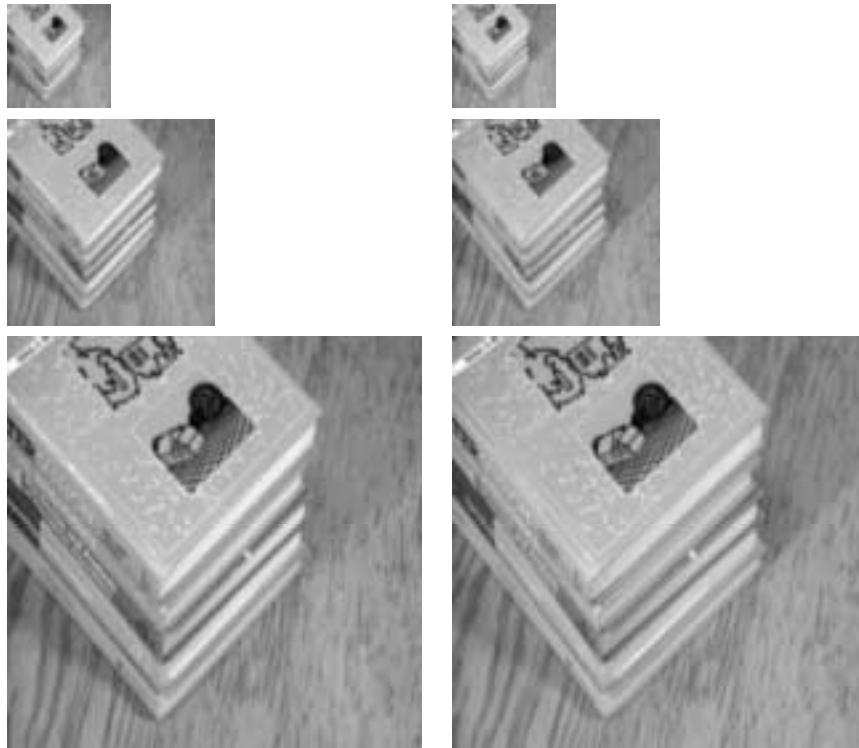


Figure 5.6: Books — Original Intensity Images.

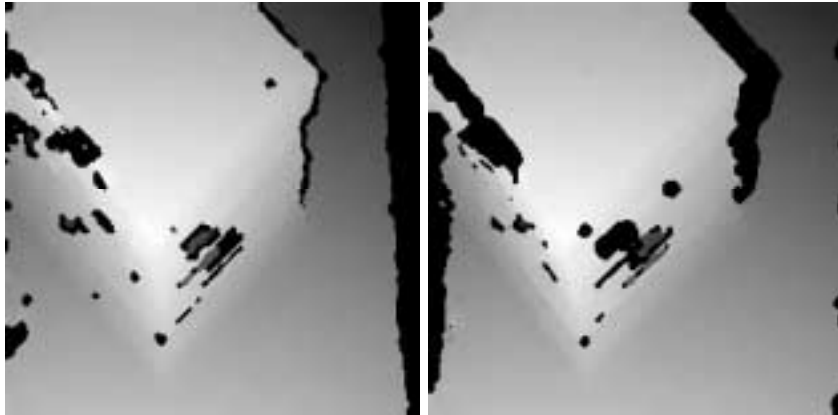


Figure 5.7: Books — Disparity Surface with only the Finest Level.

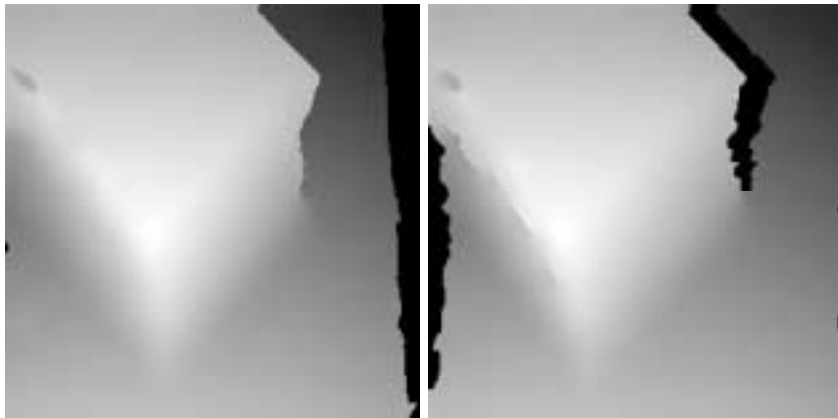
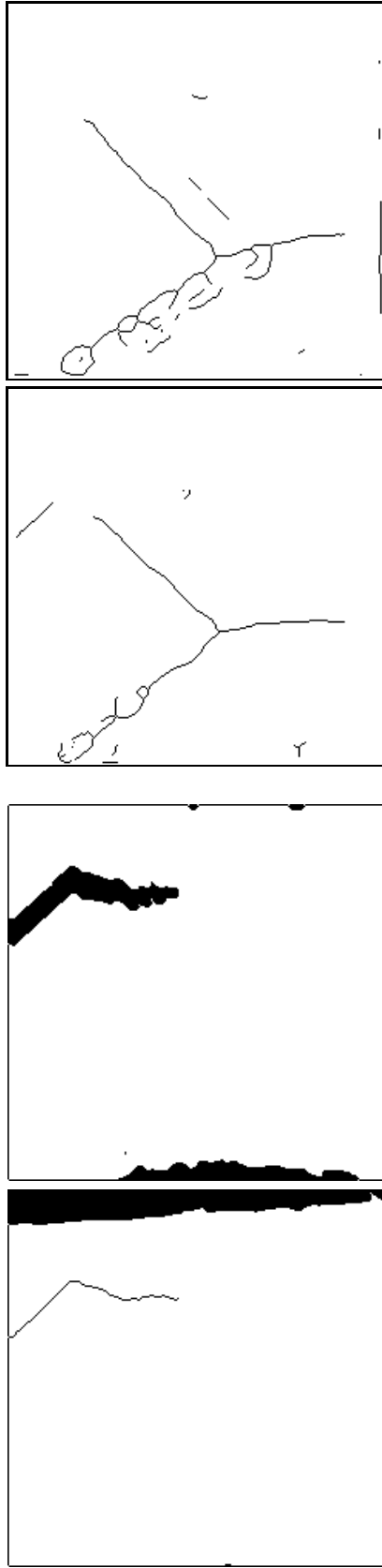
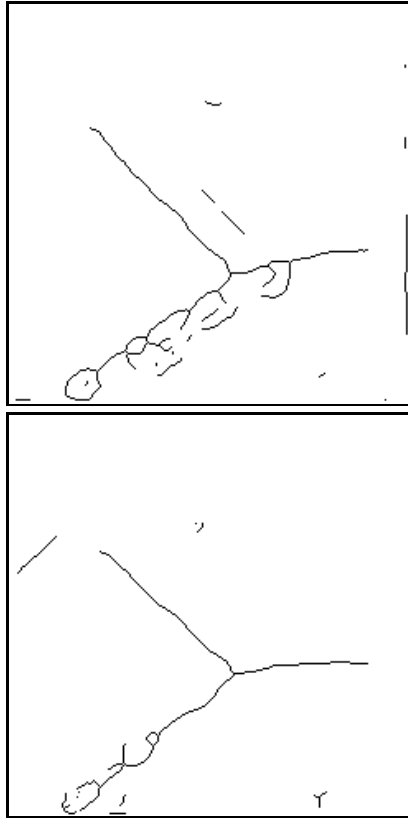


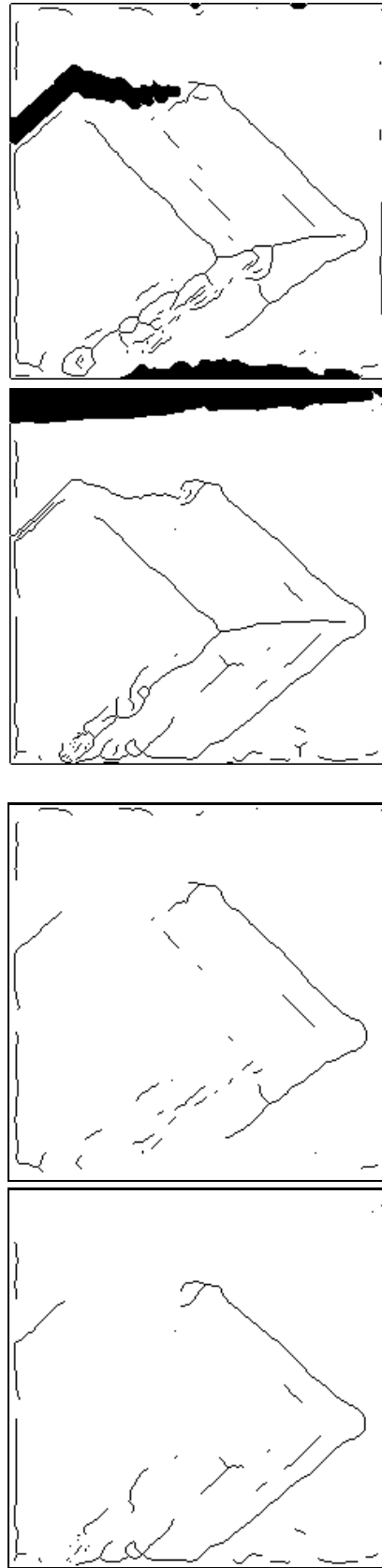
Figure 5.8: Books — Disparity Surface Using Estimates from Coarser Resolution.



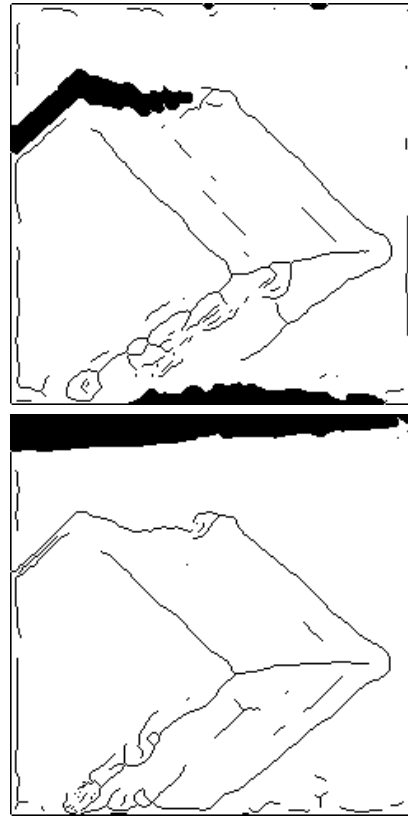
(a) Depth Discontinuities and Occluded Regions.



(b) Convex Orientation Discontinuities.



(c) Concave Orientation Discontinuities.



(d) Combined Discontinuities.

Figure 5.9: Books — Surface Features.

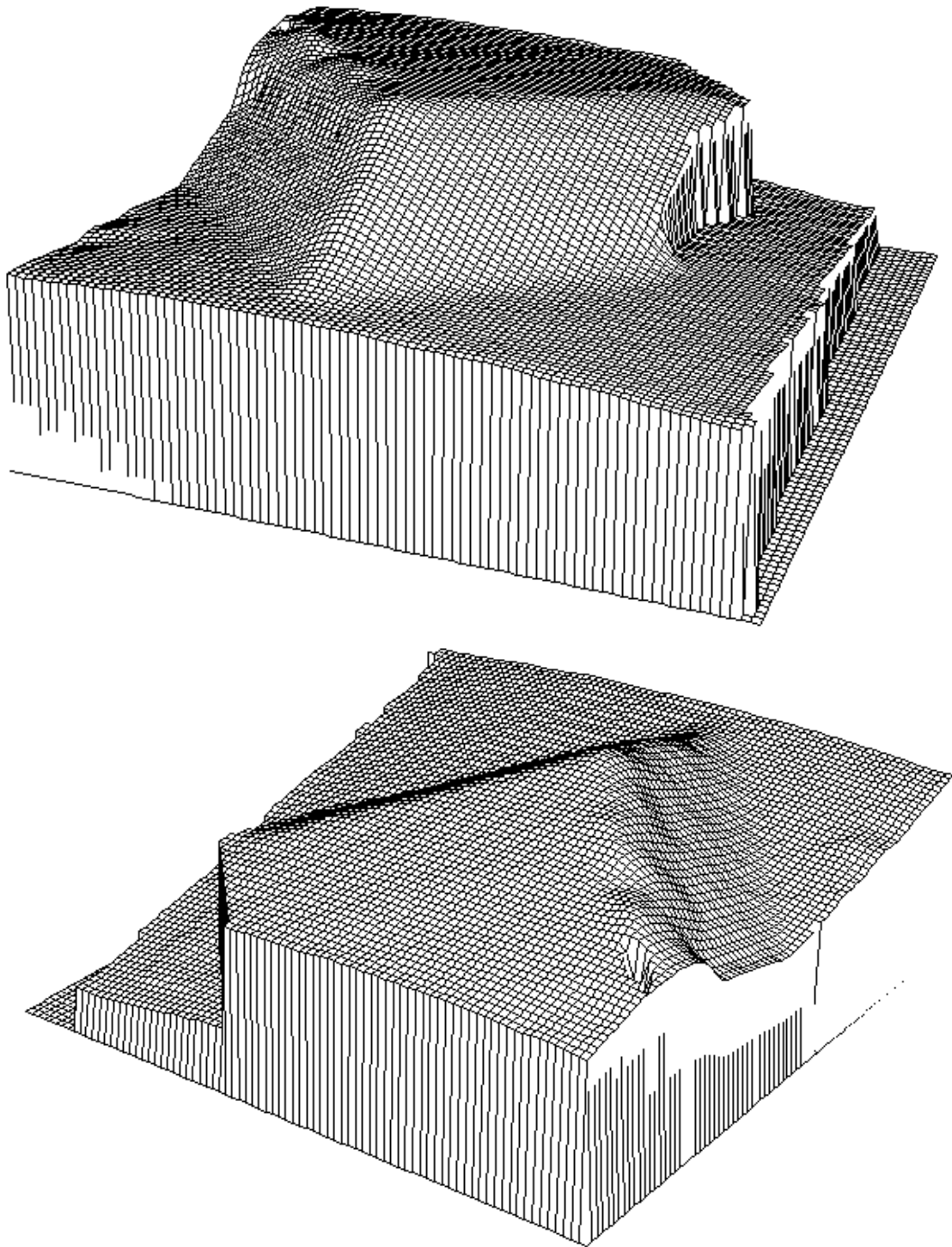


Figure 5.10: Books — 3-D Plot of the Integrated Results.



Figure 5.11: Books — 3-D Rendered View of the Integrated Results.

5.2.3 Jussieu

The stereo pair in figure 5.12 shows an aerial view of an urban scene, part of a French University located in Paris at about 48.8° North Latitude by 2.3° East Longitude. This image is $256 \times 256 \times 8$ bits, cropped from a larger image obtained from Jean-Luc Jezouin of Matra MS2I, France.

Figure 5.13 shows the edges and figure 5.14 shows the integrated results. The discontinuities in this image often produce poor edges, and the sharp angles do not allow the adaptive smoothing to work very effectively. This gives us a very rough contour which does *not* follow the intensity edges as we desire.

How can we solve this? One approach that has proved effective is to represent the estimated depth discontinuities as dynamic splines or “snakes” [41]. We use a variant of these dynamic splines developed by Menet *et al.* [55] in which the curves

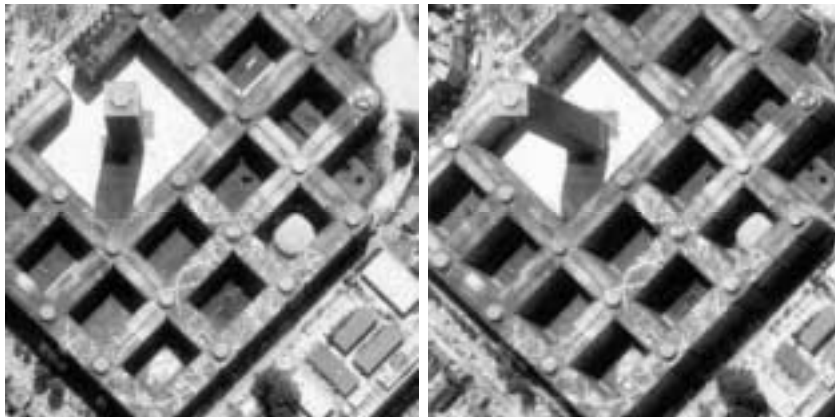


Figure 5.12: Jussieu — Original Intensity Images.

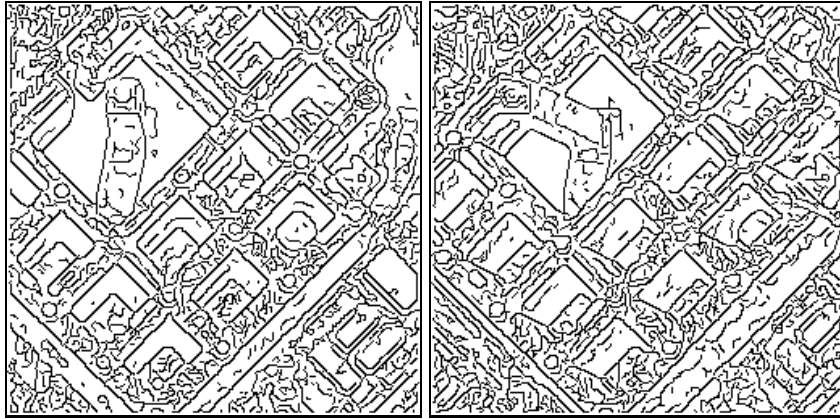


Figure 5.13: Jussieu — Edgels.

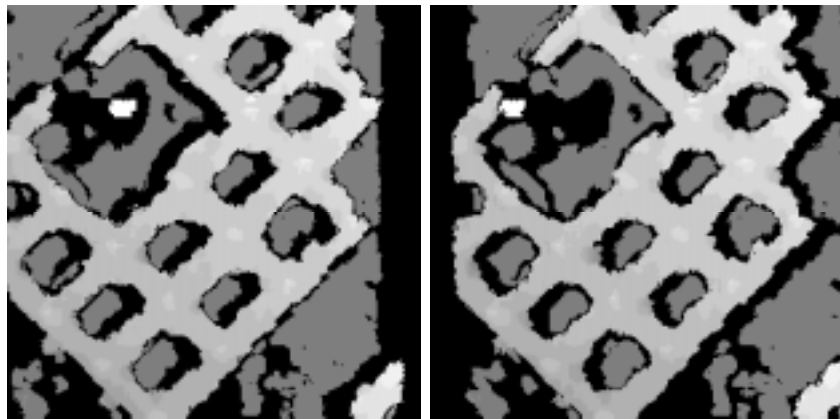


Figure 5.14: Jussieu — Disparity Surface.

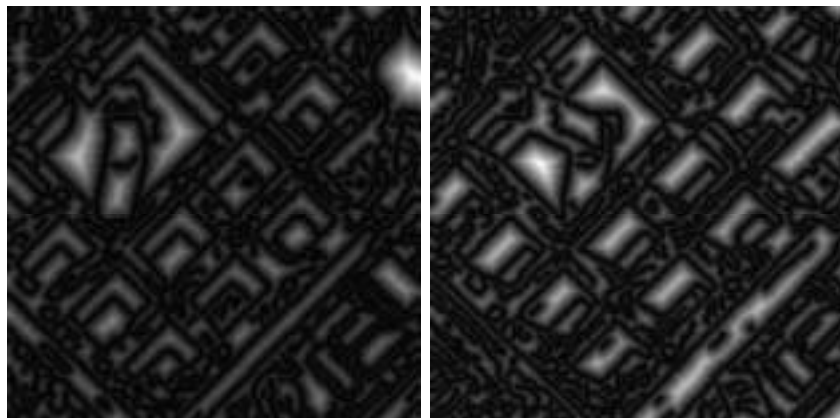


Figure 5.15: Jussieu — Chamferred Edges.

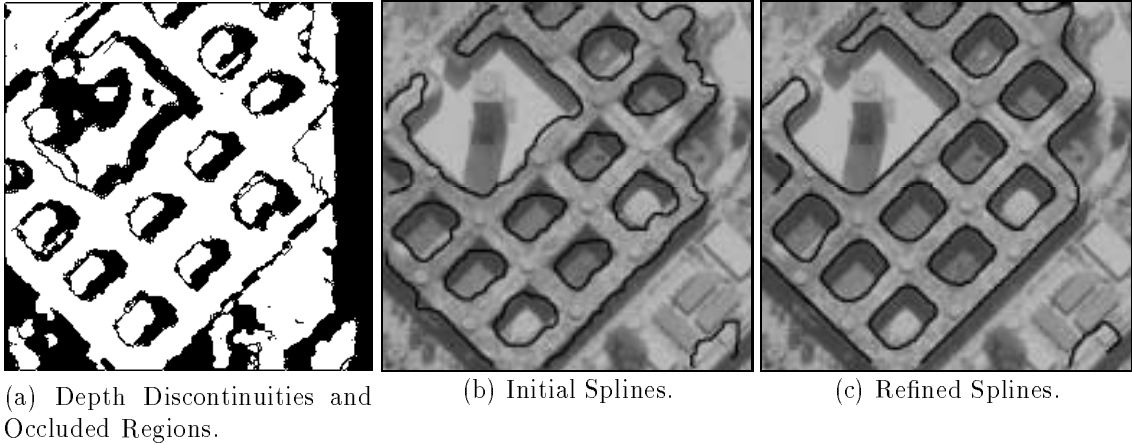


Figure 5.16: Jussieu — Refinement of Depth Discontinuities.

are approximated by a B-spline to speed the processing, but the idea is the same. Each depth discontinuity may be represented as a linked chain in the form $V(s) = (X(s), Y(s))$, and we let

$$\begin{aligned}
 E^* &= \int_0^1 E(V(s)) ds \\
 &= \frac{1}{2} \int_0^1 [\alpha |V_s(s)|^2 + \beta |V_{ss}(s)|^2] ds + \int_0^1 E_c ds
 \end{aligned}$$

where the first term, regulated by α , makes the curve act as a string and adjusts the tension of the spline; the second term, controlled by β , makes it act as a thin rod and adjusts the bending energy of the spline. The last term is the external constraint energy, which is used to force the discontinuity (represented by the spline) to lie along the edges. Figure 5.15 shows the chamfer distance [12], that we use for this external constraint, in which the value at each point is the distance to the nearest edge. Figure 5.16 shows the depth discontinuities, the initial splines representing the discontinuities, and the splines after they have converged to a minimal energy state.

Figures 5.17 and 5.18 show the 3-D views and the rendered scene of the disparity shown in Figure 5.16(a) respectively.

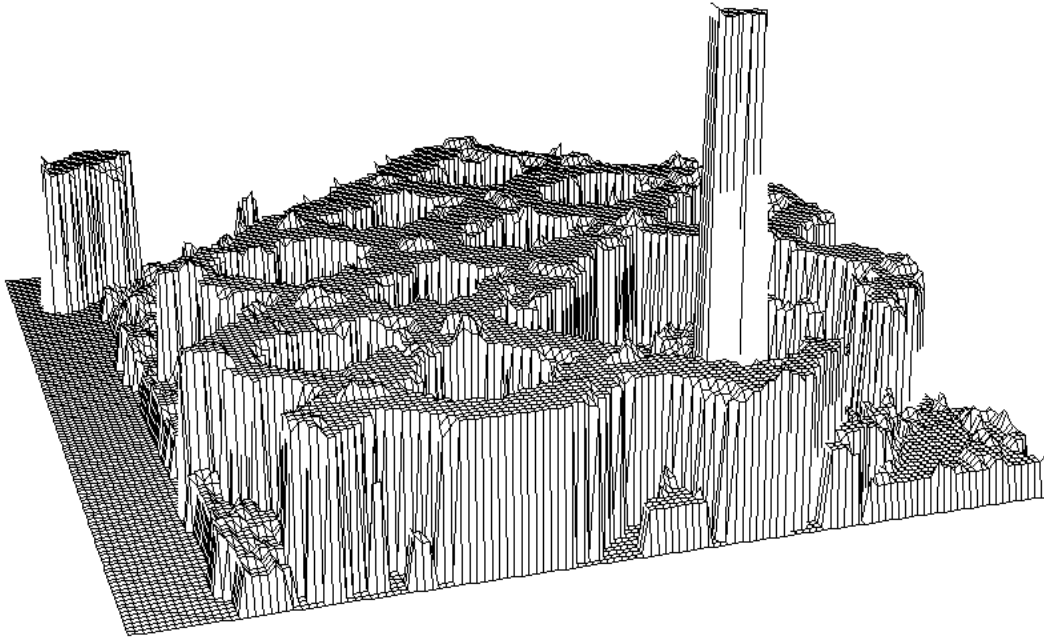
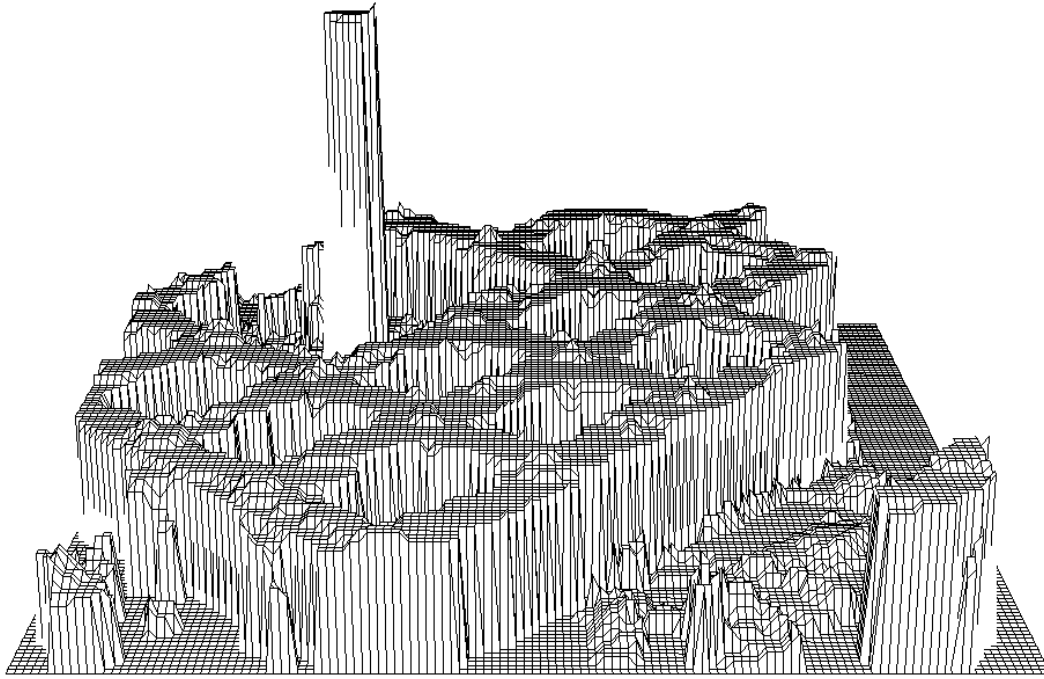


Figure 5.17: Jussieu — 3-D Plot of the Integrated Results.

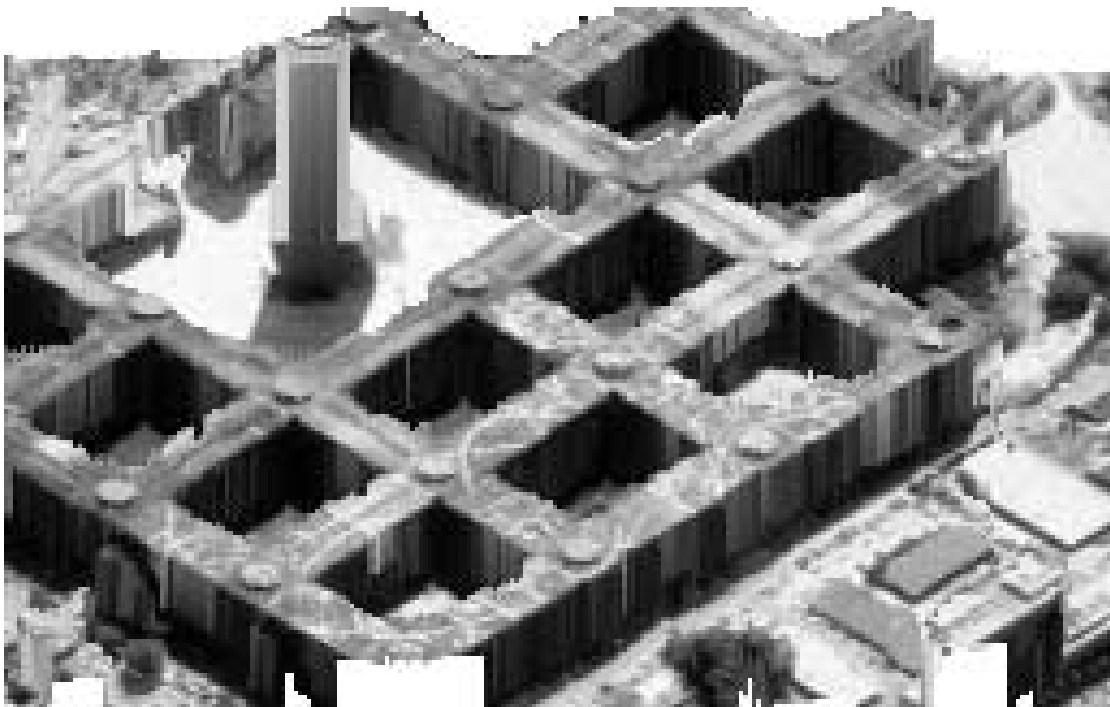


Figure 5.18: Jussieu — 3-D Rendered View of the Integrated Results.

5.2.4 Blocks

Figure 5.19 is a $295 \times 295 \times 8$ bits image of blocks on a table taken with a large angular separation with areas of both high and low texture. In addition the fusion interval was intentionally chosen so that the back wall falls outside of it, so the correct value cannot be obtained there. Also, the “Play-doh” container is more strongly textured than the wood figures. Finally, the lower wooden cylinder has a grain pattern which is approximately parallel to the epipolar lines of the image. Thus, most of the adverse conditions for our approach are present in the image.

Figure 5.20 shows the recovered disparity surfaces, while figure 5.21(a) shows the unknown regions along with a few depth discontinuities figure 5.21(b) shows the

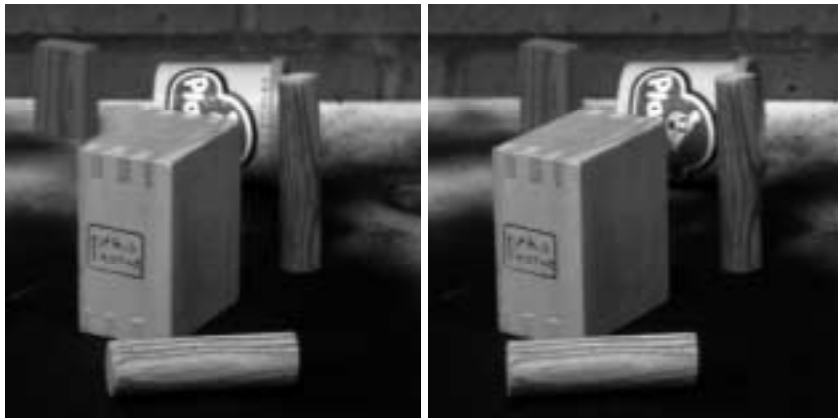


Figure 5.19: Blocks — Original Intensity Images.

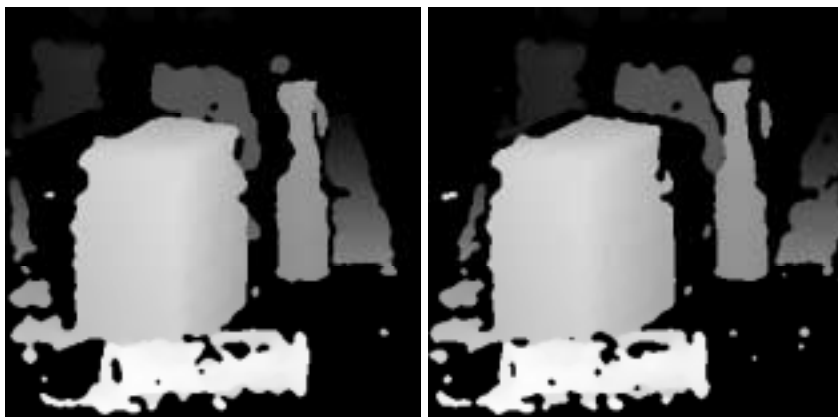
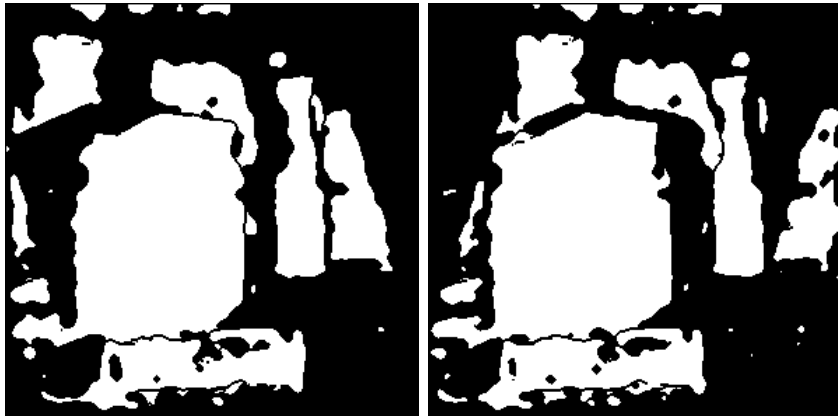
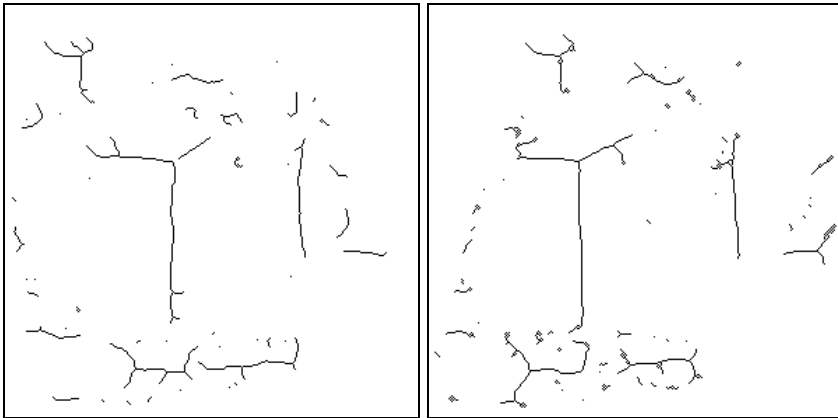


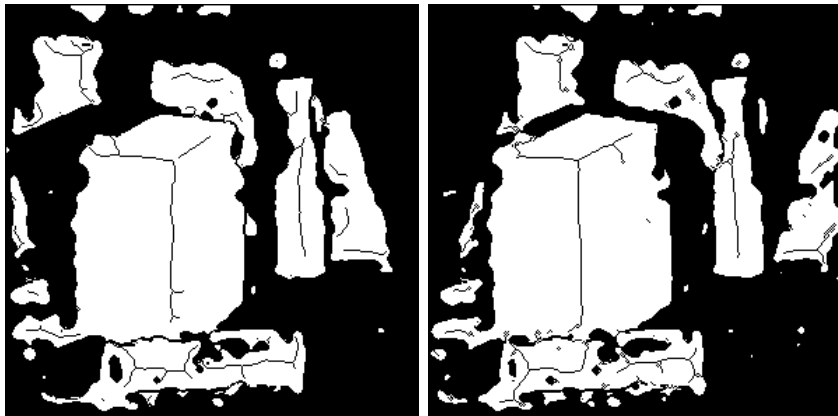
Figure 5.20: Blocks — Disparity Surface.



(a) Depth Discontinuities and Unknown Regions.



(b) Orientation Discontinuities.



(c) Combined Discontinuities.

Figure 5.21: Blocks — Surface Features.

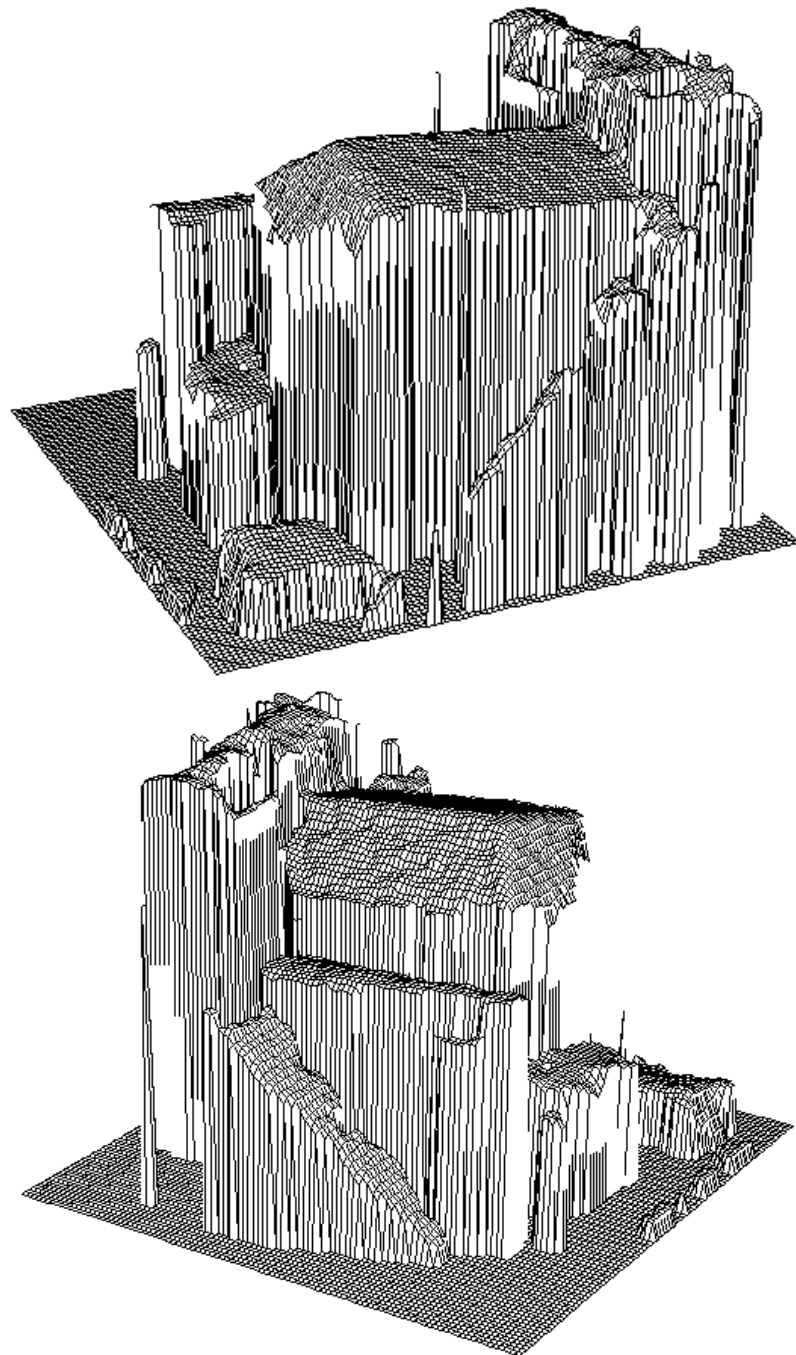


Figure 5.22: Blocks — 3-D Plot of the Integrated Results.

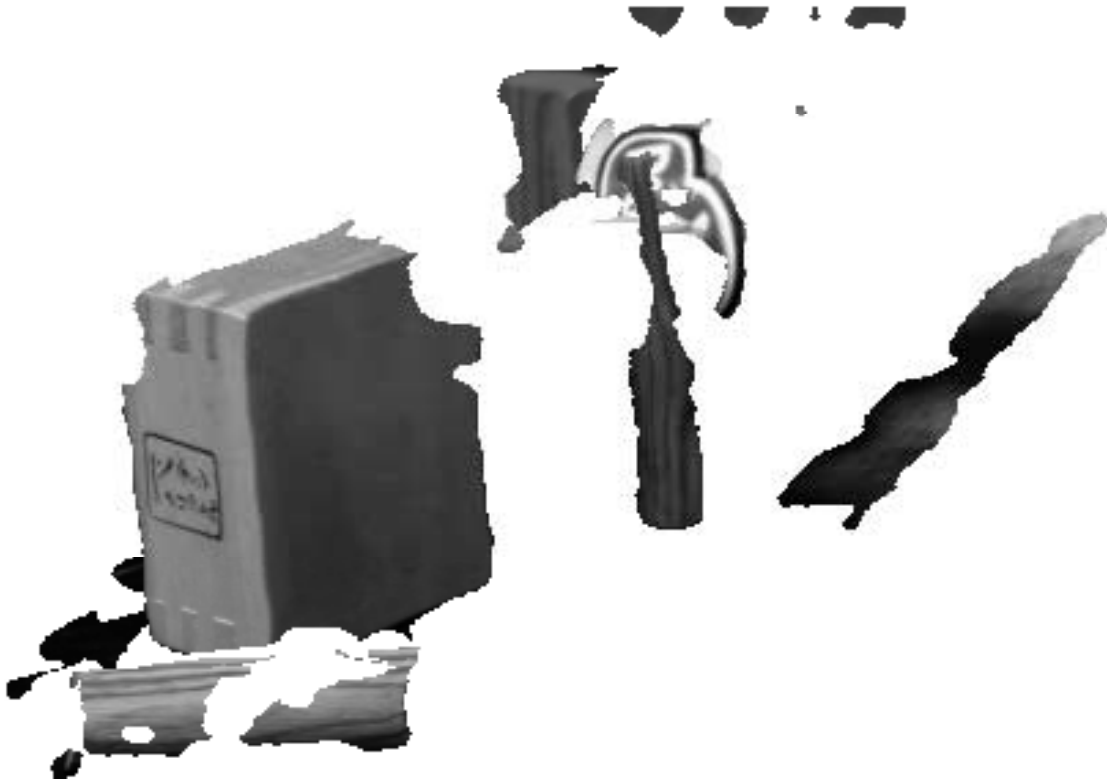


Figure 5.23: Blocks — 3-D Rendered View of the Integrated Results.

orientation discontinuities and they are combined in figure 5.21(c). Although much of the scene was discarded as unmatchable due to occlusion, or lack of texture, the remainder is present and has accurate disparity values. Even though most of the brick wall is absent, some small parts that were just within the fusion interval were correctly matched, and can be seen in the 3-D plots in figure 5.22. Finally figure 5.23 shows a rendered view of the matched parts of this scene. This example provides an illustration of the graceful degradation of our algorithm in the presence of a difficult image pair.

5.3 Other Results

In the following sections we show the results of six other image pairs from three domains: Three aerial scenes, an indoors scene, and two outdoors scenes.

5.3.1 Montagne du Lubéron (SPOT Image)

Figure 5.24 shows a pair of SPOT images of natural terrain at about 43.5° North Latitude by 5.2° East Longitude (about 50Km north of Marseille, France) obtained from Jean-Luc Jezouin at Matra MS2I, France. This scene is $256 \times 256 \times 8$ bits and has few useful edges, but lots of texture. The disparity ranges from -12 (near) to 7 (far) pixels.

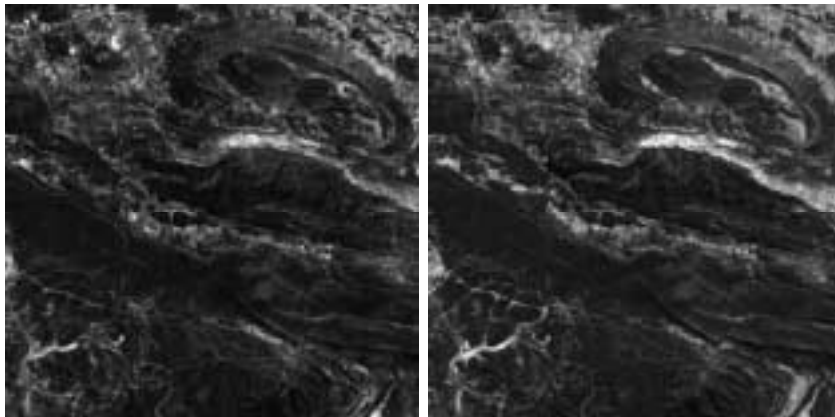


Figure 5.24: Lubéron — Original Intensity Images.

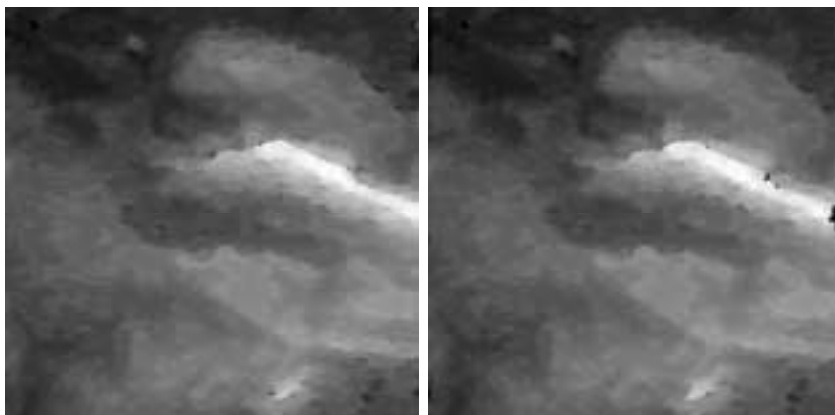
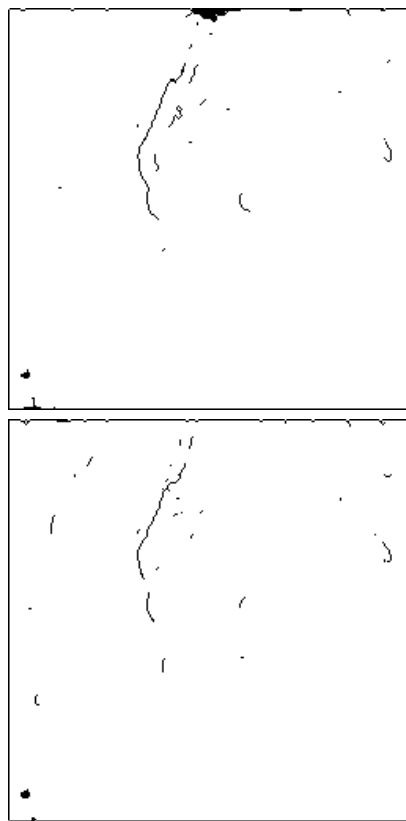
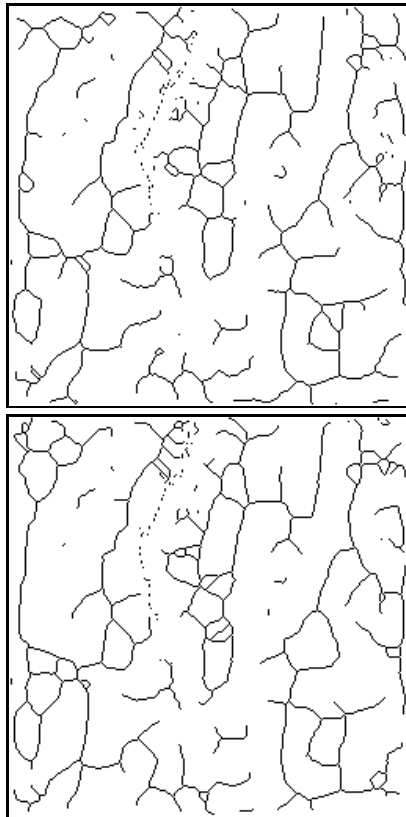


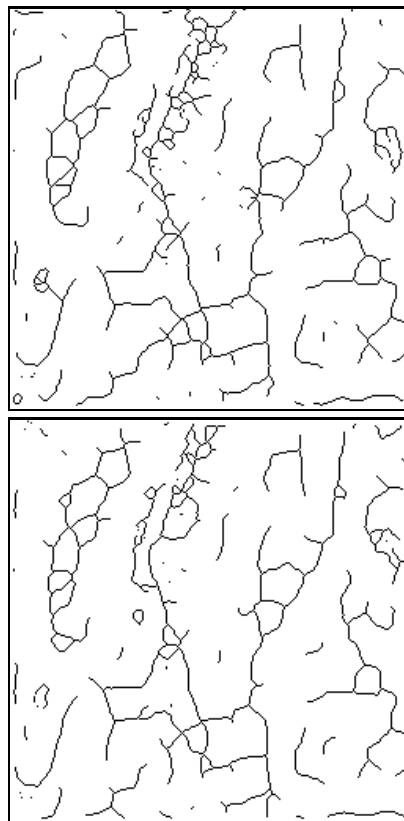
Figure 5.25: Lubéron — Disparity Surface.



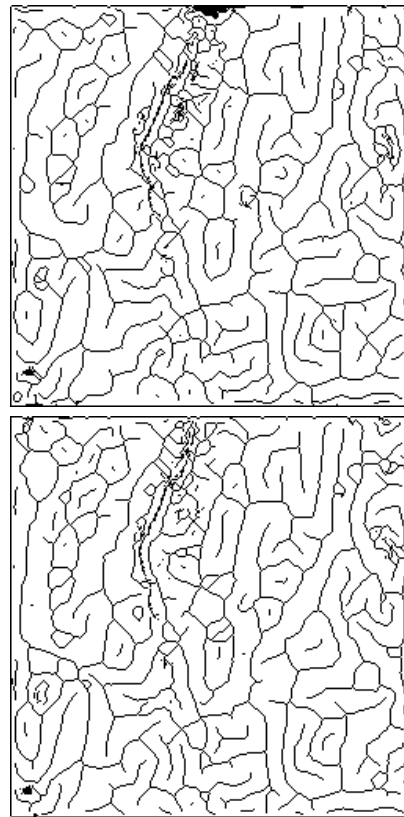
(a) Depth Discontinuities and Occluded Regions.



(b) Convex Orientation Discontinuities.



(c) Concave Orientation Discontinuities.



(d) Combined Discontinuities.

Figure 5.26: Lubéron — Surface Features.

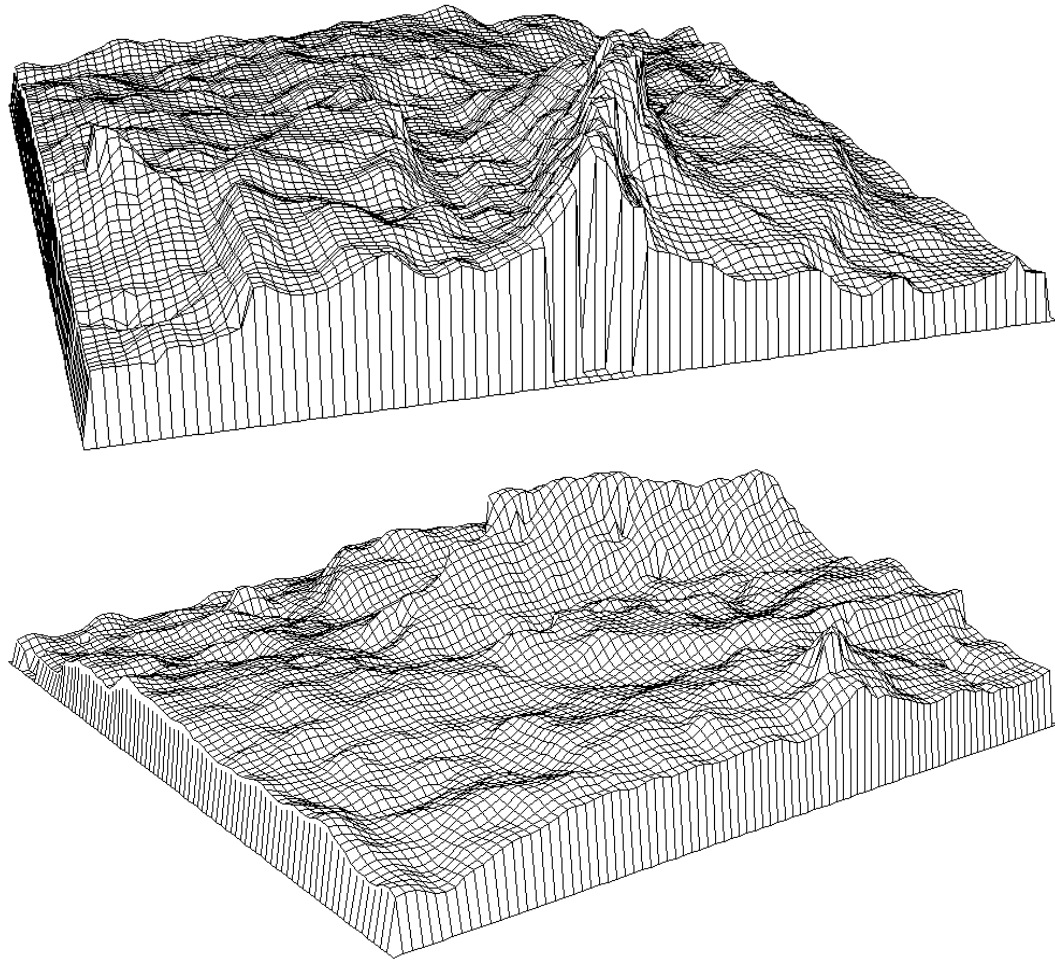


Figure 5.27: Lubéron — 3-D Plot of the Integrated Results.



Figure 5.28: Lubéron — 3-D Rendered View of the Integrated Results.

Note that the images were taken at different times and the the amount of snow cover is different — still the matching and constraints are sufficient to generate the correct matching over most of the scene (Figure 5.25).

Figure 5.26 shows the surface features extracted from this scene. There are no real depth discontinuities in the scene, but the steep mountain at the center generates a gradient which exceeds the threshold for depth discontinuities — it also shows up as the dotted convex orientation discontinuity.

Figures 5.27 and 5.28 show the 3-D plot of the disparity surface and the reconstruction of the scene respectively.

5.3.2 Pentagon

Figure 5.29 shows an aerial view of the Pentagon in Washington D.C., USA (about 38.5° North Latitude and 77.0° West Longitude) which was obtained from Professor Takeo Kanade of the Carnegie-Mellon University Computer Science Department. The image is $512 \times 512 \times 8$ and the disparity ranges from -9 (near) to 8 (far) pixels. Figure 5.30 show the results following the integration of the area and feature based processing. The results are quite impressive, but also point out two weaknesses with the method. The first is that when the disparity difference between two surfaces is only about 1 pixel, they cannot be accurately separated as can be seen around the underpass of the bridge in the upper right corner. Also, when two edges are very close together, the “blurring” can extend past more than one edge. We have no solution for this except to work at a higher level of resolution or to attempt to obtain sub-pixel accuracy. Note that the occluded areas around the walls are well localized except for a false match in the area-based processing of the inside wall at the left.

Figure 5.31 shows the location of the depth discontinuities. Orientation discontinuities are not very meaningful for this image since almost everything is flat and there is very little relative disparity.

Figure 5.32 shows a 3-D plot of figure 5.30(b) and figure 5.33 shows a rendered view of the same figure.



Figure 5.29: Pentagon — Original Intensity Images.

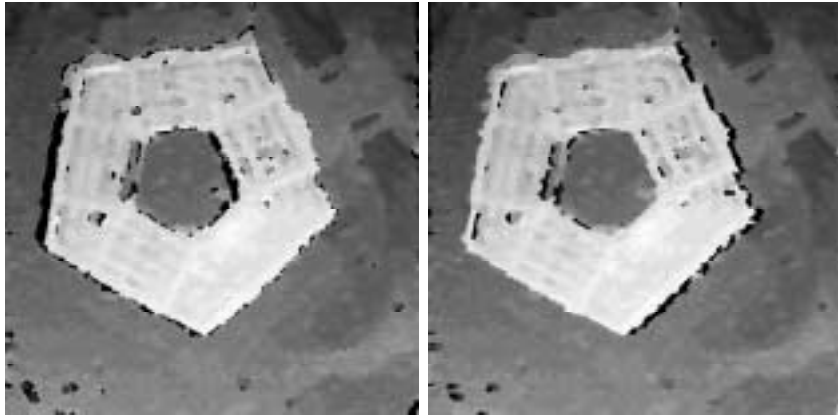


Figure 5.30: Pentagon — Disparity Surface.

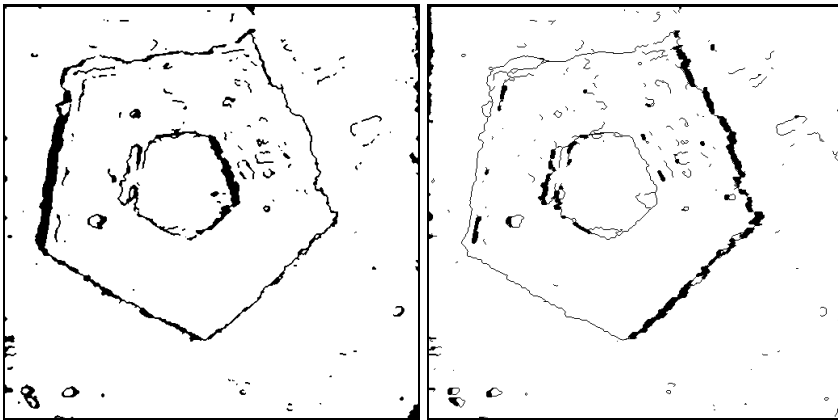


Figure 5.31: Pentagon — Depth Discontinuities and Occluded Regions.

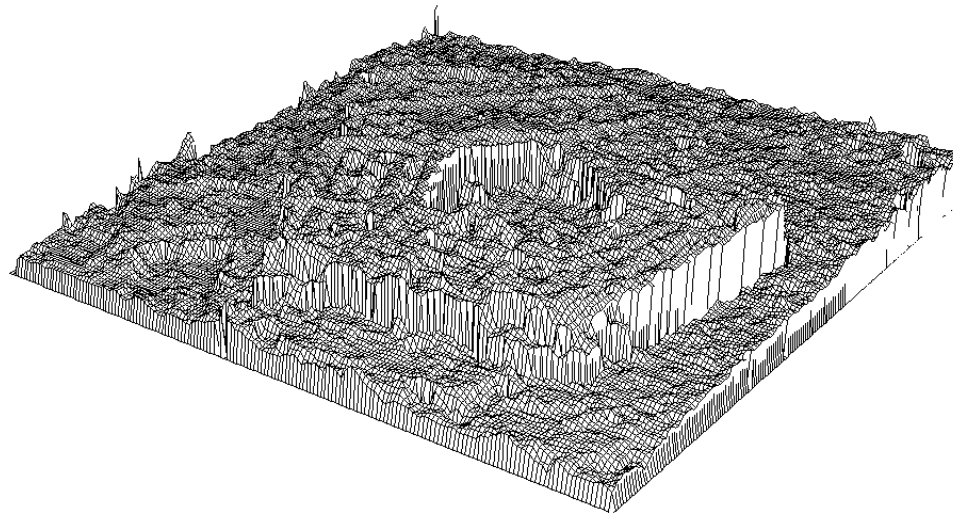
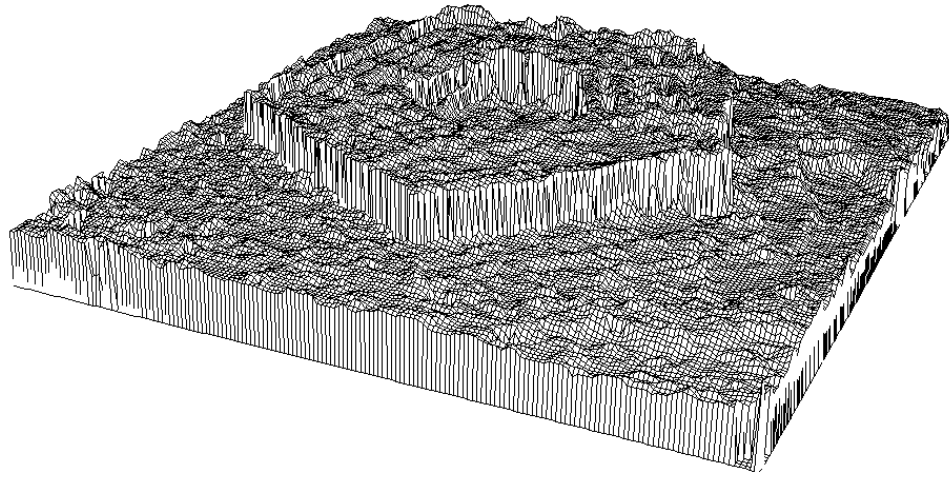


Figure 5.32: Pentagon — 3-D Plot of the Integrated Results.



Figure 5.33: Pentagon — 3-D Rendered View of the Integrated Results.

5.3.3 Nuclear Power Plant

The stereo pair in figure 5.34 show an aerial view of another cultural feature, part of a French Nuclear Power plant. This image is $256 \times 256 \times 8$ cropped from a larger image obtained from Jean-Luc Jezouin at Matra MS2I, France. Figure 5.35 shows the edges and figure 5.36 shows the integrated results.

Figure 5.38 shows the depth discontinuities extracted from the results along with the occluded regions and the fine-tuning of those depth discontinuities using the dynamic splines. The external energy is supplied by the chamfered edges shown in figure 5.37.

Figures 5.39 and 5.40 show the 3-D views and the rendered scene respectively, of the disparity shown in the left view image of Figure 5.36(left).



Figure 5.34: Power Plant — Original Intensity Images.

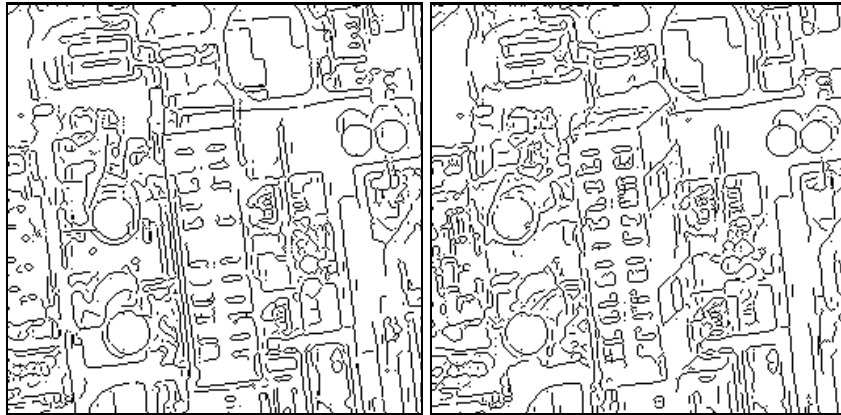


Figure 5.35: Power Plant — Edgels.

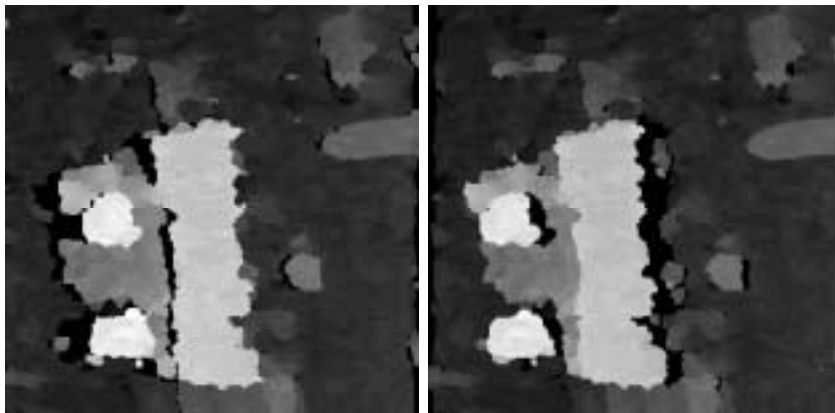


Figure 5.36: Power Plant — Disparity Surface.

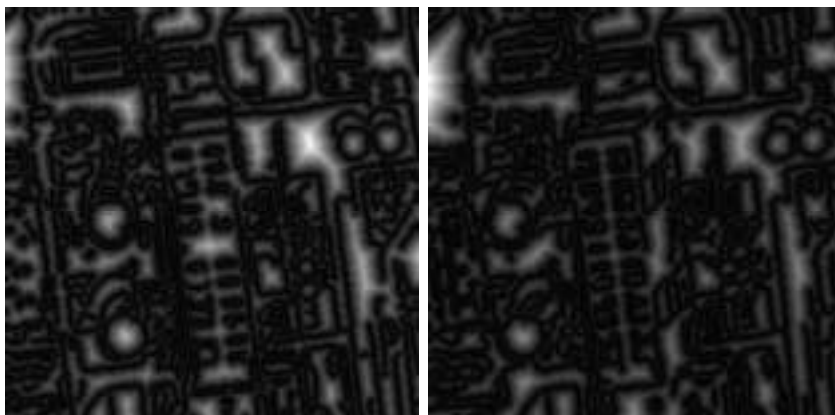


Figure 5.37: Power Plant — Chamferred Edges.

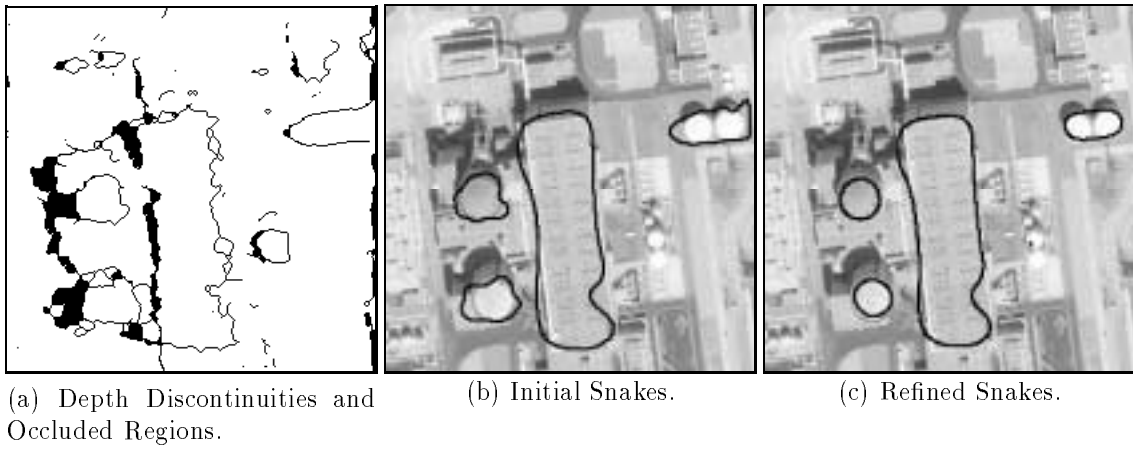


Figure 5.38: Power Plant — Refinement of Depth Discontinuities.

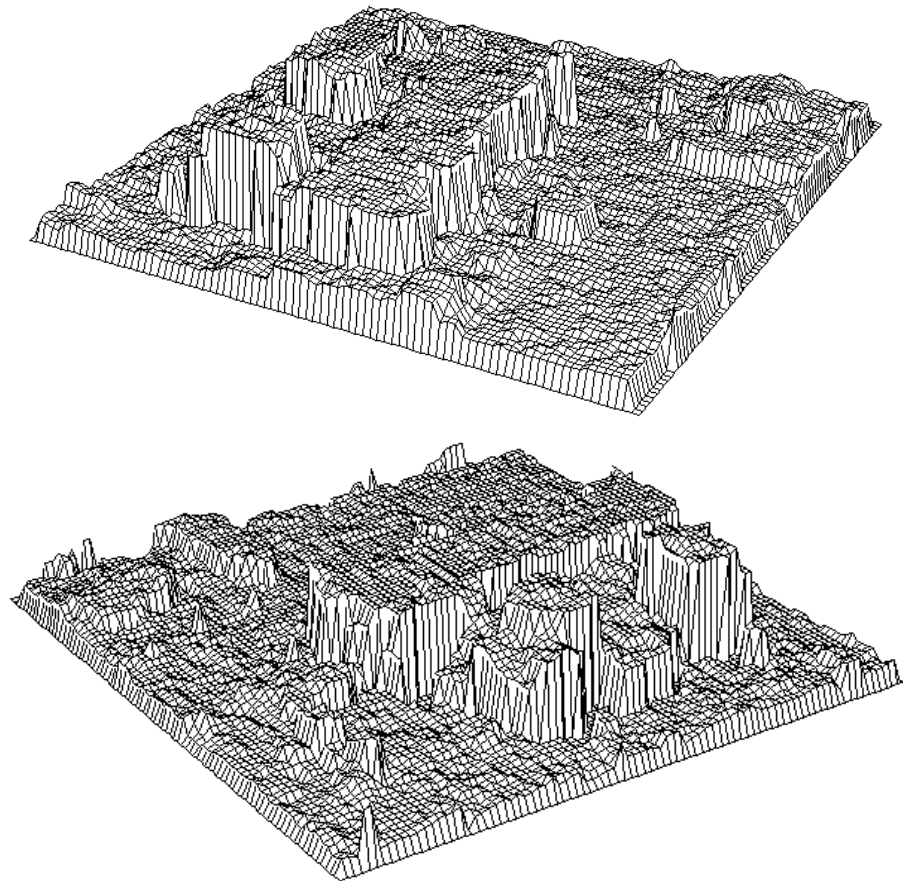


Figure 5.39: Power Plant — 3-D Plot of the Integrated Results.

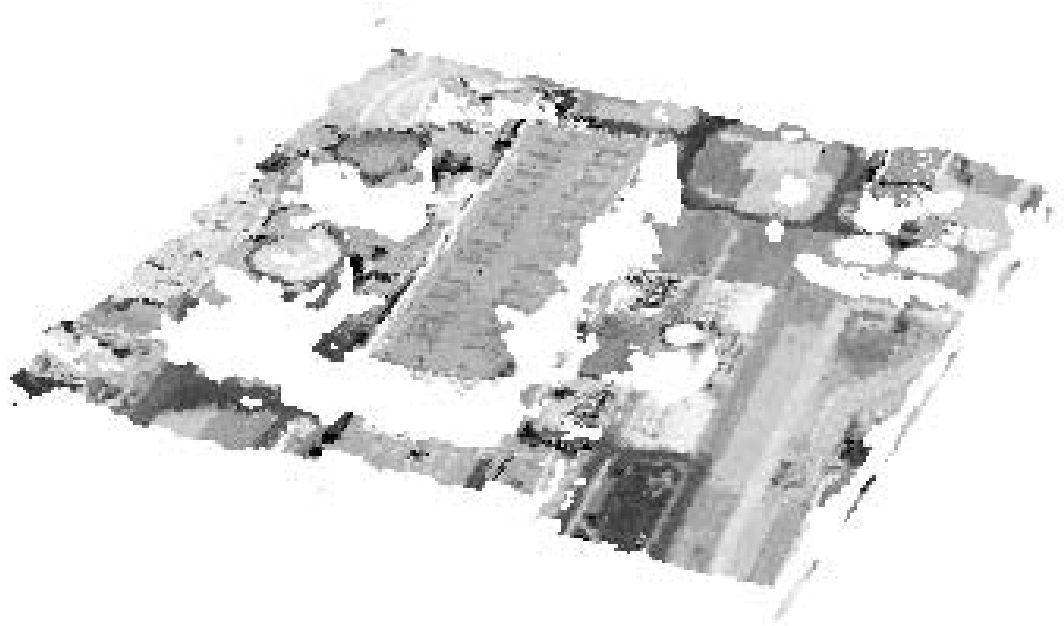


Figure 5.40: Power Plant — 3-D Rendered View of the Integrated Results.

5.3.4 Fruit on a Table

Figure 5.41 is another image obtained from the University of Illinois, courtesy of Dr. W. Hoff, now with Martin Marietta. This image is $232 \times 256 \times 8$ bits. Here again the background is more textured than the foreground. In addition, the tablecloth pattern ripples with a period of about 16 pixels which generates aliasing within the fusion interval. The multi-level matching succeeds and the image matches are generally correct (figure 5.42), except along the lower part of the cantaloupe. In addition, the border of the cantaloupe is rough due to the edge cutback as was seen with the Power Plant example. This process did however leave a small bad match at the upper left edge.

Figure 5.43(a) shows the depth discontinuities and the occluded regions while (b) and (c) show the orientation discontinuities. Note that the maximum of curvature

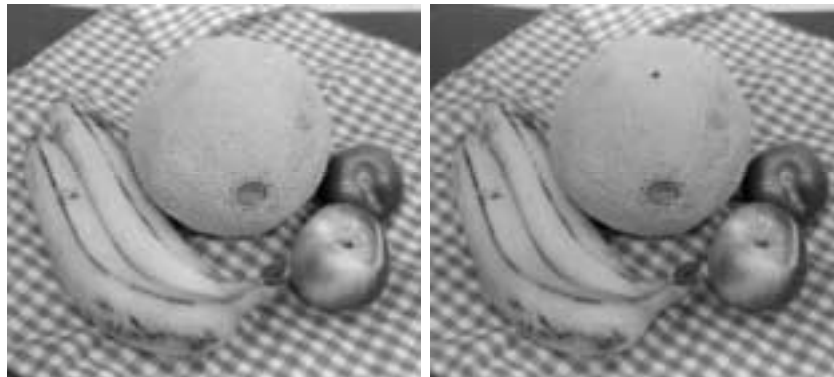


Figure 5.41: Fruit Scene — Original Intensity Images.

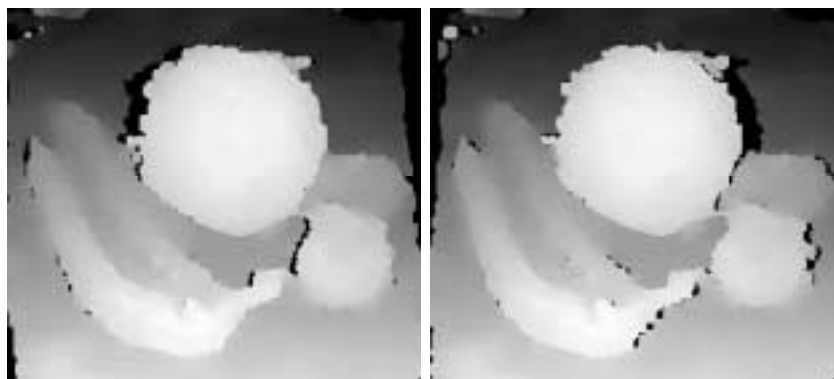


Figure 5.42: Fruit Scene — Disparity Surface.

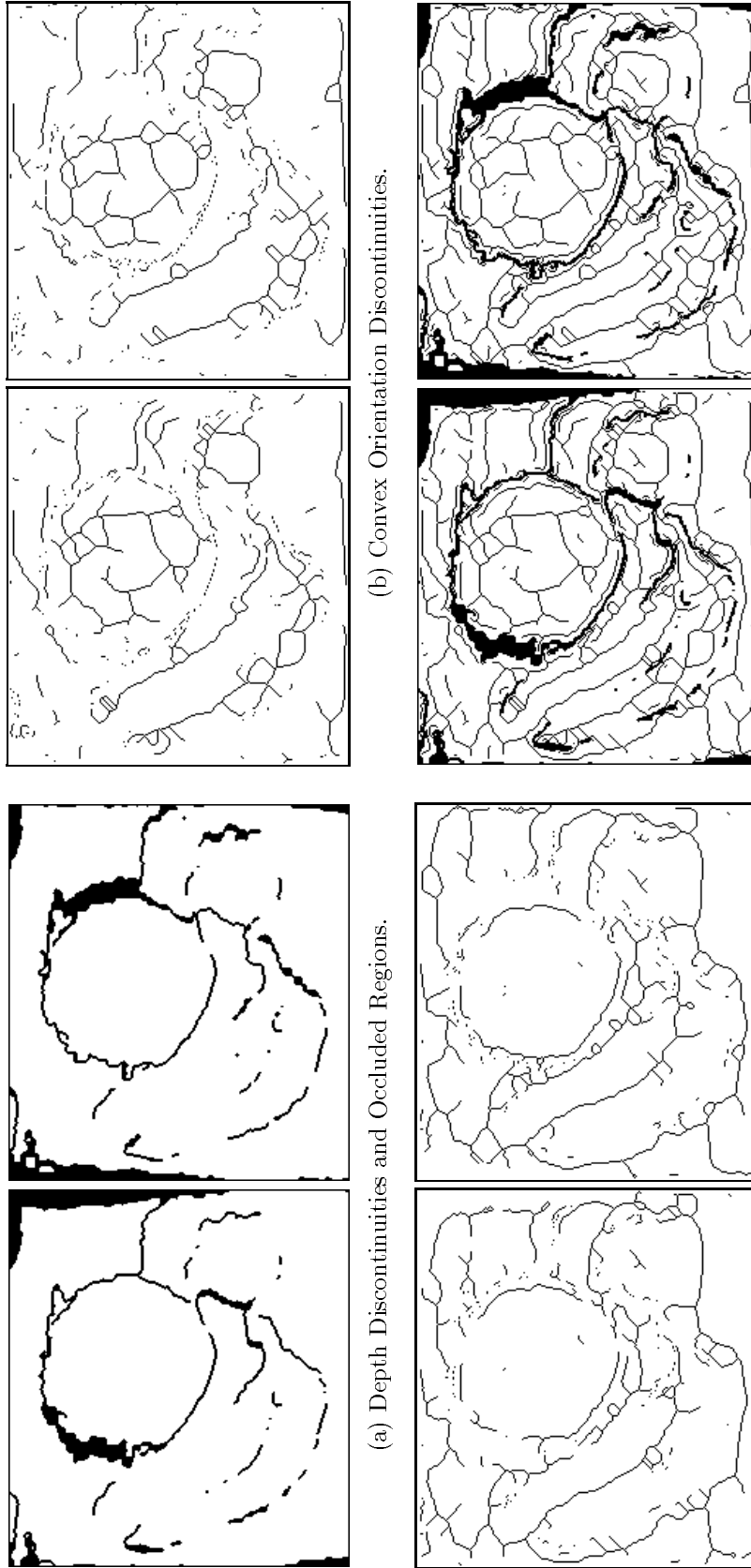


Figure 5.43: Fruit Scene — Surface Features.

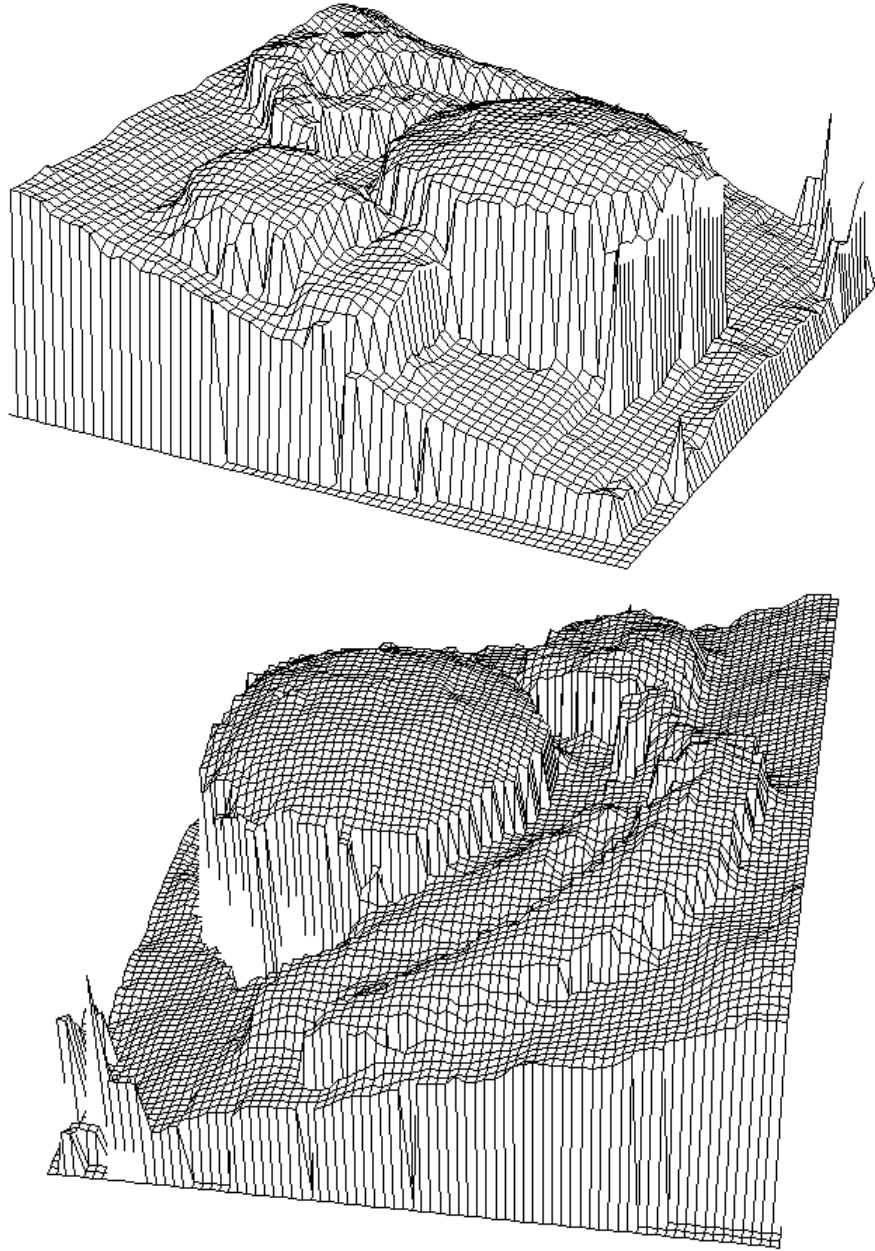


Figure 5.44: Fruit Scene — 3-D Plot of the Integrated Results.



Figure 5.45: Fruit Scene — 3-D Rendered View of the Integrated Results.

surrounding the stem-end of the nearer apple and the cantaloupe show as possible convex discontinuities.

Figure 5.44 shows two 3-D plots of this scene while figure 5.45 shows a rendered view of the scene. Note the depth discontinuity along the lower-left fold in the tablecloth.

5.3.5 Quarry Wall

The Quarry Wall scene shown in figure 5.46 is one of the International Society of Photogrammetry and Remote Sensing Working Group III/4 stereo test images [29]. In this scene there are not any real depth discontinuities (except near the top left corner), but there are a large number of orientation discontinuities along the edges of the “blocks” in the wall.

Figure 5.47 shows the disparity map and figure 5.48 show the surface features, while figures 5.49 and 5.50 show the reconstruction of the wall. Note that the wall is pretty accurate, although a bit smoothed.

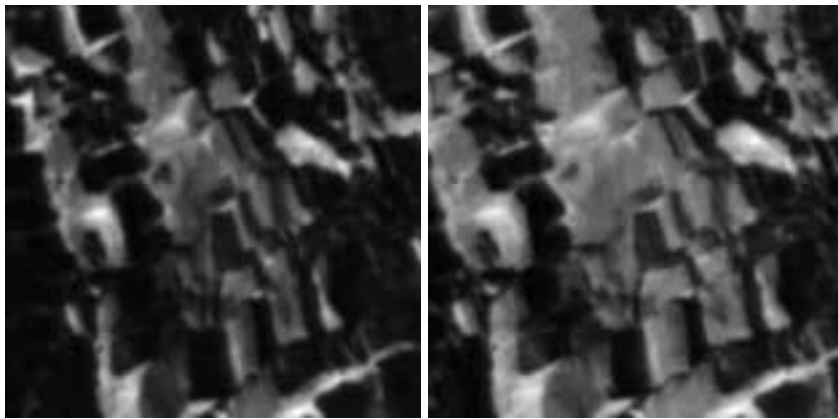


Figure 5.46: Quarry Wall — Original Intensity Images.

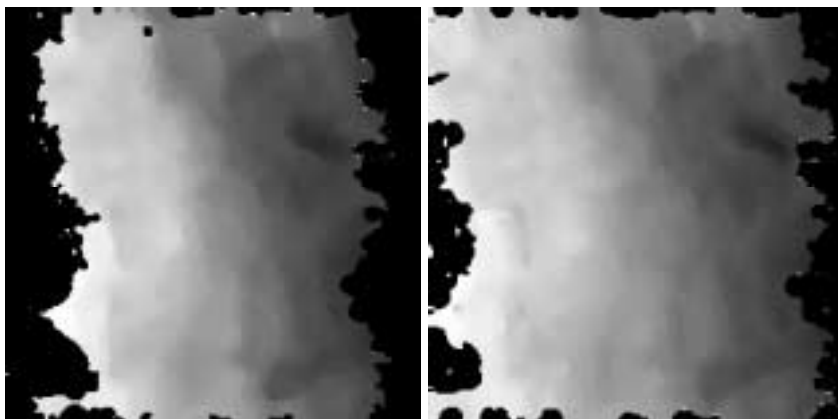
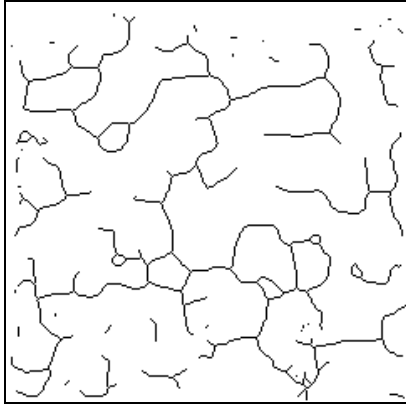
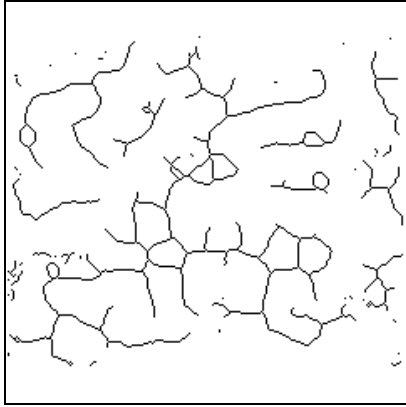


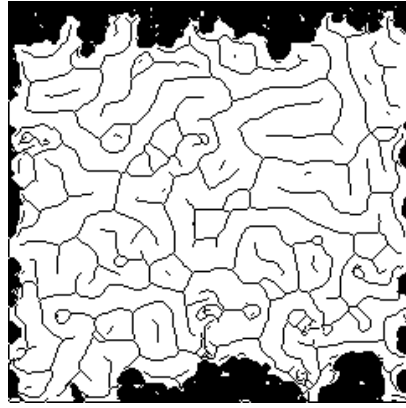
Figure 5.47: Quarry Wall — Disparity Surface.



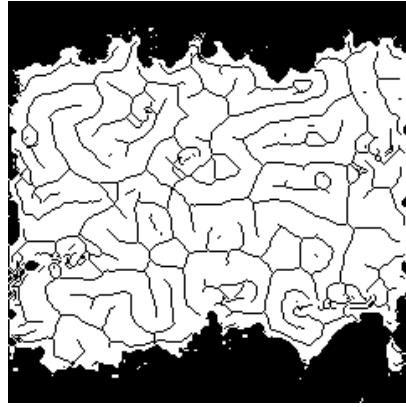
(a) Depth Discontinuities and Occluded Regions.



(b) Convex Orientation Discontinuities.



(c) Concave Orientation Discontinuities.



(d) Combined Discontinuities.

Figure 5.48: Quarry Wall — Surface Features.

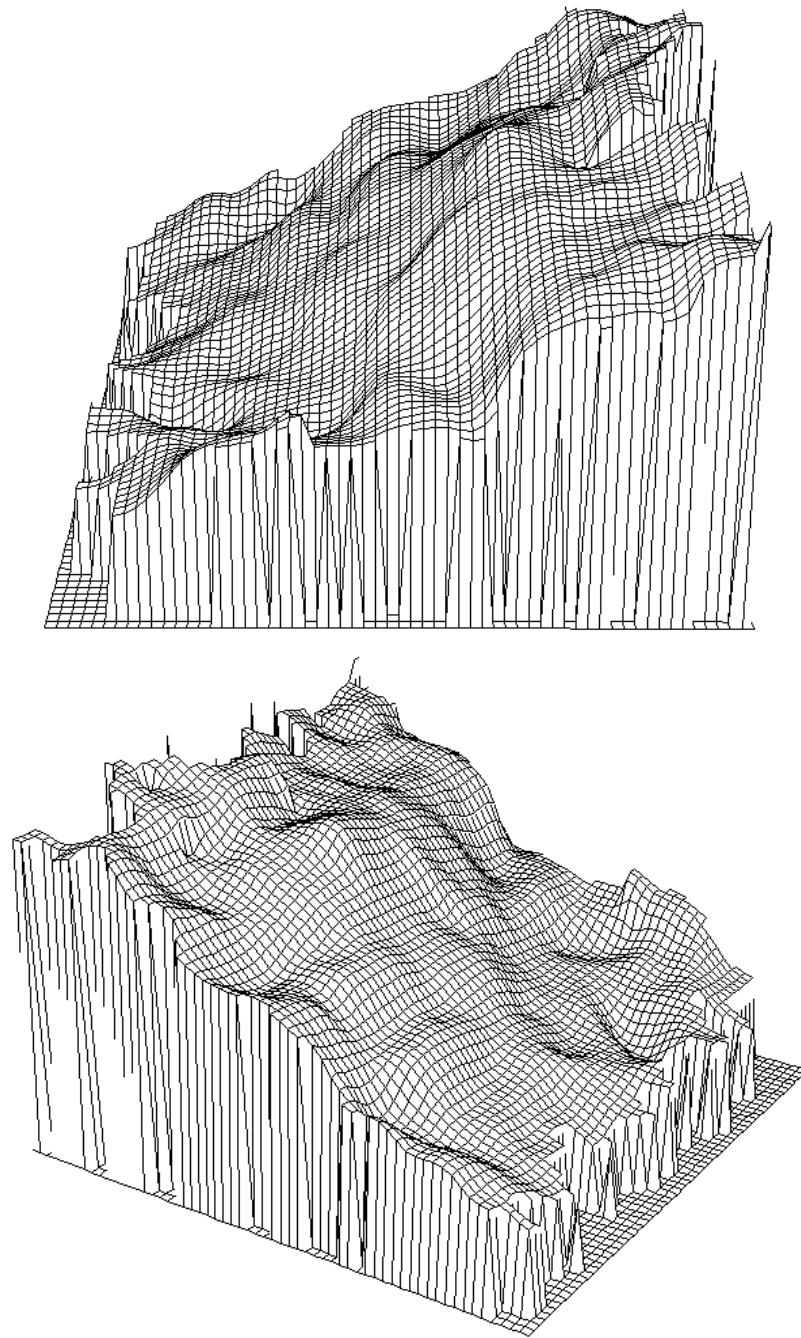


Figure 5.49: Quarry Wall — 3-D Plot of the Integrated Results.

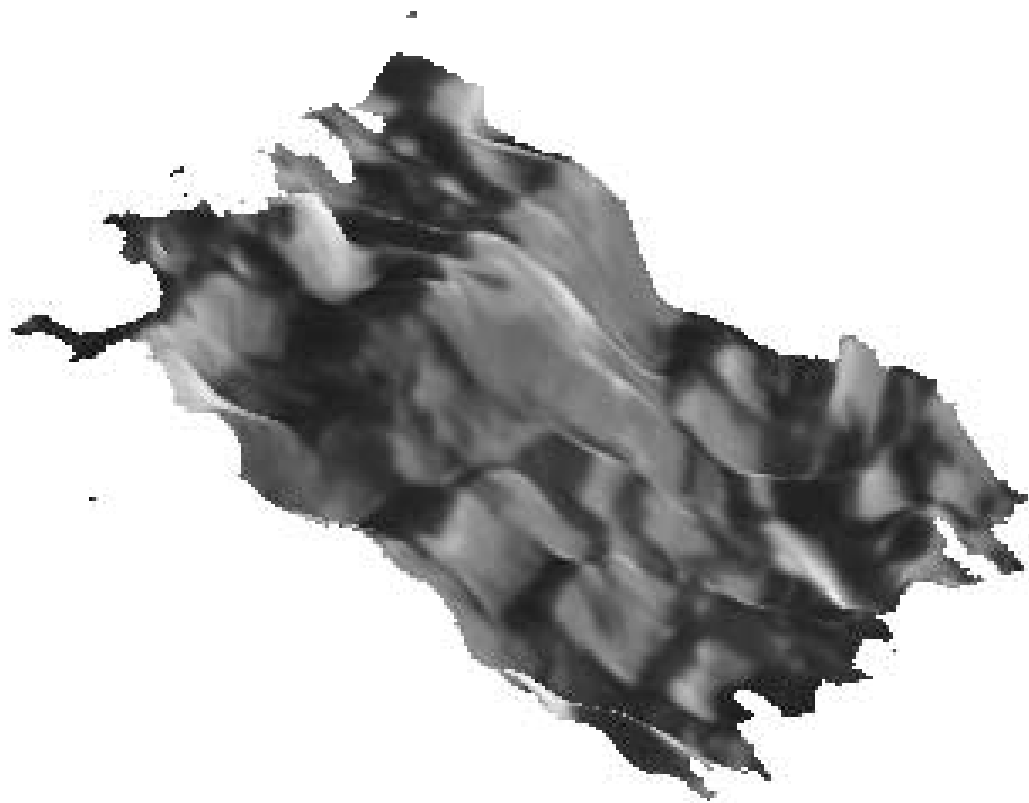


Figure 5.50: Quarry Wall — 3-D Rendered View of the Integrated Results.

5.3.6 Suspension Bridge

The Suspension Bridge scene shown in figure 5.51 is another of the International Society of Photogrammetry and Remote Sensing Working Group III/4 stereo test images [29]. In this scene there are not any depth discontinuities, unless we count the “transparent” surface formed by the chain-link fence on the right of the image. The links form one surface and the shadow of the bridge on the landscape below forms a second surface.

Figure 5.52 shows the disparity map and figure 5.53 show the surface features. Our algorithm cannot handle transparent surfaces, since the chain-link fence is (globally) the more highly textured aspect, it was selected as the surface. In the early processing,



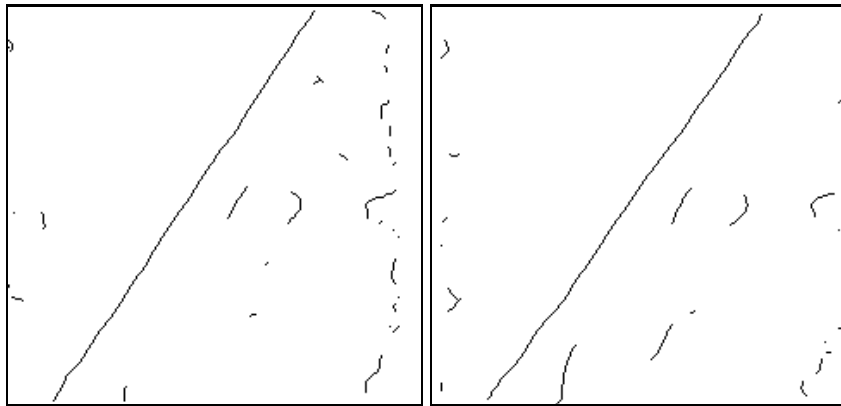
Figure 5.51: Suspension Bridge — Original Intensity Images.



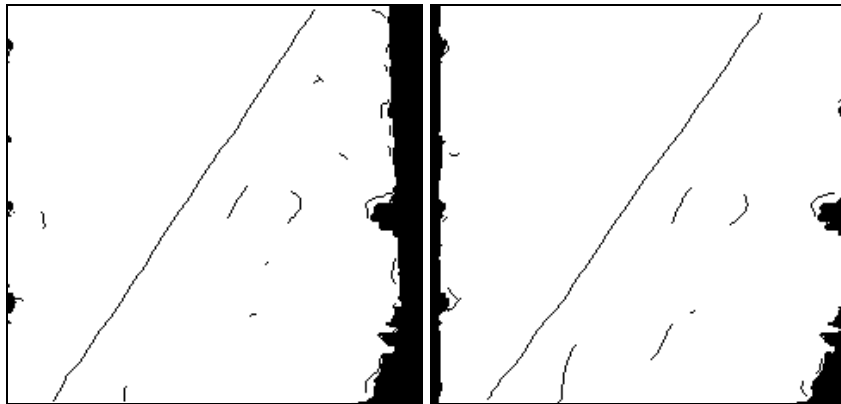
Figure 5.52: Suspension Bridge — Disparity Surface.



(a) Depth Discontinuities and Unknown Regions.



(b) Orientation Discontinuities.



(c) Combined Discontinuities.

Figure 5.53: Suspension Bridge — Surface Features.

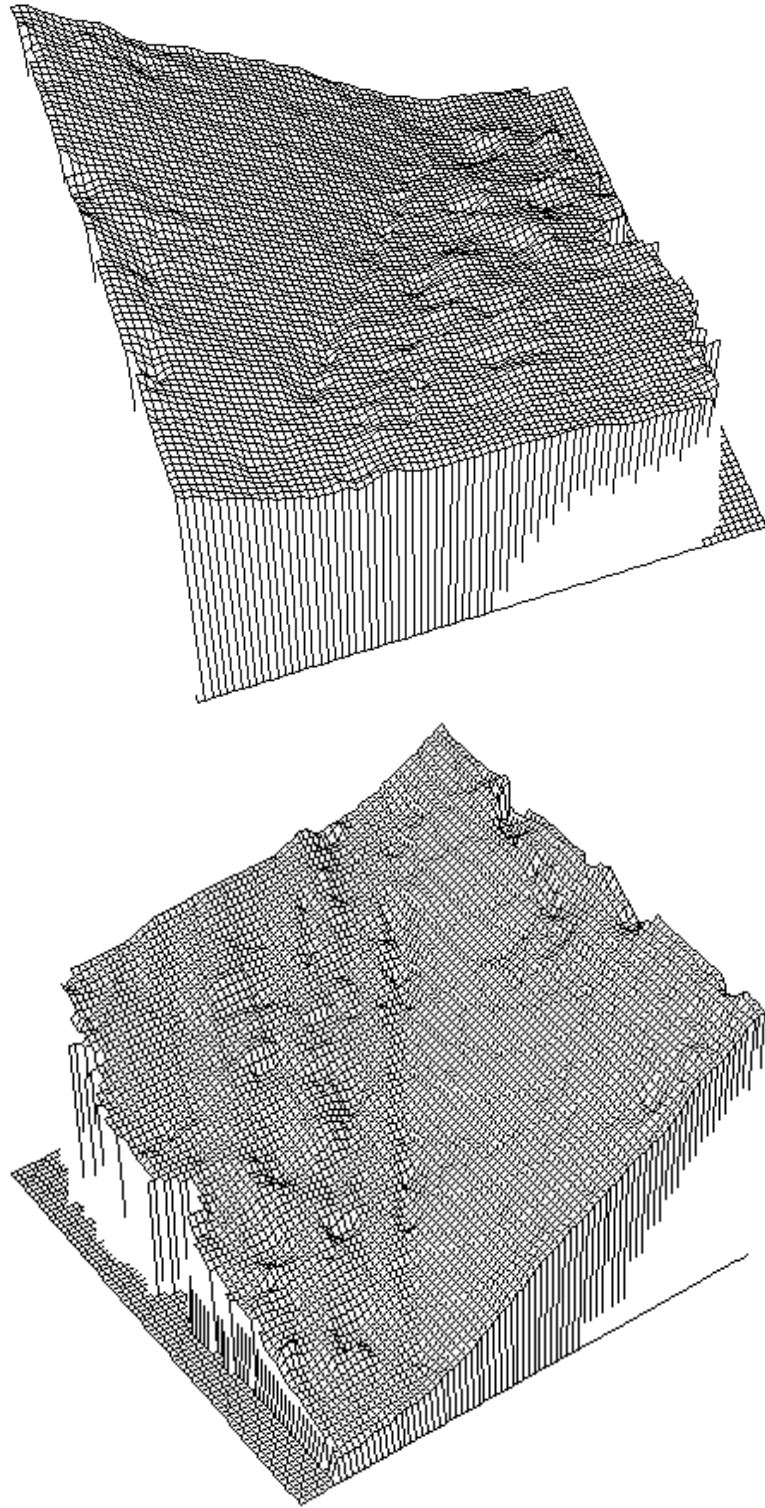


Figure 5.54: Suspension Bridge — 3-D Plot of the Integrated Results.

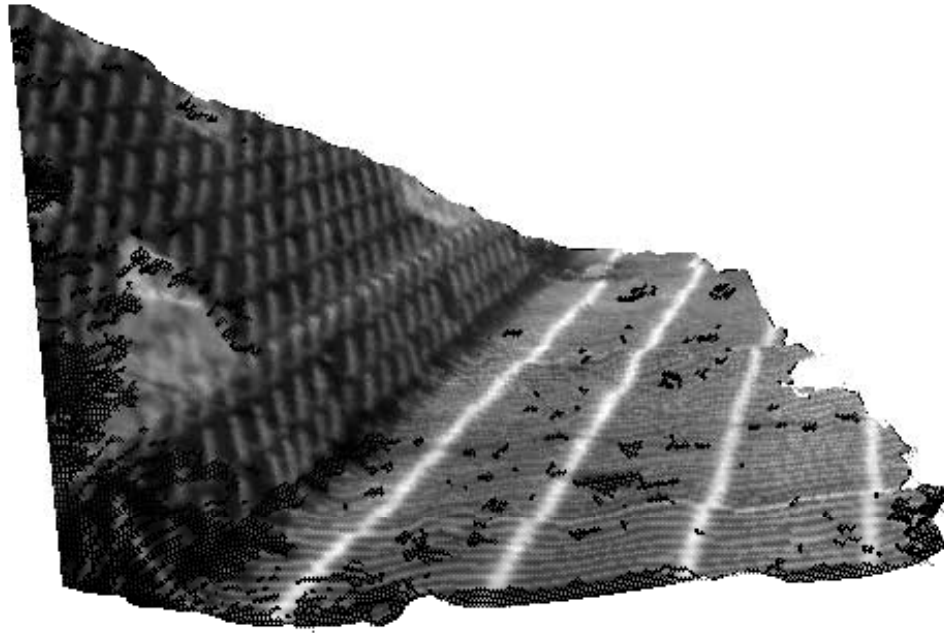


Figure 5.55: Suspension Bridge — 3-D Rendered View of the Integrated Results.

the regions near the shadow edge were matched to the more distant surface, but these spots were removed in favor of a single more globally acceptable surface. Figures 5.54 and 5.55 show the reconstruction of the scene.

Chapter 6

Complexity, Run Times and Error Analysis

6.1 Complexity Analysis

In this section we present an partial analysis of the complexity of the algorithm presented in chapter 4. Unless otherwise stated, we perform a worse case analysis. We begin after the manual portions have been completed and the stereo images are aligned. In the following examples, all of the times are for a serial implementation written in Lisp and running on a Symbolics 3650, with a floating-point accelerator, operating under Genera 7.2. To put the runtimes in perspective, table 6.1 shows the percentage of the overall processing time used by each of the following steps.

6.1.1 Local Variation Estimate

Assume that the stereo images are rectangular and of the same size, with a height of H rows and a width of W columns. The variation estimate examines the pixels in a local 3×5 window. Each step is approximately constant time with the only variation being the presence or absence of valid data (*e.g.* near the edge), thus the serial complexity is $\mathcal{O}(H \cdot W)$. On the connection machine, the summation over the windows is limited only by the communication between the processors across the entire image leaving a parallel complexity of $\mathcal{O}(\log H \cdot \log W)$. Table 6.2 shows the serial runtimes for the local variation estimate for each of the example image pairs.

6.1.2 Estimate from Lower Level

When an estimate pair is expanded from a coarser level of the image pyramid, all values are simply doubled and offset by a constant, thus one pass over the expanded image is sufficient for the serial machine giving a complexity of $\mathcal{O}(H \cdot W)$ and yields a constant processing time on the Connection Machine. Table 6.3 shows the serial runtimes for those passes which made use of a coarser level of the pyramid.

Processing Step	Percent Runtime
Local Variation Estimate	2.02%
Pyramid Estimate	0.07%
Correlation	35.70%
Peak Extraction	2.93%
Initial Disparity	13.09%
Order Reversal	0.24%
Viewpoint Constraint	0.05%
Singleton Removal	0.29%
Area Interpolation	11.93%
Edge Extraction	4.58%
Area/Feature Integration	27.95%
Other Processing	0.43%

Table 6.1: Breakdown of Overall Timing.

6.1.3 Correlation

The correlation complexity is similar to the local variation except that it has a local window size of 9×9 pixels and it is applied once for each disparity value. The range of disparity $D = |L_{\max} - L_{\min}| + 5$ (inclusive of the range plus 2 pixels of padding at the extremes) which gives a complexity of $\mathcal{O}(H \cdot W \cdot D)$ on a the serial machine and $\mathcal{O}(\log H \cdot \log W \cdot D)$ on the Connection Machine. Table 6.4 shows the serial runtimes for the correlation of each of the stereo pairs. The Renault Part was also processed on the Connection Machine and took 6m 51s to process the $251 \times 256 \times 50$ volume.

6.1.4 Peak Extraction

Once the correlation volume is available, the peak extraction marks those points which meet the requirements for a peak as discussed in section 4.3.3 on page 41. For each row of the volume (H) which corresponds to a slice across the image, the routine makes two passes, first for the left view of the data and then for the right view. Both of these are the same as the image width or W and the depth of the volume from each of these views is D . Therefore the serial complexity is $\mathcal{O}(H \cdot W \cdot D)$, while on the Connection Machine, in order to simplify the pass through the volume, it is better to limit the maximum number of peaks that can exist from any one viewpoint to some fixed number which is small enough to allow the locations and size of the peaks to be retained in the small amount of local memory available. Since there is

Image Pair	<i>H</i>	<i>W</i>	Run Time
Renault Part	62	64	28s
	125	128	1m 56s
	251	256	7m 47s
Wedding Cake	32	32	7s
	64	64	30s
	128	128	2m 04s
Books	122	122	1m 41s
	244	244	6m 48s
	488	488	26m 51s
Jussieu	64	64	40s
	128	128	1m 59s
	256	256	8m 12s
Blocks	73	73	40s
	147	147	2m 45s
	295	295	10m 39s
Lubéron	64	64	33s
	128	128	2m 09s
	256	256	7m 50s
Pentagon	128	128	2m 04s
	256	256	7m 55s
	512	512	31m 37s
Power Plant	64	64	31s
	128	128	2m 04s
	256	256	7m 46s
Fruit Scene	58	64	26s
	116	128	1m 50s
	232	256	7m 01s
Quarry Wall	60	60	26s
	120	120	1m 47s
	240	240	6m 53s
Bridge	57	60	24s
	115	120	1m 41s
	231	240	6m 36s

Table 6.2: Local Variation Estimate Times.

Image Pair	H	W	Run Time
Renault Part	125	128	6.4s
	251	256	11.7s
Wedding Cake	64	64	4.0s
	128	128	5.5s
Books	244	244	10.6s
	488	488	30.5s
Jussieu	128	128	5.6s
	256	256	13.4s
Blocks	147	147	5.9s
	295	295	13.1s
Lubéron	128	128	4.5s
	256	256	10.9s
Pentagon	256	256	10.9s
	512	512	33.9s
Power Plant	128	128	5.6s
	256	256	11.6s
Fruit Scene	116	128	5.6s
	232	256	11.9s
Quarry Wall	120	120	7.6s
	240	240	9.7s
Bridge	115	120	5.2s
	231	240	10.0s

Table 6.3: Pyramid Estimate Times.

Image Pair	H	W	D	Serial Run Time
Renault Part	62	64	17	3m 47s
	125	128	28	25m 10s
	251	256	50	3h 01m 23s
Wedding Cake	32	32	9	32s
	64	64	13	3m 15s
	128	128	21	21m 08s
Books	122	122	22	12m 15s
	244	244	39	1h 26m 30s
	488	488	73	10h 47m 26s
Jussieu	64	64	23	5m 08s
	128	128	35	33m 04s
	256	256	61	3h 51m 48s
Blocks	73	73	23	6m 40s
	147	147	42	47m 15s
	295	295	80	5h 57m 55s
Lubéron	64	64	10	2m 35s
	128	128	15	15m 39s
	256	256	20	1h 40m 44s
Pentagon	128	128	10	9m 49s
	256	256	14	54m 13s
	512	512	22	5h 38m 45s
Power Plant	64	64	13	3m 00s
	128	128	19	17m 46s
	256	256	33	2h 12m 48s
Fruit Scene	58	64	12	2m 33s
	116	128	17	14m 31s
	232	256	29	1h 46m 47s
Quarry Wall	60	60	17	3m 25s
	120	120	29	23m 21s
	240	240	51	2h 44m 57s
Bridge	57	60	12	2m 21s
	115	120	19	14m 58s
	231	240	31	1h 38m 35s

Table 6.4: Correlation Times.

Image Pair	H	W	D	Run Time
Renault Part	62	64	17	20s
	125	128	28	1m 53s
	251	256	50	14m 35s
Wedding Cake	32	32	9	3s
	64	64	13	19s
	128	128	21	2m 46s
Books	122	122	22	1m 30s
	244	244	39	10m 03s
	488	488	73	1h 35m 05s
Jussieu	64	64	23	26s
	128	128	35	2m 15s
	256	256	61	15m 16s
Blocks	73	73	23	34s
	147	147	42	3m 32s
	295	295	80	33m 33s
Lubéron	64	64	10	20s
	128	128	15	1m 07s
	256	256	20	6m 53s
Pentagon	128	128	10	50s
	256	256	14	3m 55s
	512	512	22	22m 45s
Power Plant	64	64	13	19s
	128	128	19	1m 34s
	256	256	33	10m 01s
Fruit Scene	58	64	12	14s
	116	128	17	1m 09s
	232	256	29	6m 19s
Quarry Wall	60	60	17	19s
	120	120	29	1m 34s
	240	240	51	10m 01s
Bridge	57	60	12	19s
	115	120	19	1m 22s
	231	240	31	7m 24s

Table 6.5: Peak Extraction Times.

usually only 0 to 5 peaks selected, this is not too much of a restriction, and gives a complexity of $\mathcal{O}(D)$, one pass through the disparity. Table 6.5 gives the serial times for the peak extraction of the example image pairs.

6.1.5 Disparity Estimate

The selection of the initial, raw, estimate is done by a single pass over the peaks. Since the peaks are stored in a volume similar to the correlation and the peaks are subsequently sorted, the complexity is $\mathcal{O}(H \cdot W \cdot D \cdot \log D)$ on a serial machine and is related to the number of peaks retained on the parallel machine, giving a worse case of $\mathcal{O}(D)$. However, since the number of peaks is variable, the $\log D$ term represents a worse case that is rarely met in real-world scenes. The actual times, shown in table 6.6, vary widely due to the differing actual number of peaks found.

6.1.6 Order Reversal

The order-reversal also simply traverses the planar image twice, once to mark the ordering and a second time to remove reversals. The complexity, then is simply $\mathcal{O}(H \cdot W)$ on a serial machine and is performed in $\mathcal{O}(\log W)$ on a parallel machine. The non-constant time on the parallel machine is due to the need to communicate the ordering across the scanline. The actual time (table 6.7) varies considerably from the worse case since there are two paths through the algorithm; one for constant disparity and one for where the disparity changes. When the disparity surface is constant, the processing is much faster.

6.1.7 Multiple Viewpoint Constraint

The **two-view** constraint, limiting the disparity estimate to the estimated matches that are in agreement, is a 2-D operation over the height and width of each view, giving a complexity of $\mathcal{O}(H \cdot W)$ on a serial machine and using constant time on the parallel. Table 6.8 shows the serial runtimes for each of the example stereo pairs.

6.1.8 Singleton Removal

The last constraint is the removal of isolated edgels. This is a very local process which is constant on the parallel machine (involving only the neighboring processors) and a simple scan over the image for the serial implementation, giving a complexity of $\mathcal{O}(H \cdot W)$. The serial runtimes for the singleton removal are shown in table 6.9.

6.1.9 Interpolation

The interpolation involves three passes using slightly different constraints each time. The first pass is tightly constrained and uses information local the existing “seed” points, giving a complexity is $\mathcal{O}(H \cdot W)$. The second pass allows some freedom to

Image Pair	<i>H</i>	<i>W</i>	<i>D</i>	Run Time
Renault Part	62	64	17	2s
	125	128	28	11m 58s
	251	256	50	1h 12m 39s
Wedding Cake	32	32	9	1s
	64	64	13	21s
	128	128	21	7m 23s
Books	122	122	22	5m 22s
	244	244	39	35m 36s
	488	488	73	3h 08m 28s
Jussieu	64	64	23	3s
	128	128	35	12m 19s
	256	256	61	3h 49m 32s
Blocks	73	73	23	4s
	147	147	42	49m 59s
	295	295	80	3h 28m 48s
Lubéron	64	64	10	2s
	128	128	15	1m 49s
	256	256	20	10m 03s
Pentagon	128	128	10	6s
	256	256	14	4m 28s
	512	512	22	1h 14m 57s
Power Plant	64	64	13	2s
	128	128	19	8m 43s
	256	256	33	57m 06s
Fruit Scene	58	64	12	2s
	116	128	17	4m 24s
	232	256	29	24m 27s
Quarry Wall	60	60	17	2s
	120	120	29	4m 10s
	240	240	51	32m 36s
Bridge	57	60	12	2s
	115	120	19	1m 52s
	231	240	31	30m 24s

Table 6.6: Initial Disparity Times.

Image Pair	<i>H</i>	<i>W</i>	Run Time
Renault Part	62	64	2.1s
	125	128	7.8s
	251	256	28.8s
Wedding Cake	32	32	0.8s
	64	64	4.4s
	128	128	8.2s
Books	122	122	6.4s
	244	244	59.1s
	488	488	5m 29.1s
Jussieu	64	64	2.6s
	128	128	7.7s
	256	256	1m 52.2s
Blocks	73	73	2.6s
	147	147	8.7s
	295	295	1m 25.3s
Lubéron	64	64	2.2s
	128	128	7.6s
	256	256	45.4s
Pentagon	128	128	7.4s
	256	256	25.8s
	512	512	5m 22.7s
Power Plant	64	64	2.1s
	128	128	35.1s
	256	256	28.9s
Fruit Scene	58	64	2.2s
	116	128	21.4s
	232	256	1m 18.8s
Quarry Wall	60	60	2.0s
	120	120	6.9s
	240	240	25.1s
Bridge	57	60	2.2s
	115	120	6.0s
	231	240	24.1s

Table 6.7: Order Reversal Times.

Image Pair	<i>H</i>	<i>W</i>	Run Time
Renault Part	62	64	0.7s
	125	128	2.6s
	251	256	9.7s
Wedding Cake	32	32	0.3s
	64	64	0.8s
	128	128	2.6s
Books	122	122	2.5s
	244	244	9.6s
	488	488	35.2s
Jussieu	64	64	0.9s
	128	128	2.5s
	256	256	9.9s
Blocks	73	73	1.0s
	147	147	3.3s
	2950	295	12.7s
Lubéron	64	64	0.8s
	128	128	2.7s
	256	256	10.1s
Pentagon	128	128	2.9s
	256	256	10.2s
	512	512	38.8s
Power Plant	64	64	0.8s
	128	128	2.6s
	256	256	9.5s
Fruit Scene	58	64	0.8s
	116	128	2.8s
	232	256	9.0s
Quarry Wall	60	60	0.7s
	120	120	2.4s
	240	240	8.3s
Bridge	57	60	0.7s
	115	120	2.3s
	231	240	8.3s

Table 6.8: Viewpoint Constraint Times.

Image Pair	H	W	Run Time
Renault Part	62	64	3.6s
	125	128	14.6s
	251	256	58.0s
Wedding Cake	32	32	1.0s
	64	64	3.8s
	128	128	14.8s
Books	122	122	14.1s
	244	244	56.9s
	488	488	3m 45.8s
Jussieu	64	64	3.9s
	128	128	14.9s
	256	256	1m 00.1s
Blocks	73	73	4.8s
	147	147	19.1s
	295	295	1m 15.6s
Lubéron	64	64	3.8s
	128	128	15.0s
	256	256	59.7s
Pentagon	128	128	15.4s
	256	256	59.6s
	512	512	3m 54.4s
Power Plant	64	64	3.7s
	128	128	15.3s
	256	256	58.0s
Fruit Scene	58	64	3.9s
	116	128	14.1s
	232	256	55.8s
Quarry Wall	60	60	3.4s
	120	120	13.0s
	240	240	50.9s
Bridge	57	60	3.2s
	115	120	12.6s
	231	240	49.8s

Table 6.9: Singleton Removal Times.

Image Pair	H	W	d	Run Time	Run Time	Run Time
Renault Part	62	64	2	26s	29s	28s
	125	128	3	3m 13s	2m 03s	1m 55s
	251	256	4	9m 41s	9m 22s	7m 25s
Wedding Cake	32	32	2	8s	8s	8s
	64	64	3	33s	33s	31s
	128	128	4	2m 37s	2m 16s	1m 53s
Books	122	122	2	1m 45s	2m 23s	2m 18s
	244	244	3	10m 47s	8m 42s	7m 03s
	488	488	4	40m 45s	39m 30s	32m 13s
Jussieu	64	64	2	25s	31s	29s
	128	128	3	1m 27s	2m 23s	1m 46s
	256	256	4	12m 09s	9m 39s	6m 08s
Blocks	73	73	2	32s	36s	35s
	147	147	3	2m 26s	2m 12s	2m 05s
	295	295	4	15m 53s	15m 32s	7m 18s
Lubéron	64	64	2	31s	34s	32s
	128	128	3	1m 59s	2m 14s	2m 04s
	256	256	4	9m 14s	8m 54s	8m 03s
Pentagon	128	128	2	2m 03s	2m 12s	2m 07s
	256	256	3	8m 42s	8m 47s	8m 41s
	512	512	4	38m 56s	39m 57s	30m 45s
Power Plant	64	64	2	25s	30s	29s
	128	128	3	2m 16s	1m 54s	1m 50s
	256	256	4	9m 38s	10m 14s	6m 50s
Fruit Scene	58	64	2	27s	33s	30s
	116	128	3	1m 55s	1m 54s	1m 48s
	232	256	4	7m 57s	7m 50s	7m 01s
Quarry Wall	60	60	2	24s	30s	28s
	120	120	3	1m 29s	1m 48s	1m 45s
	240	240	4	9m 02s	9m 45s	6m 16s
Bridge	57	60	2	26s	28s	26s
	115	120	3	1m 37s	1m 47s	1m 41s
	231	240	4	6m 39s	6m 54s	6m 15s

Table 6.10: Area Interpolation Times.

select a match within some window of allowed local gradient. If we let the dimension of this window be d , we have a worse-case complexity of $\mathcal{O}(d \cdot H \cdot W)$. The last pass uses a local median filter within a 3×3 window, giving it a complexity of $\mathcal{O}(H \cdot W)$. Table 6.10 gives the serial runtimes for the area interpolation for each of the example image pairs.

6.1.10 Edge Extraction

The edge extraction may be performed by any of a number of off-the-shelf processes. We have used USC's LINEAR [62] in all of these examples. This process is complex, and varies with the number of edges found, but the worst case complexity is $\mathcal{O}(H \cdot W)$. The serial runtimes for LINEAR are shown in table 6.11.

6.1.11 Integration

The integration is performed by a simple linear pass over the disparity image two smoothing steps. The first interpolates through the quantized image and its complexity varies with the size of the image and the number of actual disparity steps that exist and the second smoothing is performed by a cascade of convolutions with a 3×3 mask. The number of convolutions is kept constant (20 at each level), so the overall complexity becomes $\mathcal{O}(H \cdot W \cdot D)$. The serial runtimes for the example images are shown in table 6.12.

6.1.12 Overall Complexity

The overall complexity is in practice simpler than much of the above detailed examples, since several of the parameters are fixed. The only factors that vary for the images shown in this thesis are H , W and D . The complexity is given by $\mathcal{O}(H \cdot W \cdot D)$, which has been used to estimate actual times for different images to within a few minutes. Most of the processing in this algorithm is inherently local, however, and can be performed in constant time, except for the search for the strongest peak of the disparity which is $\mathcal{O}(D)$. The running time may be estimated as $H \cdot W \cdot D \cdot 12 \text{ milliseconds/pixel}$.³

Typical times for a serial implementation are therefore about 1 hour for a 128×128 image with 25 pixels total disparity, or about 5 hours for a 256×256 image with 25 pixels total disparity. Table 6.12 shows the overall runtimes for each of the examples.

Image Pair	<i>H</i>	<i>W</i>	Run Time
Renault Part	62	64	1 <i>m</i> 03 <i>s</i>
	125	128	4 <i>m</i> 03 <i>s</i>
	251	256	16 <i>m</i> 09 <i>s</i>
Wedding Cake	32	32	25 <i>s</i>
	64	64	1 <i>m</i> 08 <i>s</i>
	128	128	4 <i>m</i> 28 <i>s</i>
Books	122	122	3 <i>m</i> 19 <i>s</i>
	244	244	13 <i>m</i> 17 <i>s</i>
	488	488	55 <i>m</i> 33 <i>s</i>
Jussieu	64	64	1 <i>m</i> 00 <i>s</i>
	128	128	3 <i>m</i> 39 <i>s</i>
	256	256	16 <i>m</i> 13 <i>s</i>
Blocks	73	73	55 <i>s</i>
	147	147	3 <i>m</i> 31 <i>s</i>
	295	295	16 <i>m</i> 05 <i>s</i>
Lubéron	64	64	58 <i>s</i>
	128	128	3 <i>m</i> 46 <i>s</i>
	256	256	16 <i>m</i> 55 <i>s</i>
Pentagon	128	128	3 <i>m</i> 46 <i>s</i>
	256	256	16 <i>m</i> 12 <i>s</i>
	512	512	1 <i>h</i> 15 <i>m</i> 51 <i>s</i>
Power Plant	64	64	59 <i>s</i>
	128	128	3 <i>m</i> 36 <i>s</i>
	256	256	15 <i>m</i> 22 <i>s</i>
Fruit Scene	58	64	56 <i>s</i>
	116	128	3 <i>m</i> 33 <i>s</i>
	232	256	14 <i>m</i> 29 <i>s</i>
Quarry Wall	60	60	52 <i>s</i>
	120	120	3 <i>m</i> 09 <i>s</i>
	240	240	11 <i>m</i> 24 <i>s</i>
Bridge	57	60	50 <i>s</i>
	115	120	3 <i>m</i> 15 <i>s</i>
	231	240	13 <i>m</i> 32 <i>s</i>

Table 6.11: Edge Extraction Times.

Image Pair	<i>H</i>	<i>W</i>	<i>D</i>	Run Time
Renault Part	62	64	17	4m 05s
	125	128	28	20m 20s
	251	256	50	1h 45m 56s
Wedding Cake	32	32	9	1m 09s
	64	64	13	6m 16s
	128	128	21	40m 43s
Books	122	122	22	18m 40s
	244	244	39	1h 39m 14s
	488	488	73	8h 48m 22s
Jussieu	64	64	23	3m 39s
	128	128	35	15m 17s
	256	256	61	1h 28m 31s
Blocks	73	73	23	6m 19s
	147	147	42	32m 20s
	295	295	80	2h 46m 18s
Lubéron	64	64	10	4m 00s
	128	128	15	18m 48s
	256	256	20	1h 24m 20s
Pentagon	128	128	10	19m 03s
	256	256	14	1h 27m 26s
	512	512	22	6h 05m 21s
Power Plant	64	64	13	3m 58s
	128	128	19	19m 01s
	256	256	33	1h 23m 33s
Fruit Scene	58	64	12	4m 20s
	116	128	17	17m 27s
	232	256	29	1h 27m 37s
Quarry Wall	60	60	17	3m 25s
	120	120	29	14m 45s
	240	240	51	1h 12m 29s
Bridge	57	60	12	4m 47s
	115	120	19	17m 50s
	231	240	31	1h 12m 30s

Table 6.12: Area/Feature Integration Times.

Image Pair	<i>H</i>	<i>W</i>	<i>D</i>	Run Time	Total Time
Renault Part	62	64	17	12m 26s	8h 53m 46s
	125	128	28	1h 17m 02s	
	251	256	50	7h 24m 18s	
Wedding Cake	32	32	9	3m 29s	1h 48m 51s
	64	64	13	15m 02s	
	128	128	21	1h 30m 20s	
Books	122	122	22	54m 03s	34h 40m 52s
	244	244	39	4h 57m 16s	
	488	488	73	28h 49m 32s	
Jussieu	64	64	23	13m 37s	12h 10m 48s
	128	128	35	1h 18m 42s	
	256	256	61	10h 38m 29s	
Blocks	73	73	23	18m 43s	17h 09m 49s
	147	147	42	2h 32m 33s	
	295	295	80	14h 18m 33s	
Lubéron	64	64	10	11m 38s	5h 38m 53s
	128	128	15	54m 49s	
	256	256	20	4h 32m 26s	
Pentagon	128	128	10	47m 12s	22h 44m 16s
	256	256	14	3h 39m 05s	
	512	512	22	18h 17m 59s	
Power Plant	64	64	13	11m 48s	7h 08m 58s
	128	128	19	1h 03m 50s	
	256	256	33	5h 53m 20s	
Fruit Scene	58	64	12	11m 27s	5h 52m 36s
	116	128	17	53m 28s	
	232	256	29	4h 47m 41s	
Quarry Wall	60	60	17	11m 11s	6h 51m 10s
	120	120	29	58m 20s	
	240	240	51	5h 41m 39s	
Bridge	57	60	12	11m 20s	5h 27m 16s
	115	120	19	50m 23s	
	231	240	31	4h 25m 33s	

Table 6.13: Overall Run Times.

Comparison	Agreement within Δ			σ
	$\Delta \leq 1$	$1 < \Delta \leq 2$	$2 < \Delta$	
Human-vs-Feature	66.2%	14.8%	19.0%	3.42
Human-vs-Area	49.5%	36.2%	14.3%	0.83

Table 6.14: Comparison of Matching Results.

6.2 Error Analysis

To obtain an estimate of how well the matching performed we compare the results of a human operator, a feature-based matcher and the area-based matcher at 1675 points along the edges in the Renault Part scene in table 6.14. This provides some quantitative basis for the actual matches beyond the empirical evidence for the relative disparity changes provided by the discontinuity extraction. While the edge-based matches are closer to the human selected ones, when correct, the standard deviation is much larger since when it is wrong, it is off by much more than the area based matches. The standard deviation for repeated matches, at 300 scattered points, by the human operator was 0.26 pixels.

Chapter 7

Conclusions and Further Research

7.1 Summary and Conclusions

Our goal was to generate a dense disparity map from passively acquired stereo pairs which is sufficiently accurate to allow the extraction of surface discontinuities and the labelling of occluded areas. It was also important to suppress incorrect matches as much as possible, even at the expense of losing some correct ones. We have produced an approach which is able to meet these goals, and an implementation of this approach which provides impressive results on several images from different domains. More importantly, it demonstrates the validity of some important observations about stereo vision and shows some promising areas for further research.

In particular, we show one use of those edgels which are parallel to the epipolar lines and, along with all edgels, serve as boundaries to limit the search for discontinuities in depth or orientation. We have also demonstrated an important paradigm in stereo matching: the agreement by two views for each disparity value to resolve the ambiguity and verify the matches as well as constraining the interpolation of the surface through the disparity space.

These have allowed us to produce a system which infers more than the disparity at each point, as we can also mark boundary contours which correspond to depth discontinuities, and thereby label points as to whether or not they are occluded. In addition, we have made progress in locating orientation-discontinuities which should allow image segmentation based on surface properties detected from passive stereo.

This methodology is robust. When there are no edge features, it acts as a good area-based process, and when there is no texture, as a feature-based system without surface interpolation. Furthermore, many existing systems may be easily modified to search for discontinuities in the manner that we suggest, by adding a strong preference for the existence of the discontinuity contour (either depth or orientation) to occur at edgels or along edges.

One novel aspect of the Stereo Vision System (SVS) is that we maintain pixel mappings from each of the views. This allows us to map more than just a one-to-one match between pixels as was shown in figure 4.15. We can have a many-to-one match, which is very helpful in the case of a smooth surface, as opposed to a flat surface constraint. Using either view alone would be insufficient.

7.2 Problems

There are several problems or deficiencies that still exist in this implementation. Most of these are minor in the sense that their solution would be necessary for a production system, but would not add much in terms of theoretical advancement.

- First, depth discontinuities are detected as a change in disparity between 4-connected pixels which is greater than some selected threshold value (we have used 3 pixels in the examples in this paper). This causes a small amount of smoothing as the depth discontinuity terminates at an orientation discontinuity which is best illustrated by the plot in figure 5.10(a), where the books meet the table. In order to improve the extraction of the discontinuity, some sort of hysteresis should be added to extend the discontinuity to where the adjacent surfaces meet.
- When an order reversal occurs, there are usually two patches involved: for instance, in the strings **ABCD** and **ACBD** the letters “B” and “C” are the reversed pair. One or both of the patches must be removed, since we do not allow the reversals to occur within the fusion region. We have chosen to remove the **nearer** of the two since this is usually the incorrect match. Occasionally, this yields an incorrect result and a better solution would be to remove the one with less global support: the smaller patch or the one which is the most separated from its neighbors.
- When the occluding edges meet at acute angles, neither the adaptive smoothing nor the snakes are able to make the necessary sharp bend. This has already led to some follow-on work by Menet *et al.* [55] on breakable splines, which “sense” the corner and adapt themselves so that the “blurring” is fully and correctly removed.
- The orientation discontinuities, as found by SVS, are noisy and poorly located. Again, snakes seem an appropriate tool to adapt to allow better accuracy in locating these features, and by making use of the multiple viewpoint paradigm so that only those discontinuities which occur in both scenes are allowed (at

least in the initial estimate). Recently there has been some work by Parvin and Medioni[65] on extending orientation discontinuities in range data which could be applied to these estimates and which would then yield a complete segmentation of the object ready for a higher-level matcher.

- Finally, we only find point matches to pixel accuracy. It is possible to improve this to subpixel accuracy.

7.3 Suggestions for Further Research

What new ideas for study have been found in the course of this research? Well, there are still several hard problems remaining, and each of these represents an important area in which further research may provide significant new advances.

So far we have only used the simplest integration of area and feature-based methods to demonstrate the feasibility of a more complete system. Although we have used intensity edges in this exposition, texture edges, such as those produced by Perry and Lowe [66] or Malik and Perona [46], would be a better feature for most scenes. But, even if there now existed a reliable feature extractor, there would still be the control problem of selecting the proper scale of the texture which is appropriate for a given part of the scene.

So the current (serial) implementation is much too slow: It currently requires about 8 hours for a typical $256 \times 256 \times 8$ pixel scene. Half of this time is spent performing the cross-correlation. Experiments with a Connection Machine show that this can be reduced to less than seven minutes. We have, with one exception, selected processes which are local in scope so that the routine may be implemented on a parallel-architecture machine. That one exception is the search for the peaks of the cross-correlation, which is bound by the depth of the fusion interval.

The coarse-to-fine global optimization strategy implemented in SVS usually works fine and is often not needed at all. But a multi-level cooperative process should be superior to either, since it would be less likely for a mistake at one level to be propagated on “blindly” so that the correct answer is excluded from the search space. This is more along the lines of the strategy proposed by Mayhew and Frisby [53]. Another approach to finding the globally optimal disparity surfaces is to search for them in the disparity space volume generated by the correlation. The peaks form rough 2-D sheets through this space as can be seen in cross section in figure 4.3(b).

Reference List

- [1] R. D. Arnold. Local context in matching edges for stereo vision. In *Proceedings of the DARPA Image Understanding Workshop*, pages 65–72, Boston, Massachusetts, May 1978.
- [2] R. D. Arnold. *Automated Stereo Perception*. PhD thesis, Stanford University, Stanford, California, March 1983. Technical Report AIM-351 and STAN-CS-83-961.
- [3] N. Ayache, O. D. Faugeras, B. Faverjon, and G. Toscani. Matching depth maps obtained by passive stereo. In *Proceedings of the IEEE Workshop on Computer Vision: Representation and Control*, pages 197–204, Bellaire, Michigan, October 13–16 1985.
- [4] N. Ayache and B. Faverjon. Fast stereo matching of edge segments using prediction and verification of hypotheses. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 662–664, San Francisco, California, June 19–23 1985.
- [5] N. Ayache and B. Faverjon. A fast stereovision matcher based on prediction and recursive verification of hypothesis. In *Proceedings of the IEEE Workshop on Computer Vision: Representation and Control*, pages 27–37, Bellaire, Michigan, October 1985.
- [6] N. Ayache and F. Lustman. Fast and reliable passive trinocular stereovision. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 422–427, London, England, June 1987.
- [7] H. H. Baker. Depth from edge and intensity based stereo. Technical Report AIM-347 and STAN-CS-82-930, Stanford University, Stanford, California, September 1982. Based on the author’s thesis (Ph.D. — Illinois).
- [8] H. H. Baker, T. O. Binford, J. Malik, and J. Meller. Progress in stereo mapping. In *Proceedings of the DARPA Image Understanding Workshop*, pages 327–335, Arlington, Virginia, June 23 1983.

- [9] S. Barnard. A stochastic approach to stereo vision. In *Proceedings of the National Conference on Artificial Intelligence*, pages 676–680, Philadelphia, Pennsylvania, 1986.
- [10] S. Barnard and M. Fischler. Computational stereo. *ACM Computing Surveys*, 14(4):553–572, December 1982.
- [11] S. Barnard and W. Thompson. Disparity analysis of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(4):333–340, July 1980.
- [12] H. Barrow, J. Tenenbaum, R. C. Bolles, and H. C. Wolf. Parametric correspondence and chamfer matching. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 659–663, Cambridge, Massachusetts, August 1977.
- [13] P. Bishop. *Adler's Physiology of the Eye*, chapter Binocular Vision, pages 558–614. The C. V. Mosby Company, St. Louis, 1975. R. Moses (ed.).
- [14] A. Blake. Reconstructing a visible surface. In *Proceedings of the National Conference on Artificial Intelligence*, pages 23–26, University of Texas at Austin, August 6–10 1984. American Association for Artificial Intelligence.
- [15] A. Blake and A. Zisserman. *Visual Reconstruction*. Artificial Intelligence. MIT Press, Cambridge, Massachusetts, 1987.
- [16] R. C. Bolles, H. H. Baker, and D. H. Marimont. Epipolar-plane image analysis: An approach to determining structure from motion. *International Journal of Computer Vision*, 1(1):7–55, 1987.
- [17] T. Boult and J. Kender. Visual surface reconstruction using sparse depth data. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 68–76, Miami Beach, Florida, June 22–26 1986.
- [18] T. E. Boult and L.-H. Chen. Synergistic smooth surface stereo. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 118–122, Tampa, Florida, December 1988.
- [19] A. Brandt. Multi-level adaptive finite element methods: Variational problems. In J. Rice, editor, *Special Topics of Applied Mathematics*, pages 91–128. North-Holland, New York, 1980.
- [20] A. Brandt. Multigrid solvers on parallel computers. In M. Schultz, editor, *Elliptic Problem Solvers*, pages 39–83. Academic Press, New York, 1981.
- [21] R. A. Brooks. Model-based three-dimensional interpretations of two-dimensional images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-5(2):140–150, March 1983.

- [22] J. F. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, November 1986.
- [23] S. D. Cochran and M. Medioni. Implementation of a multiresolution surface reconstruction algorithm. Technical Report ISG-108, University of Southern California, Los Angeles, California, 1985.
- [24] S. Das and N. Ahuja. Integrating multiresolution image acquisition and coarse-to-fine surface reconstruction from stereo. In *Proceedings of the Workshop in Interpretation of 3D Scenes*, pages 9–15, Austin, Texas, November 27–29 1989.
- [25] U. R. Dhond and J. K. Aggarwal. Structure from stereo — a review. *IEEE Transactions on Systems, Man & Cybernetics*, 19(6):1489–1510, November/December 1989.
- [26] M. Drumheller and T. Poggio. On parallel stereo. In *Proceedings of the IEEE Conference on Robotics and Automation*, pages 1439–1448, San Francisco, California, April 1986.
- [27] T.-J. Fan, G. Medioni, and R. Nevatia. Recognizing 3-D objects using surface descriptions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(11):1140–1157, November 1989.
- [28] P. J. Flynn and A. K. Jain. On reliable curvature estimation. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 110–116, San Diego, California, June 1989.
- [29] W. Förstner and Eberhard Gülch. International society of photogrammetry and remote sensing working group III/4: Experimental test A on image matching. Institut für Photogrammetrie, Universität Stuttgart, November 1986.
- [30] D. B. Gennery. *Modeling the Environment of an Exploring Vehicle by Means of Stereo Vision*. PhD thesis, Stanford University, Stanford, California, June 1980. Technical Report AIM-339 and STAN-CS-80-805.
- [31] W. E. L. Grimson. *From Images to Surfaces: A Computational Study of the Human Early Visual System*. Artificial Intelligence. MIT Press, Cambridge, Massachusetts, 1981. Based on the author’s thesis (Ph.D. — MIT).
- [32] M. J. Hannah. *Computer Matching of Areas in Stereo Images*. PhD thesis, Stanford University Computer Science Department, Stanford, California, July 1974. Technical Report STAN-CS-74-438.
- [33] M. J. Hannah. Bootstrap stereo. In *Proceedings of the DARPA Image Understanding Workshop*, pages 201–208, College Park, Maryland, April 1980.

- [34] M. J. Hannah. SRI's baseline stereo system. In *Proceedings of the DARPA Image Understanding Workshop*, pages 149–155, Miami Beach, Florida, December 1985.
- [35] M. Herman and T. Kanade. The 3D MOSAIC scene understanding system: Incremental reconstruction of 3D scenes from complex images. Technical Report CMU-CS-84-102, Carnegie-Mellon University, Pittsburgh, PA, February 1984.
- [36] W. Hoff and N. Ahuja. Surfaces from stereo. In *Proceedings of the DARPA Image Understanding Workshop*, pages 98–106, Miami Beach, Florida, December 9–10 1985.
- [37] W. Hoff and N. Ahuja. Extracting surfaces from stereo-images: An integrated approach. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 284–294, London, England, June 1987.
- [38] W. Hoff and N. Ahuja. Surfaces from stereo: Integrating feature matching, disparity estimation, and contour detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(2):121–136, February 1989.
- [39] J.-Y. Jou and A. C. Bovik. Improved initial approximation and intensity-guided discontinuity detection in visible-surface reconstruction. *Computer Vision, Graphics, and Image Processing*, 47(3):292–326, September 1989.
- [40] B. Julesz. *Foundations of Cyclopean Perception*. The University of Chicago Press, Chicago, 1971.
- [41] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 259–268, London, England, June 1987.
- [42] D. J. Langridge. Curve encoding and the detection of discontinuities. *Computer Graphics and Image Processing*, 20:58–71, 1982.
- [43] H. S. Lim. *Stereo Vision: Structural Correspondence and Curved Surface Reconstruction*. PhD thesis, Stanford University, Stanford, California, September 1987.
- [44] H. S. Lim and T. O. Binford. Stereo correspondence: Features and Constraints. In *Proceedings of the DARPA Image Understanding Workshop*, pages 373–38, Miami Beach, Florida, December 9–10 1985.
- [45] H. S. Lim and T. O. Binford. Structural correspondence in stereo vision. In *Proceedings of the DARPA Image Understanding Workshop*, pages 794–808, Cambridge, Massachusetts, April 1988.

- [46] J. Malik and P. Perona. A computational model of texture segmentation. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 326–332, San Diego, California, June 1989.
- [47] E. Mansfield. *The Bending and Stretching of Plates*. Macmillan, New York, 1964.
- [48] D. Marr. *Vision*. W. H. Freeman and Company, 1982.
- [49] D. Marr and T. Poggio. A theory of human stereo vision. Technical Report AI Memo 451, Massachusetts Institute of Technology Artificial Intelligence Laboratory, November 1977.
- [50] D. Marr and T. Poggio. A theory of human stereo vision. *Proc. R. Soc. Lond., B*, Volume 204:301–328, 1979.
- [51] L. Matthies, R. Szeliski, and T. Kanade. Incremental estimation of dense depth maps from image sequences. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 366–374, Ann Arbor, Michigan, June 1988.
- [52] J. E. W. Mayhew. Stereopsis. In O. Braddick and A. Sleight, editors, *Physical and Biological Processing of Images*, pages 204–216. Springer-Verlag, New York, NY, September 27–29 1982. From the Proceedings of an International Symposium Organized by the Rank Prize Funds, London, England.
- [53] J. E. W. Mayhew and J. P. Frisby. Psychophysical and computational studies towards a theory of human stereopsis. *Artificial Intelligence*, 17:349–385, August 1981.
- [54] G. Medioni and R. Nevatia. Segment-based stereo matching. *Computer Vision, Graphics, and Image Processing*, 31:2–18, 1985.
- [55] S. Menet, P. Saint-Marc, and G. Medioni. Active contour models: Overview, implementation and applications. In *IEEE International Conference on Systems, Man, and Cybernetics*, pages 194–199, Los Angeles, California, November 4-7 1990.
- [56] V. Milenkovic and T. Kanade. Trinocular vision using photometric and edge orientation constraints. In *Proceedings of the DARPA Image Understanding Workshop*, pages 163–175, Miami Beach, Florida, December 1985.
- [57] R. Mohan, G. Medioni, and R. Nevatia. Stereo error detection, correction and evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(2):113–120, February 1989.

- [58] R. Mohan and R. Nevatia. Using perceptual organization to extract 3-D structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(11):1121–1139, November 1989.
- [59] H. P. Moravec. *Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover*. PhD thesis, Stanford University, Stanford, California, September 1980. Technical Report AIM-340 and STAN-CS-80-813.
- [60] K. Mori, M. Kidode, and H. Asada. An iterative prediction and correction method for automatic stereocomparison. *Computer Graphics and Image Processing*, 2:393–401, 1973.
- [61] R. Nevatia. Depth measurement by motion stereo. *Computer Graphics and Image Processing*, 5:203–214, 1976.
- [62] R. Nevatia and K. Babu. Linear feature extraction and detection. *Computer Graphics and Image Processing*, 13(3):257–269, July 1980.
- [63] Y. Ohta and T. Kanade. Stereo by two-level dynamic programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(2):139–154, April 1985.
- [64] D. Panton, C. Grosch, D. DeGryse, J. Ozils, A LaBonte, S. Kaufmann, and L. Kirvida. Geometric reference studies. Final Technical Report RADC-TR-81-182, December 1981. Volume 44, Number 12.
- [65] B. Parvin and G. Medioni. A dynamic system for object description and correspondence. USC-IRIS Internal Report, 1990.
- [66] A. Perry and D. G. Lowe. Segmentation of textured images. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 319–323, San Diego, California, June 1989.
- [67] K. Price. Hierarchical matching using relaxation. *Computer Vision, Graphics, and Image Processing*, 34:66–75, 1986.
- [68] L. Quam. Hierarchical warp stereo. In *Proceedings of the DARPA Image Understanding Workshop*, pages 149–155, New Orleans, Louisiana, October 3–4 1984.
- [69] J. J. Rodríguez and J. K. Aggarwal. Stochastic analysis of stereo quantization error. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5):467–470, May 1990.
- [70] P. Saint-Marc, J. S. Chen, and G. Medioni. Adaptive smoothing: A general tool for early vision. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 618–624, San Diego, California, June 1989.

- [71] P. Saint-Marc and G. Medioni. Adaptive smoothing for feature extraction. In *Proceedings of the DARPA Image Understanding Workshop*, pages 1100–1113, Cambridge, Massachusetts, April 1988.
- [72] A. R. de Saint Vincent. A 3D perception system for the mobile robot Hilare. In *Proceedings of the IEEE Conference on Robotics and Automation*, pages 1105–1111, San Francisco, California, April 7–10 1986. Volume 2.
- [73] S. S. Sinha and B. G. Schunck. Discontinuity preserving surface reconstruction. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 229–234, San Diego, California, June 1989.
- [74] D. Terzopoulos. *Multiresolution Computation of Visible-Surface Representations*. PhD thesis, Massachusetts Institute of Technology, Departments of Computer Science and Electrical Engineering, Cambridge, Massachusetts, January 1984.
- [75] D. Terzopoulos. Computing visible surface representations. Technical Report AI Memo 800, Massachusetts Institute of Technology Artificial Intelligence Laboratory, Cambridge, Massachusetts, 1985.
- [76] D. Terzopoulos. The computation of visible-surface representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(4):417–438, July 1988.
- [77] S. Timoshenko and S. Woinowsky-Krieger. *Theory of Plates and Shells*. McGraw-Hill, New York, 1959.
- [78] S. Tsuji, J. Zheng, and M. Asada. Stereo vision of a mobile robot: World constraints for image matching and interpretation. In *Proceedings of the IEEE Conference on Robotics and Automation*, pages 1594–1599, San Francisco, California, April 7–10 1986. Volume 3.
- [79] R. P. Wildes. *On Interpreting Stereo Disparity*. PhD thesis, Massachusetts Institute of Technology Artificial Intelligence Laboratory, Cambridge, Massachusetts, April 1989. Technical Report AI-TR-1112.
- [80] M. Yachida, Y. Kitamura, and M. Kimachi. Trinocular vision: New approach for correspondence problem. In *Proceedings of the International Conference on Pattern Recognition*, pages 1041–1044, Paris, France, October 1986.
- [81] A. L. Yuille and T. Poggio. A generalized ordering constraint for stereo correspondence. Technical Report AI Memo 777, Massachusetts Institute of Technology Artificial Intelligence Laboratory, 1984.

- [82] Y. T. Zhou and R. Chellappa. Stereo matching using a neural network. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, pages 940–943, New York, N.Y., April 11–14 1988.
- [83] Y. T. Zhou, R. Chellappa, and B. K. Jenkins. A novel approach to image restoration based on a neural network. In *Proceedings of the International Conference on Neural Networks*, San Diego, California, June 1987.