A Multiresolution Approach to Range-Guided Stereo Matching

by

Kevin James Tate

BSc, University of Saskatchewan, 1985

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
in the School
of
Computing Science

© Kevin James Tate 1990
SIMON FRASER UNIVERSITY
April 1990

All rights reserved. This thesis may not be
reproduced in whole or in part, by photocopy
or other means, without the permission of the author.
Approval

Name:
  Kevin James Tate

Degree:
  Master of Science

Title of Thesis:
  A Multiresolution Approach to Range-guided Stereo Matching

Examining Committee:
  Dr. Brian Funt, Chairman

Dr. Ze-Nian Li
Senior Supervisor

Dr. Tom Calvert
Supervisor

Dr. Kamal Gupta
Department of Engineering Sciences
Simon Fraser University
External Examiner

April 10, 1990
Date Approved
PARTIAL COPYRIGHT LICENSE

I hereby grant to Simon Fraser University the right to lend my thesis, project or extended essay (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this work for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this work for financial gain shall not be allowed without my written permission.

Title of Thesis/Project/Extended Essay

A Multiresolution Approach To Range-Guided Stereo Matching

Author: ____________________________

(signature)

Kevin James Tate

(name)

April 10, 1990

(date)
Abstract

The acquisition of high-resolution three dimensional data is an important problem in computer vision. Two of the most common techniques for acquiring 3-D data are stereo imagery and laser range finding. Stereo image intensity data can be acquired rapidly at high-resolution but the complexity of determining the correspondence between the features in the stereo image pair and accurately determining depth have lead many vision researchers to turn to other approaches. One common alternative is to use a laser range finder. These devices provide direct but noisy 3-D data within a limited range and at relatively slow acquisition rates. Due to their acquisition rate, researchers reduce the resolution of the acquired data and limit the size of the area scanned.

This thesis describes an approach where stereo image intensity data processing techniques are combined with laser range finding methods. By drawing on previous research results and problems inherent in both types of data and their respective method for depth information derivation, it is shown that these methods are complementary sources of information. This thesis presents research outlining one possible method of combining the two types of data. In this method, edge correspondences are found, and the corresponding depth information along the edge from the range data is used to drive stereo matching of these edges. This initial match reduces the search space of potential edge matches for the remainder of the edges. Finally, surface interpolation is performed, drawing on depth information from both sources of information to develop a more complete, noise-free, and accurate depth map.
Acknowledgements

I would like to particularly thank my supervisor, Dr. Ze-Nian Li for getting me interested in the area of computer vision, for our interesting technical discussions, and his consistent and sound critiques of my ideas and work. I am particularly grateful for his enthusiasm, constant support, and encouragement which kept me on track when my motivation was low or my goals were not clear. I must also thank my future wife Claudia Morawetz, who not only proved to me what the true definition of unfailing confidence is, but who also provided many extremely useful insights and solutions to many problems. Finally, I must thank my friends; those I knew before arriving at SFU and those I have met since, all of whom eased my introduction to Vancouver and made my time here extremely enjoyable.

I would also like to acknowledge the School of Computing Science and Simon Fraser University for providing financial support in the forms of teaching assistantships and Graduate Fellowships. I also appreciated receiving research assistantships from my supervisor and Dr. Robert Hadley.

I would like to thank various people at Michigan State University for the data used in this thesis. Patrick Flynn in particular was very helpful, and graciously provided the registered intensity images for the PRIP (Pattern Recognition and Image Processing) range images.
Table of Contents

Approval ii
Abstract iii
Acknowledgements iv
Table of Contents v
1. Introduction 1
   1.1. Motivation
      1.1.1. Laser Range Imagery and its Weaknesses 1
      1.1.2. Stereo Intensity Imagery 3
   1.2. Proposed Work 4
2. Related Research 6
   2.1. Combination of Intensity and Range Data 6
   2.2. Three-Dimensional Edge Detection 9
   2.3. Stereo Matching 14
   2.4. Surface Interpolation 16
3. A Model for Range-Guided Stereo Matching 19
   3.1. Method Overview 19
   3.2. Proposed Camera Geometry 21
   3.3. Image Data Acquisition 22
4. Implementation of Range-Guided Stereo Matching 26
   4.1. Range Image Edge Detection 26
   4.2. Intensity Pyramid Construction 31
   4.3. Edge Combination 35
   4.4. Stereo Matching 38
   4.5. Hierarchical Dense Depth Map Generation with Boundary Improvement 46
5. System Evaluation 49
   5.1. Image Formation 49
   5.2. Pyramid Construction 58
   5.3. Range Image Edge Detection 60
   5.4. Edge Combination 62
   5.5. Stereo Matching 64
   5.6. Hierarchical Dense Depth Map Generation with Boundary Improvement 71
6. Conclusion 84
   6.1. Future Research Issues 86
Appendix A. Glossary 90
References 94
List of Figures

Figure 1-1: Typical laser system geometry (from [Jarvis 83]) 2
Figure 2-1: The three types of missing parts described in [Chen 89] 8
Figure 2-2: The different types of range edges (from [Besl 86]) 10
Figure 3-1: Order of processing used in the implementation. 20
Figure 3-2: Proposed camera geometry 22
Figure 3-3: Synthetic data creation 23
Figure 4-1: An overview of the order of processing used in the implementation 27
Figure 4-2: Edge detection masks for the calculation of \( f_x \) and \( f_y \). 28
Figure 4-3: The classification of edge gradient into a section number. 29
Figure 4-4: Neighbours used for edge thinning 30
Figure 4-5: The range image data structure 32
Figure 4-6: Intensity pyramid base structure 32
Figure 4-7: Neighbours used for intra-level edge linking by edge section 34
Figure 4-8: Order of search for the prioritised local AND 37
Figure 4-9: Combined image pyramid base structure 38
Figure 4-10: Algorithm for stereo matching with depth as a guide 40
Figure 4-11: Algorithm for stereo matching without depth as a guide 42
Figure 4-12: Stereo matching, example 1: An illustration of retrying matching 43
Figure 4-13: Stereo matching, example 2: An illustration of matching a long edge segment containing a horizontal segment 44
Figure 4-14: Stereo matching, example 3: An illustration of dealing with ambiguity and in setting disparity values for unmatched edge points 45
Figure 4-15: Stereo matching, example 4: An illustration of the role of inter-level figural continuity 47
Figure 5-1: Original range images: the "hand" images. 50
Figure 5-2: Original range images: the "cup" images. 51
Figure 5-3: An example of the problem of generating stereo intensity data. 52
Figure 5-4: Synthetic intensity image for "cup2". 53
Figure 5-5: Reduced resolution range images, shown by gray-level. 54
Figure 5-6: Effects of k-nearest neighbour smoothing on a range image. 55
Figure 5-7: Smoothed versions of the "hand1" and "hand2" images. 56
Figure 5-8: MSU PRIP lab registered range and intensity image of a cone ("cone1"). 56
Figure 5-9: MSU PRIP lab registered range and intensity image of a Y pipe ("wye1"). 57
Figure 5-10: MSU PRIP lab registered range and intensity image of a Y pipe ("wye2"). 57
Figure 5-11: MSU PRIP lab registered range and intensity of a toy (joystick?) ("toy"). 58
Figure 5-12: Top and middle levels of the intensity image pyramid for the "toy" image. 59
Figure 5-13: Bottom-most levels of the intensity image pyramid for the "toy" image. 59
Figure 5-14: Top and middle levels of the intensity image pyramid for the "hand1" image. 60
Figure 5-15: Bottom level of the intensity image pyramid for the "hand1" image.
Figure 5-16: Unthinned edge maps from the top and middle levels for the "hand1" intensity image pyramid.
Figure 5-17: Unthinned edge maps from the bottom level for the "hand1" intensity image pyramid.
Figure 5-18: Thinned edge maps from the top and middle levels for the "hand1" intensity image pyramid.
Figure 5-19: Thinned edge maps from the bottom level for the "hand1" intensity image pyramid.
Figure 5-20: Thinned edge maps from the top and middle levels for the "hand2" intensity image pyramid.
Figure 5-21: Thinned edge maps from the bottom level for the "hand2" intensity image pyramid.
Figure 5-22: Jump edge detection in the "hand1" image.
Figure 5-23: Jump edge detection in the "hand2" image.
Figure 5-24: Jump edge detection in the "bottle" image.
Figure 5-25: Jump and roof edge maps for two artificial test images designed to highlight roof edge detection.
Figure 5-26: Depth values obtained at common edges in the hand images.
Figure 5-27: Depth values obtained at common edges in the "toy" image.
Figure 5-28: Disparity maps obtained for the "hand1" image when stereo matching starts at the 64x64 level of resolution.
Figure 5-29: Disparity maps obtained for the "hand1" image when stereo matching starts at the 128x128 level of resolution.
Figure 5-30: Disparity maps obtained for the "hand2" image when stereo matching starts at the 64x64 level of resolution.
Figure 5-31: Disparity maps obtained for the "hand2" image when stereo matching starts at the 128x128 level of resolution.
Figure 5-32: Disparity maps obtained for the "toy" image when stereo matching starts at the 64x64 level of resolution.
Figure 5-33: Disparity map obtained for the "toy" image when stereo matching is performed without depth as a guide.
Figure 5-34: Disparity maps obtained for the "wyel" image when stereo matching starts at the 128x128 level of resolution.
Figure 5-35: Depth maps for the "hand1" and "hand2" images resulting from hierarchical depth map generation.
Figure 5-36: Depth maps for the "toy" image resulting from boundary improvement and the role of the minimum depth threshold.
Figure 5-37: Depth maps for the "toy" image resulting from boundary improvement and the blurring of depth information.
Figure A-1: Child/parent relationships in a non-overlapped pyramid
Figure A-2: Child/parent relationships in an overlapped pyramid
Chapter 1
Introduction

1.1. Motivation

The acquisition of accurate high-resolution 3-D data is an important problem in computer vision. Two of the most common techniques for acquiring 3-D data are stereo imagery and laser range finding. Both techniques have appealing aspects making them attractive candidates for 3-D data acquisition. Unfortunately, each also has associated problems which make them unattractive as a definitive means of 3-D data acquisition. The goal of this thesis is to examine ways to combine the advantages of these two approaches.

1.1.1. Laser Range Imagery and its Weaknesses

The use of laser range finders for the acquisition of 3-D data is an increasingly common approach in computer vision research. The primary reason for this is that 3-D data is acquired in a well-defined, easily understood and easily implemented manner. There are two main implementations of laser range finders, one based on triangulation, the other on time-of-flight [Besl 85a, Besl 85b, Jarvis 83].

The most common type of range finder is the so-called triangulation range finder, since they are cheaper than time-of-flight range finders and easily implemented. A common configuration for this type of range finder is shown in figure 1-1. In this system, a laser projecting a dot or row of dots is directed onto a mirror, which is swept across a scene. The known orientations of the camera and mirror and the baseline distance between them are sufficient to calculate depth at the instant the camera detects a light spot at a point. The major drawback of these systems is the so-called "missing parts" problem, where a portion of a surface or an entire surface are undetected due to
surface absorption or reflectance, or by directional occlusion\(^1\). As with other triangulation systems, the larger the separation between laser and camera the more accurate will be the ranging, but at the same time, the "missing parts" caused by directional occlusion will be more prevalent. A typical commercially-available triangulation system provides resolution of 128 by 128 points with a depth resolution of approximately one centimetre for objects in the one to four metre range at the rate of one frame per second [Besl85a].

\[ \text{depth} = \text{baseline} \times \tan(\text{orientation}) \]

**Figure 1-1: Typical laser system geometry (from [Jarvis 83])**

Triangulation systems have limited range and resolution. Range is limited by power (and thus cost) of the laser as well as surface curvature and absorption properties since increasing range requires increasing surface planarity to allow detection of the projected dot(s). Resolution of these devices is limited by the power of the laser (higher power is required for greater distance and to project finer, more intense dots) and the accuracy of the stepping motor used to sweep the laser

\(^1\)Portions of a surface undetected due to "unfriendly" surface properties are often called "nulls".
across the image. If a mirror is used to reflect the laser beam (thus allowing the laser to be fixed and the reflected beam from the mirror swept across the scene), then the quality of the mirror is also a factor.

Time-of-flight range finders represent distinct contrasts to triangulation-based range finders. In these systems, a laser is used to bounce a signal off a surface to a receiver which detects the signal and measures its return time. Time-of-flight range finders have distinct advantages over triangulation-based range finders. Their main advantage is that absolute range is immediately available since depth is proportional to the time-of-flight and therefore requires no calculations. In addition, energy source and detection windows can be coaxial; therefore, there need not be any viewpoint perspective problems. And finally, range accuracy can be maintained over depth until reliable signal detection is no longer possible. Unfortunately, time-of-flight range finders are also very expensive since accuracy and speed are determined both by the power of the laser and sensitivity of the receiving device. High-powered lasers are required to project a light wave which is still finely detailed on return to the receiver. The receiving device must be sensitive enough to detect returned signals and be very fast since measurements on the order of the speed of light are required.

This thesis considers triangulation-based range finders due to their larger, and almost exclusive, usage in the research literature.

1.1.2. Stereo Intensity Imagery

The concept of stereo imagery is based on the human 3-D vision mechanism. In computer vision research, stereopsis is most commonly implemented using the lateral camera model which, although not the most accurate camera model, requires the simplest mathematics [Alvertos 89]. The primary advantage of stereo imagery is that image intensity data can be acquired rapidly (many frames per second) and at high resolution (resolution of 512x512 or better). The main disadvantage of intensity data is that it is subject to lighting and surface reflectance effects. These can lead to homogeneous intensity values across edges, making detection of actual edges such as object boundaries difficult, and anomalies such as false edges due to bright or dull spots in objects. Stereo matching uses a pair of intensity images and attempts to detect a correspondence between them,
providing an indication of depth by the amount an area shifts, or is displaced between the images. Unfortunately, the most difficult component of stereo vision is this so-called correspondence problem.

The complexity of the correspondence problem is determined by many factors. In particular, when attempting to detect correspondences, an approach must be capable of dealing with similarity in features (there may be many possible matches), while attempting to compensate for occlusion (due to the difference in perspective), distortion (e.g. circles can become spheres, lines can be of different length, etc.), and perspective differences (where surface shape, lighting, and surface reflectance effects can manifest themselves differently in the images). The most common solution to the correspondence problem is a feature-based approach, where features of the images (most commonly, edge points) are matched, thus reducing the possible search space.

To improve the smoothness of resulting surfaces, researchers have introduced various constraints. The most common are compatibility, figural continuity, and uniqueness. However, these constraints are sometimes too weak to enforce surface smoothness completely since edges may not be present at major surface discontinuities where two adjacent but overlapping objects produce a homogeneous intensity or where surface discontinuities are present [Hoff 89].

1.2. Proposed Work

This thesis describes an approach where stereo image intensity data processing techniques are combined with laser range finding methods. In this thesis, stereo intensity data is used as a supplementary source of depth information, providing depth information for surfaces or portions of surfaces the laser range finder is unable to view due to surface reflectance or absorption. Also, it provides smoother depth values at edges (where many laser reflectance problems occur) and finer detail (since in general, intensity data is capable of providing finer detail from an object than laser data). For its part, the depth information from the laser range data is used at edges common to the intensity and range data to guide the search for a matching edge, thus reducing the stereo correspondence problem. In addition, the weaknesses of the traditional matching constraints are compensated for by the presence of depth information gathered without any dependence on lighting.
conditions. Therefore, where a major discontinuity may be invisible in the intensity data due to homogeneous intensity, it will be present in the range data. The primary goal of this thesis is to demonstrate that the complementary nature of these two sources of data can result in more accurate and complete depth information than depth obtainable by either method alone. The secondary goal is to demonstrate that this method is efficient, and may with sufficient hardware support result in an alternative method for rapid high-resolution depth map acquisition.

The method described uses edge-based correspondences. It is assumed that one of the intensity images is registered with the range image. The registered images are used as the foundation for the remainder of the processing. The first step is edge detection in the range image. Then, since the image intensity data used is of higher resolution than the range data, a data structure called an image intensity pyramid is used to bring the stereo intensity data to the same level of resolution as the range data. Edge detection is then performed in the image pyramid. Edge correspondences between the registered intensity and range image are then used to guide a multiresolution stereo matching and surface interpolation process which uses the power and flexibility of the intensity pyramid to perform more accurate matching at the required level of resolution. By matching intensity edges with a known estimated disparity first, the remaining edges are easier to match since the number of possible candidate matches has been reduced. Once stereo matching is complete, the original range data is used to guide the interpolation of the final surface description, ensuring that surface information from both the laser range data and the stereo matching depth data appear in the final surface description. This final surface description will be more complete and accurate than a comparable description derived from either method alone.

The remainder of this thesis is organised as follows. Chapter 2 contains a survey of recent literature relevant to the work performed in this thesis. Chapter 3 presents the framework for the combination of range and stereo intensity data used in this thesis, and chapter 4 discusses the implementation of this framework. Results and an evaluation of this implementation are presented in chapter 5. The concluding chapter summarises the results of the implementation and discusses important issues raised together with some possible further research directions.
Chapter 2
Related Research

Many researchers have explored the combination of single-camera intensity and range data. The research outlined in this thesis depends on various research issues, each described in a separate section. These issues are: the combination of monocular intensity and range data (to highlight how monocular intensity research can benefit from stereo data), 3-D edge detection (this thesis proposes variations on common methods), and stereo matching and its associated problem of surface interpolation (since variations on the traditional stereo matching and surface interpolation approaches are proposed in this thesis).

2.1. Combination of Intensity and Range Data

Image intensity and laser range data have been used as complementary information sources in various ways [Magee 85a]. One application is motivated by the comparative difference in speed of data acquisition between the laser and intensity-only systems. In these approaches, intensity data is examined to find key features of objects such as corners and center-points. Range data is then acquired only where these key features occur, thus significantly reducing range data acquisition time but adding increased processing of the intensity data. This technique has been applied by researchers for the purposes of object model matching [Magee 85b], motion-parameter detection [Magee 85c], and industrial parts differentiation [Orrock 86].

Boyter and Aggarwal attempt object recognition of synthetic objects using registered intensity and range data [Boyter 84]. They first obtain a number of object models, which are registered range and intensity images of 240 pixels each taken from 256 different views. Then, they present the system with a single view of one of the objects and attempt to match it with one of the object models by calculating a statistical correlation coefficient for the range and intensity image with the corresponding image for all views of each object model. The largest object-model coefficient is
Related Research

taken as the optimal match. Their results indicate that the range data coefficients were more accurate since they were not subject to lighting effects.

Duda et al present a technique for the detection of planes of arbitrary orientation using an ordered search procedure seeking horizontal, vertical, and arbitrarily oriented planar surfaces [Duda 79]. First, horizontal surfaces are removed using a range data histogram. Then, vertical surfaces are extracted from the remaining scene data using the Hough transform, and finally, arbitrarily oriented surfaces are identified through a histogram of the intensity image.

Gil et al use registered range and intensity images to obtain more reliable edges [Gil 83]. Edges are defined to be more reliable in the sense that if an edge is in both images it implies strong evidence of an edge. They claim that if an edge is in the range map only it is a false edge, and an edge in the intensity image only is a possible edge. However, this does not properly consider lighting effects in the intensity map; as mentioned previously, it is possible for a true edge to be present in the range but not intensity maps due to homogeneous intensities across the edge.

Edges in the intensity map are detected using a standard 3x3 Kirsch operator. Jump edges in the range map are detected first using first-order differencing. Roof edges are found by calculating the angle of curvature at every point in the range image using a 9x9 neighbourhood and examining the 3-D vectors between neighbouring points. Once the edges are detected in both images, two methods are described for extracting common edges, which occur wherever an intensity edge corresponds to either type of range edge. The first method, local AND, performs a logical AND for each intensity edge point with a 3x3 neighbourhood at the corresponding position in the range edge map. The second method, global AND, extracts edge segments from the intensity map, then attempts to match to the range map edge segments. The global AND method’s main drawback is that edge segments may be different in the images; they found it difficult to deal with situations such as a single line segment in one image corresponding to multiple segments in the other image. In the local AND method, it is possible for an edge correspondence to appear where there should be none since no consideration is made for possible edge similarities such as orientation.

Chen et al place their main emphasis on the processing of range data, using intensity information
to recover missing parts for the derivation of more complete object descriptions for planar, spherical, and cylindrical objects [Chen 89]. They describe three types of missing parts: those which occur at the outer boundary of an object view (type-A), those which lie in the interior of an object view (type-B), and those which occur at the outer boundary of a particular object view and occlude other objects (type-C) (refer to Figure 2-1). They proceed by first segmenting the range image, classifying regions by analyzing the distribution of unit surface normals on the Gaussian sphere, and producing a region adjacency graph where nodes represent regions and arcs specify relations between them (e.g. roof or jump edge separation). Then, the intensity image is thresholded and line drawings extracted using a Hough transform tuned to the detection of linear and elliptical features.

**Figure 2-1:** The three types of missing parts described in [Chen 89]

Missing parts are extracted in a three-part process. First, jump edges adjacent to the background (i.e. object borders) are projected onto the intensity image. If compatible intensity edges in the neighbourhood of a jump edge are found, this gives strong evidence of no missing part. If no intensity edges are in the neighbourhood then the possibility of a type A missing part is noted.
Next, the possibility of type B missing parts are noted where jump edges join nodes in connected subgraphs of the region adjacency graph. All remaining edges correspond to jump edges connecting two distinct subgraphs and reveal a type B missing part or a type C missing part implicitly embedded in stacked multiple objects. Missing parts are inferred using evidence from the collection of edges marked with the possibility of a missing part. Based on assumptions about the types of recognizable objects, surfaces adjoining a missing part are used to infer vertices to estimate surface parameters for the applicable model. Surface parameters for spheres and cylinders cannot be extracted since a sufficient number of vertices is difficult to obtain.

All of this research demonstrates the complimentary nature of monocular intensity and laser range data. The main conclusions which can be drawn from the available research are that intensity data is useful for discovering missing parts from a range image [Duda 79, Chen 89]. Also, the compatibility of range and intensity edges provides important edge information [Gil 83], and that these edge correlations can be useful in discovering missing parts [Chen 89]. Of all these approaches, the work of Chen et al is the most comprehensive approach to combining range and intensity data. The primary drawbacks in their work is their decision to limit the types of surfaces which can be dealt with, and their inability to derive accurate surface parameters for non-planar surfaces. The approach presented in this thesis addresses these concerns by presenting a stereo camera geometry whereby depth information can be derived for any surface, regardless of type or shape.

2.2. Three-Dimensional Edge Detection

Edge detection is a problem common to both range and intensity image processing. However, since range data is 3-dimensional, edge detection methods for image intensity data are not in general applicable to range data. This is because intensity image edge detection methods are concerned with the detection of areas where large changes in intensity occur. Range images have a similar feature, which are so-called jump or step edges, occurring where large changes in depth occur, such as where one surface or object occludes another. Due to this similarity, intensity edge detection methods are generally capable of jump edge detection. However, they can not detect other important 3-D edges: the roof or crease edge which occurs at surface peaks or where two
surfaces meet forming a peak, and smooth or ramp edges which occur at discontinuities in surface curvature in a smoothly curving surface. The different types of edges identifiable in range images is shown in figure 2-2. This thesis is only concerned with jump and roof edge detection, which is consistent with their importance relative to other edge types in tasks such as object recognition and image segmentation [Besl 85b].

![Figure 2-2: The different types of range edges(from [Besl 86])](image)

(a) Convex roof edge (b) concave roof edge (c) concave ramp edge (d) step (jump) edge (e) convex ramp edge.

Approaches to range image edge detection described in the research literature can be broken down as follows: techniques based on first order differencing (for jump edge detection only), techniques drawing on intensity edge detection methods (and thus not capable of roof edge detection), techniques allowing the detection of jump and roof edges by making limiting assumptions regarding surface types and/or orientation, and techniques based on differential geometry.

A common technique in the research literature for the detection of jump edges is first-order differencing. This is a simple method where the maximum difference in depth between a point and its immediate neighbours is used. This technique generally provides good results, but has been shown to be error-prone for planar surfaces highly oblique to the line of sight and for steeply curved surfaces [Mitchie 83].

Various researchers have had success detecting jump edges by decomposing the 3-D edge detection problem into finding the maxima of the three underlying 2-D edge detection problems [Liu 77, Zucker 81, Monga 86, Monga 89]. These so-called "gradient-based" approaches
Related Research

are effective for detecting jump edges, especially in noisy data [Bhanu 86]. Laurendeau and Boussart use a modified image intensity edge detection algorithm which allows them to detect jump edges and some roof edges in scenes containing a single object [Laurendeau 87]. They first detect jump edges using first-order differencing, producing a sub-image containing the outline of the object. Within this sub-image they apply four edge operators, each applied using three neighbourhood sizes. Their results indicate that limiting the scene to one object and using variably-sized neighbourhoods allows the detection of some roof edges. Unfortunately, their technique is extremely computationally intensive and limited due to the restriction to one object in a scene.

The simplest approach which is successful in detecting both jump and roof edges involves assuming that all surfaces are planar. Milgram and Bjorklund for example, fit planes using 5x5 neighbourhoods to noise-free range data, detecting edges by the difference in surface normals between adjacent planes [Milgram 80]. Their results indicate that this method is extremely susceptible to noise, is very slow, and is inaccurate near jump edges. Mitchie and Aggarwal attempt to address these issues by using a probabilistic model to account for noise and by fitting planes only in the horizontal and vertical orientations to increase edge detection accuracy [Mitchie 83]. Also, jump edges are detected first using first-order differencing, fitting planes only on the interior of an object, thus reducing inaccuracy near edges. Their method seems capable of dealing with noise but is also slow. Sugihara describes an approach using an edge detection operator which determines the presence of jump and roof edges but has no mechanism to distinguish between these edge types or to provide an indication of edge direction [Sugihara 79, Sugihara 87].

The most mathematically rigorous approach to 3-D edge detection is based on the use of differential geometry [Besl 86]. Despite the fact that differential geometry is useful only for smoothly differentiable surfaces [Kreyszig 68] and that laser range data for an object is not smooth, good results have been obtained in each of the following approaches. These approaches generally calculate first and second partial derivatives at all points in the image, using these to calculate surface normals, the principal curvatures $K_1$ and $K_2$ (the maximum and minimum curvatures respectively), and the Gaussian and Mean curvatures for each point. The information from the Gaussian and Mean versus from the principal curvatures is the same [Besl 86] so either pair have been used for edge detection.
The primary differences between the approaches that use differential geometry is in their method of calculating the partial derivatives. The most common approaches are: multi-scale methods, where coarse filters are used to detect features and increasingly finer filters to perform localisation, numerical estimation techniques, and methods based on fitting surfaces to the data then calculating surface curvature from the fit surface.

Ponce and Brady describe a multi-scale approach capable of detecting roof, jump, and smooth edges [Ponce 87]. They first smooth the image with Gaussian masks at different scales, then at each scale they calculate derivatives using a set of characteristic contour masks, each tuned to the detection of edges of different orientation and type. The scale-space behaviour of the detected zero crossings is then used to differentiate the type of edge and its exact location.

In the first of a series of papers, Medioni and Nevatia describe a numerical estimation technique where four 1-D windows are used to compute the first and second order differences for directions $0^\circ$, $45^\circ$, $90^\circ$, $135^\circ$ as an estimate of the principal curvatures [Medioni 84]. In subsequent research, they outline various problems with their original work, noting especially its susceptibility to noise. To deal with these problems, they introduce a multi-scale approach using variable sized edge filters for each of their four 1-D windows, producing a single feature description from the OR of all scales [Fan 86, Fan 87]. Subsequently, this method was expanded by introducing a technique to locally find the most appropriate scale to describe a feature [Parvin 89].

Other approaches numerically approximate surface curvature by fitting planes in a 5x5 neighbourhood to estimate local surface normals [Ittner 85, Hoffman 86, Hoffman 87]. A surface curvature estimate is then obtained by taking the maximum difference in surface normal between a point and its 5x5 neighbour normals. Roof edges are taken as local maxima or minima of the surface curvature estimates; jump edges are detected using a thresholded gradient operator. Yokoya and Levine outline a similar numerical approximation technique which detects jump edges using first-order differencing [Yokoya 89]. Roof edges are located by calculating surface normals from a 5x5 neighbourhood selective biquadratic surface fit (see below) and using the maximum angular difference between adjacent unit surface normals.
A common technique for obtaining differentials is to perform a local surface fit using quadratic surfaces then differentiating the resulting quadratic. Parvin and Medioni use a third-order polynomial fit using linear regression [Parvin 89]. Another technique is to compute a least-square local surface fit using a quadratic model in a 9x9 neighbourhood [Besl 86, Faugeras 87]. All of these authors note that the major problem with this technique is that the surface fit is very poor near edges and is susceptible to noise. Also, the calculated derivatives are not uniform, a result of the fact that there is no smoothness constraint between adjacent surface patches. The selective biquadratic surface [Yokoya 89] fit could be used to reduce surface fit errors due to noise, although the issue of uniform surface fitting is not addressed.

Some researchers fit B-spline surface patches to range data then differentiate the splinar surface. Unlike local surface fitting using quadratics, B-spline surface patches enforce smoothness between adjacent surface patches, thus providing a more uniform fit. Yang and Kak derive a set of 3x3, 5x5, and 7x7 masks approximating the derivatives of a local rational B-spline surface patch, demonstrating good results using the 5x5 masks except near jump edges [Yang 86]. An improvement on this approach is to first detect jump edges using first-order differencing, then use linear regression to fit B-splines under tension to the surface patches defined by dividing the image into overlapped L x L windows (L is unspecified), and excluding jump edges [Vemuri 84, Vemuri 87]. This technique gives good results but involves extremely intensive calculations. The major disadvantage with spline methods is the tradeoff between patch size and computation time. Large surface patches require fewer calculations but provide a poor surface fit, thus making roof edge detection almost impossible; small surface patches provide better surface fit but require more calculations.

The problems of calculating surface curvature for range images are highlighted by Flynn and Jain [Flynn 89]. They compare five methods for surface curvature calculation, three based on calculating surface curvature from a fit surface: orthogonal polynomials [Besl 86] (described above), a surface fit using standard linear regression techniques, and spline-based estimation

---

2It must be noted that this paper is not concerned with surface curvature calculations for edge detection but with the accuracy of the calculations. However, these results should still be considered relevant since more accurate surface curvature calculations provide a firm foundation for accurate edge maps.
(which does not account for fitting surfaces in the vicinity of jump edges), and two numerical estimation techniques: surface normal changes [Hoffman 87] (described above), and directional curvature from derivatives [Fan 86] (described above). Their study involves comparing the surface curvatures calculated by each of the methods on synthetic planar, spherical, and cylindrical data as well as two real range data images. Significant results of their study are: the calculation of differentials is sensitive to noise introduced by the sensor, the numerical estimation techniques provide approximately the same results as the surface fitting techniques, and the calculations are also sensitive to data truncation when the data is scaled to (typically) 8-bit values. To deal with noise, many researchers have used Gaussian smoothing (e.g. [Besl 86, Yang 86]), but Flynn and Jain show that this technique can obliterate or smooth edges. Another common approach is to use large edge masks (5x5 or larger), but these require intensive computations.

The approach presented in this thesis for 3-D edge detection is based on the numerical approximation technique of Yokoya and Levine [Yokoya 89]. Also, an intensity edge mask is used for jump edge detection, as in [Hoffman 87]. One of the goals of the implementation is to maintain efficiency, so an attempt has been made to use small edge masks and rely on image smoothing to reduce the susceptibility of edge detection to noise. The k-nearest neighbour smoothing approach is used due to its reliability in removing noise while still preserving object boundaries [Hoffman 87, Hurt 86].

2.3. Stereo Matching

There are two approaches to solving the stereo image correspondence problem. The first is feature-based stereo, which involves the extraction and matching of features such as edges. The second approach is area-based matching, where corresponding points are found on the basis of area correspondence and similarity measures. Feature-based matching is recognised as being faster than area-based techniques, since the number of candidates for corresponding points are fewer and less sensitive to photometric variations of imaging systems. Since this thesis is concerned with techniques to combine the edge maps from laser range and stereo intensity images, this survey does not cover area-based matching techniques.
There have been many studies of the mechanism of human stereo vision. The key problem lies in determining how image fusion takes place, or equivalently, how corresponding regions or features in two images are matched. The most heavily referenced approach is work by Marr and associates who propose a theory of stereo vision computation which is motivated by the human low-level visual mechanisms [Marr 80, Marr 82]. In their theory, they use compatibility and uniqueness as matching conditions between the images. Their approach is based on initially extracting edges with mask operators of various sizes convolved over both images and extracting zero crossings for each. The correspondence problem is approached by using the disparity measures obtained at lower levels of resolution of the images to guide matching at finer, higher resolutions. Mayhew and Frisby introduce the concept of figural continuity in a similar coarse-to-fine approach [Mayhew 81]. Grimson presents a modified Marr-Poggio matcher using figural continuity which permits vertical disparity adjustment to allow for imprecise location of the zero-crossings due to small variations in epipolar (camera) geometry [Grimson 85]. Grimson also introduces an interesting control strategy for matching and controlling the required search space. Kim and Bovik present a stereo matching algorithm which first matches extremal features (such as endpoints, junctions, and high curvature points), then uses figural continuity to match the remainder of the edges along the contour [Kim 88].

Baker and Binford present a matching strategy based on dynamic programming. They use an intra-scanline search to enforce the uniqueness constraint [Baker 81] on a per-scanline basis. Improvements on their results were obtained by adding an inter-scanline search [Ohta 85]. This search is used to enforce figural continuity along an edge, and takes place in parallel with the intra-scanline search. An alternative approach uses the Hough transform to convert the stereo line matching problem into a point matching problem in the Hough space. Dynamic programming is used in the Hough space to establish point correspondence, thus effectively reducing the required search space [Li 89a].

Some researchers attempt to match high-level features of an image, such as edge segments, junctions, etc. Approaches based on matching straight-line segments use various similarity measures such as edge length, width, contrast, and maximum angular difference [Medioni 85, McIntosh 88, Peng 89]. Lim and Binford take this idea a step further, constructing a
Related Research

hierarchical representation of an image using bodies, surfaces, junctions, curves, and edge segments, where edge segments are the lowest-level structure and bodies are the highest [Lim 88]. In this case, image segmentation and junction classification take place before matching to identify each of the structures. Matching proceeds by attempting to match low-level structures first. If there is an ambiguity, attempts are made to match increasingly higher-level structures until the ambiguity is resolved.

Stewart and Dyer describe a multiresolution approach to edge-based matching using the uniqueness, correspondence, and figural continuity constraints [Stewart 88a]. The algorithm is implemented in a neural network, where each node represents a distinct potential match, and the constraints are implemented directly in the connections between candidate match nodes. Matching works iteratively; during each iteration, each node computes its new activation and output where the new activation is based on the old activation, the decay rate, the supporting input and uniqueness.

Feature-based stereo matching using range data as a guide makes it possible to reduce the complexity of the correspondence problem. Stereo matching can then provide smoother edge information for existing surfaces and, even more importantly, disparity information for surfaces not present in the range data due either to surface reflectance or resolution of the feature being finer than the resolution of the laser range devices.

2.4. Surface Interpolation

One of the major problems in edge-based stereo matching is the problem of surface interpolation. Interpolating from a grid of known points has been thoroughly studied. However, stereo matching produces a sparse disparity map, where the only available depth information is at the edges. Few researchers have examined the problem of surface interpolation under these conditions; it is clear from the available research that greater understanding of human surface development methods are required. Two main constraints have been introduced so far. The first, introduced as no information is information, points out that the absence of zero-crossings constrains possible surface shape, since any rapid surface fluctuation should give rise to a zero-crossing in the edge
map [Grimson 81]. The second constraint is based on the idea that the human visual system constructs the most conservative solution consistent with the data [Barrow 81], which is based on observations that we are able to interpolate smooth curves and surfaces without involving rich semantics. Even when there is more than one possible interpolation for a surface, we choose the one most consistent with the shape of the boundary. Both of these constraints have led researchers to consider the surface interpolation problem modeled in terms of variational calculus.

The idea is to choose an appropriate "performance index" $P$ and define the interpolated surface to be that which minimises the integral of $P$ subject to the boundary constraints. This is analogous to determining the surface formed by fitting a thin elastic plate over a region and using the best-fit surface. This idea has been explored by various researchers. Grimson introduces two techniques for interpolating at a single level of resolution, the quadratic variation, and the square Laplacian, demonstrating that the quadratic variation produces surfaces most consistent with human expectations [Grimson 81]. Since the techniques highlighted by Grimson involve iterative solutions, convergence can be a problem. To address this issue, an alternative multiresolution approach using the same basic mathematical foundation was developed, which uses interactions between levels of resolution of the disparity resulting in a substantial increase in the speed of convergence [Terzopoulos 84].

Several problems with these interpolation methods have been highlighted [Grimson 81]. In particular, interpolation across occlusions is difficult, since these are unmatched zero-crossings not appearing in the disparity map. Also, the interpolated surface contains bumps due to noisy disparity values. Grimson points out that in general finer disparity estimates are required. These can be obtained by increasing the resolution of the sensing devices or by finding zero-crossings at a sub-pixel resolution.

In an attempt to handle surface occlusions, arguments have been made for the integration of feature matching, contour detection, and surface interpolation [Hoff 89, Hoff 87]. Instead of being driven by the traditional smoothness constraints, their algorithm assumes that objects are piecewise smooth (i.e. within which surface normals vary smoothly), where smooth segments are separated by either ridge segments or occlusion segments (i.e. roof and jump edges, respectively). Their
approach is therefore unique, since by attempting to discover surface discontinuities during stereo matching they make the problem of integrated surface interpolation more accurate near these discontinuities. They use a coarse-to-fine multiresolution approach which uses the previous coarse level contour descriptions as a guide to edge matching at the current resolution. Matching is driven from left to right and right to left in parallel. At each resolution level, the algorithm first matches individual edge points. The surface smoothness constraint is used to identify matches, and the depth of each point estimated by a local planar surface fit. These planar patches represent a rough surface approximation. Quadratic patches are then fit at each grid point to the combination of planar patch matches found in the previous step using a standard least squares technique. The quadratic patch containing the most points is kept as the fit for that point. Next, depth and orientation contours are found by fitting bipartite planar patches and detecting discontinuities between the halves. The contours found from the left and right images are then combined and a smooth surface is interpolated away from contours to yield a piecewise smooth surface map at the given resolution. This process is repeated at a finer resolution using the current surface description to predict the location of matches. Although they demonstrate good results, their algorithm is extremely slow and is susceptible to noisy disparity estimates and non-existent matches.

Surface interpolation using range data as a guide makes it possible to deal with the issues of occlusion and noisy disparity values. However, due to the complexity of the surface interpolation problem this thesis does not thoroughly examine the possible advantages. Instead, the technique presented will be called boundary improvement, since it does not perform any interpolation and only attempts to produce a better surface description within the known object boundaries. This technique should be sufficient to demonstrate that a low-resolution range map can be used to guide some form of surface interpolation since the initial depth information gives important clues about overall surface shape.
Chapter 3
A Model for Range-Guided Stereo Matching

3.1. Method Overview

Figure 3-1 shows the overview of the approach used in this thesis. First, image intensity pyramids are constructed for the left and right intensity images to bring them to the same level of resolution as the range image. Then, edge detection is performed in each image in the intensity pyramid and in the range image. In both pyramids, bi-directional links are set up between corresponding edges at different levels of resolution. Also, edge segments are formed by linking each point on the edge with its predecessor and successor on the segment. Edge correspondences are then found between the range image and the top level of the pyramid for the registered intensity image (i.e. the image acquired by the same camera). The depth information from along these edges is then projected to the next level of the pyramid using the inter-level edge linking. Stereo matching is performed at this pyramid level using the depth values as guides to constrain the search for a potential match. Once all edges with associated depth have been matched, remaining edges are matched. Stereo matching is then followed by boundary improvement at the current level of resolution. New depth values are determined using the smoother depth values acquired through stereo matching and the depth values from the range image. These new depth values are projected to the next pyramid level for use as estimated depth values for stereo matching at that level. This process of depth projection, stereo matching, and boundary improvement is carried out to the bottom of the pyramid, where each new level of resolution introduces further levels of refinement to the range data due to increased detail.
Figure 3-1: Order of processing used in the implementation.

1. Construct an intensity pyramid for both intensity images and perform edge detection at each level of the pyramid individually.

2. Perform jump and roof edge detection in the range image.

3. Perform edge combination between the range image and the top level of the registered intensity image.

4. Project depth values from the common edges to the next level of the pyramid as a guide for stereo matching at that level.

5. Perform stereo matching and boundary improvement.

6. Project depth values resulting from boundary improvement to the next pyramid level as a guide for stereo matching at that level.

7. Perform stereo matching and boundary improvement.
3.2. Proposed Camera Geometry

Figure 3-2 shows the camera geometry proposed for use in this thesis. As will be explained later in this chapter, this camera geometry is not actually used in this thesis since the program used to generate a synthetic stereo pair dictates the actual camera geometry. However, it is included here as a possible solution to the missing parts problem. The system consists of a laser range finder and two cameras mounted in a right-angle isosceles triangle similar to the three-camera system used by Stewart and Dyer [Stewart 88b]. Note that this configuration is similar to the traditional triangulation laser system configuration with an additional camera mounted above the original camera. In this system, the camera originally associated with the laser range system collects two images, a low-resolution range map and a high-resolution intensity image. The second camera produces a high-resolution intensity image.

Although many camera configurations are possible, this one is particularly advantageous. Its primary advantage is that the recovery of depth information for missing parts is possible through stereo matching. The stereo camera viewpoints are also sufficiently similar that they can serve as a secondary source of depth information for common surfaces, thus providing a technique capable of dealing with the noise in the laser range data for these surfaces.

The time improvement for range data acquisition is expected to be linearly related to the original time of acquisition, based on the method of acquiring range data [Jarvis 83]. Therefore, it is expected that a scanner capable of acquiring an image of 128 by 128 points per second should be capable of two frames per second when data is acquired at half the original resolution. The variables in this assumption are the speed of the stepping motor which sweeps the laser beam over the scene, the software used to calculate the depth measurements from observed intensity, and determining when the intensity data is acquired with respect to range data acquisition. In the case of the stepping motor and the depth calculation software, the speedup is likely linear if the time at each image point is considered constant. The problem with intensity data acquisition is determining how intensity and range data can be acquired from the same camera’s frame buffer. Clearly, the alternative requiring the least overhead is to capture the range data and use the contents of the frame buffer during the range scan as the intensity image. The only overhead then should be
in redirecting the frame buffer contents to two destinations. Of course, these assumptions will not hold for all image acquisition techniques, but they should be sufficient for the purposes of this thesis.

![Diagram of proposed camera geometry](image)

Figure 3-2: Proposed camera geometry

### 3.3. Image Data Acquisition

One of the major problems encountered in this thesis was the problem of image data acquisition, caused by the lack of an available laser and camera system. All that was available was single-camera laser range data from various sources with no corresponding image intensity data. Therefore, it was necessary to create synthetic images from the original range data. As shown in figure 3-3, various manipulations are performed on the original range image to arrive at three images for use by the implementation: a low-resolution range image and a high-resolution stereo pair of intensity images. Note that the resolution of the original range data constrains the resulting resolution of the intensity images.

Reduced resolution range data is created using a tailor-made program called "iffshrink". When reducing the resolution of an image, groups of four points are reduced to one point in the new image. The iffshrink program sorts the appropriate four values and averages the middle two, assigning this value to the image point in the reduced resolution image. This technique has been shown to have good edge-preserving qualities in multiresolution images [Li 89b].
Figure 3-3: Synthetic data creation
The "synthetic" program [IFF library 89] is used to create a synthetic intensity image from a range image. This program is not state-of-the-art since the Lambertian lighting model is used, thus producing images with poor shading. It takes as input the range image, and optional arguments specifying the elevation and azimuth of the light source as well as a depth value scaling factor. The elevation of the light source is specified in degrees from the image plane, and the azimuth specifies the orientation in a clockwise fashion about the image plane, with 0° being "north". For the purposes of this thesis, the light source was assumed to be behind the laser and camera, thus leading to a light source elevation of 90°. At this elevation, the value specified for light-source azimuth is meaningless.

It was discovered that range image noise affects surface texture in the synthetic intensity images. Therefore, as shown in figure 3-3, to reduce the variation in texture in the synthetic intensity images due to noise in the range image, all range images are smoothed using the k-nearest neighbour approach before synthetic intensity images are generated.

The "stereo" program [IFF library 89] is used to create the required second image for the stereo pair of intensity images. It takes as input an intensity image (the synthetic image created previously), a range image (the range image from which the synthetic image was created), the difference in camera orientation in degrees, and the same range scaling factor used to generate the synthetic intensity image. Note that only one of the images produced by the stereo program is part of the final stereo pair since in order to perform image correspondence between an intensity image and the range image the two images must be registered. Thus, the intensity image generated by the "synthetic" program is used as one image in the pair.

There are obvious disadvantages with using synthetic data:

- Some of the desirable results from this thesis cannot be shown. In particular, it is desirable to show that where a surface is hidden to the laser range scanner, stereo matching can be used to provide depth information for that surface. Unfortunately, since the intensity data is generated from the range data, any such surfaces will be hidden in both images.

- Large camera separations cannot be used. As mentioned previously, the larger the camera separation the more accurate the depth estimation. Unfortunately, when even medium camera separation values are specified for the "stereo" program, the new camera orientations are often directed at a part of the scene for which is no depth information. In these cases, a "shadow" is generated.
Details prevalent in intensity data generated from laser range data are not the same as details in genuine intensity data. In particular, synthetic intensity data can not show extremely fine details since the level of detail is subject to the accuracy of the original range data.

The resolution of the range data constrains the resolution of the intensity images. Real intensity images would be 2-3 times the resolution of the images used in this thesis, providing more detail and thus more accurate disparity (and thus depth) values from stereo matching.

Synthetic data makes it almost impossible to recover the exact depth/disparity relationship, which involves calculating depth from disparity and vice versa. This problem requires the exact camera configuration, including: the focal length of the cameras, separation distance between the cameras (or alternately, angle of separation between the cameras when the viewing distance is known), and the aspect ratio of the image (i.e. the ratio between length and height of the image versus the ratio of length and height of the scene) [Grimson 81]. None of these factors are easily determined or readily available from the "stereo" program. Therefore, a simple lookup table is used for this implementation, where certain ranges of depth produce certain disparity estimates and vice versa. These lookup routines have been localised inside of dedicated modules and can easily be exchanged for routines which calculate exact depth or disparity when the required parameters are available.

The stereo program is restricted to the converging camera model, with horizontally aligned cameras. Therefore, the proposed camera geometry outlined at the beginning of this chapter can not be used. This is not a severe drawback, since the exact camera arrangement is not important to demonstrate the necessary points for this thesis.

Fortunately, registered range and intensity data have been recently acquired from Michigan State University's PRIP (Pattern Recognition and Image Processing) lab. These images are "perfectly" registered, which means that they are of the same resolution. Also, the intensity values are "active"; the values reflect the intensity of the laser stripe as seen by the camera. In practical terms, this gives a sometimes funny-looking, washed out intensity map which is brightest at concave creases and other points where the light concentrates. Of the processing steps described above, the range images still must be reduced in resolution, and a stereo pair of intensity images must be produced. Therefore, the problems associated with these steps as outlined above still apply.
Chapter 4
Implementation of Range-Guided Stereo Matching

An overview of the implementation which shows the exact order of processing used appears in figure 4-1. Each of the following sections describes the major processing steps outlined in the figure.

4.1. Range Image Edge Detection

This thesis explores a novel approach to range image edge detection. As outlined in chapter 2, typical approaches to edge detection use large mask sizes and often require multiple masks to be run over the image and are thus quite slow. Furthermore, differential geometry approaches are handicapped by the necessity for additional calculations after the application of the edge masks. Based on these observations, a unique numerical approximation approach using the intensity gradient edge detection Sobel operator was developed which minimises necessary calculations, reduces the number of required passes over the image, and reduces the operator mask sizes. Jump edges are detected using the Sobel operator, and roof edges by the difference in adjacent surface normals as in [Yokoya 89], where the surface normals are calculated using the output of the Sobel operator. This results in a fast and efficient method for jump and roof edge detection where only four passes over the image are required: the first for jump edge detection and surface normal calculation, the second for jump edge thinning, the third for roof edge detection by surface normal comparisons, and the fourth for roof edge thinning.

In differential geometry, surface normals are calculated using two vectors, $v_0$ and $v_1$, calculated from the estimates of the first derivatives $f_x$ and $f_y$, as follows [Yang 86]:

26
Implementation of Range-Guided Stereo Matching

Figure 4-1: An overview of the order of processing used in the implementation
\[
\begin{align*}
\vec{v}_0 &= [1.0 \ 0.0 \ f_x]^T \\
\vec{v}_1 &= [0.0 \ 1.0 \ f_y]^T
\end{align*}
\]

These surface normal vectors are then used to calculate the surface normal using the following equation:

\[
\vec{n} = \frac{\vec{v}_0 \times \vec{v}_1}{||\vec{v}_0 \times \vec{v}_1||}
\]

Thus, all that is required is to estimate the first-order derivatives, \(f_x\) and \(f_y\). For this task, the Sobel operator was chosen due to its similarity to the surface differential calculation masks presented in [Yang 86]. As in figure 4-2, the masks differ in the multiplicative factor for the centre row or column and the +/- orientation of the \(f_y\) mask. Neither of these differences are major. The difference in multiplicative factor is taken into account by using a different normalisation divisor and the orientation difference is due to a difference in orientation of the coordinate systems.

<table>
<thead>
<tr>
<th>Yang and Kak's operator</th>
<th>(f_x)</th>
<th>(f_y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\frac{1}{12})</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(-1)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>(-4)</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>(-1)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>(-1)</td>
<td>-4</td>
<td>-1</td>
</tr>
<tr>
<td>(0)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(1)</td>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sobel operator</th>
<th>(f_x)</th>
<th>(f_y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\frac{1}{4})</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(-1)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>(-2)</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>(-1)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>(1)</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>(0)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(-1)</td>
<td>-2</td>
<td>-1</td>
</tr>
</tbody>
</table>

*Figure 4-2:* Edge detection masks for the calculation of \(f_x\) and \(f_y\).
In the first edge detection pass, both masks of the Sobel operator are applied at each point in the image. The output of the masks are used to calculate the edge magnitude and gradient, where the magnitude is an indication of the strength of the edge, and the gradient is the edge direction in degrees. The magnitude and gradient are calculated from the $f_x$ and $f_y$ operators using the following equations:

\[
magnitude = \sqrt{f_x^2 + f_y^2} \\
gradient = \tan \left( \frac{f_y}{f_x} \right)
\]

A jump edge is present wherever the edge magnitude exceeds a threshold (a parameter of the program). The gradient of each jump edge is classified into a section number, which is an indication of which of the surrounding eight neighbours the edge "points into" as in figure 4-3. These edge section numbers are used to choose two neighbours for edge thinning using non-maximum suppression as in figure 4-4. Non-maximum suppression is performed with a neighbour if the difference between its edge gradient and the centre point's is less than a threshold (currently set at 30°). If the centre point's edge magnitude is smaller than either of these neighbours, then it is marked as not being an edge.

![Diagram showing edge section numbers](image)

1 = $-22^\circ < x \leq 23^\circ$
2 = $23^\circ < x \leq 68^\circ$
3 = $68^\circ < x \leq 113^\circ$
4 = $113^\circ < x \leq 158^\circ$
5 = $158^\circ < x \leq 180^\circ$ \|$ -180^\circ < x \leq -157^\circ$
6 = $-157^\circ < x \leq -112^\circ$
7 = $-112^\circ < x \leq -67^\circ$
8 = $-67^\circ < x \leq -22^\circ$

**Figure 4-3:** The classification of edge gradient into a section number.
Figure 4-4: Neighbours used for edge thinning.
Roof edge detection is performed by comparing the angular differences in a 3x3 neighbourhood between the centre point and its eight neighbours. This is a variation on the technique used by [Yokoya 89], who use the following angular difference calculation:

$$M_{\text{roof}} = \max \{ \cos^{-1}(\vec{n}(x, y) \cdot \vec{n}(x + k, y + l)), -1 \leq k, l \leq 1 \}$$

Unfortunately, this method requires a large number of calls to the $\cos^{-1}()$ function. Therefore, the following optimisation is used in the implementation3:

$$M_{\text{roof}} = \cos^{-1}(\min(\vec{n}(x, y) \cdot \vec{n}(x + k, y + l) \mid -1 \leq k, l \leq 1))$$

A roof edge is present wherever the value of $M_{\text{roof}}$ exceeds a threshold (a parameter of the program). For each roof edge, the section number is set to the neighbour with the minimum dot product value. Once roof edge detection is complete, roof edges are thinned using the same thinning technique as for jump edges.

Roof edge thinning is the final step in range image processing. The result of this processing is a two dimensional array of the range image data structure shown in Figure 4-5. This structure contains all of the information gathered during range image processing and is used in later processing.

4.2. Intensity Pyramid Construction

The power and flexibility inherent in the image pyramid data structure are important to the results in this thesis. In particular, the pyramid structure is used to reduce the resolution of the intensity images to that of the range image while providing the ability to link (relate) features between levels of resolution as well as at the same level of resolution. These inter and intra-level relationships allow the implementation of a hierarchical stereo matching mechanism which uses information from more than one level of resolution simultaneously.

3It must be noted that this, and the original roof edge detection method proposed by Yokoya and Levine are only effective on the interval $0^\circ \leq x \leq 180^\circ$. Thus, both techniques are effective only when all neighbours of a point are either concave or convex.
typedef struct {
    double u, v, w;  /* components of a 3D vector. */
} vector;

struct ri {
    unsigned char il,  /* Input image range value. */
    re;  /* Roof edge angular difference. */
    short jeg,  /* Jump edge gradient. */
    jem;  /* Jump edge magnitude. */
    vector norm;  /* Surface normal. */
    unsigned jsec : 4,  /* Section of 3x3 grid pointed */
    rsec : 4,  /* to by jump edge gradient. */
    /* Section of maximum angular */
    /* difference for roof edge. */
    isjEdge : 1,  /* TRUE if the jump edge has not */
    /* been thinned. */
    isrEdge : 1;  /* FALSE if the roof edge has */
    /* been thinned. */
};

Figure 4-5: The range image data structure

A pair of pyramids are built, one for each image in the stereo pair of intensity images. Pyramids are implemented as one-dimensional arrays; special pyramid manipulation routines provide the illusion that the array is actually a pyramid. The pyramid is an array of the structures shown in figure 4-6, thus providing the ability to reference all relevant information through a single pointer. The pyramid manipulation routines provide functions to determine information such as the array index for an entry at a certain row and column of a given pyramid level, or the index of a parent or sibling [Li 89c].

struct pyr_elem {
    unsigned gr : 8,  /* gray level value. */
    support_cnt : 8,  /* point's support value. */
    root : 1,  /* TRUE if a root of a */
    /* boundary tree. */
    sMatched : 1,  /* TRUE if the point is */
    /* matched with one in the */
    /* source image. */
    section : 4,  /* section number of edge */
    /* gradient. */
    par : 5,  /* parent pyramid element. */
    /* NOTE: This is set up to */
    /* handle at most 4 parents */
    isEdge : 1;  /* TRUE if not thinned. */
    int left,  /* left supporting */
    /* neighbour point. */
    right;  /* right supporting */
    /* neighbour point. */
    short egrad,  /* edge gradient value. */
    edge;  /* edge value. */
    unsigned short childEdges;  /* Bitmap of compatible */
    /* child edges. */
};

Figure 4-6: Intensity pyramid base structure
The first step in pyramid construction is creating the reduced resolution versions of the original image to a desired image resolution (this is one of the program arguments), and the pyramid constructed up to and including this level. This pyramid construction is performed using a non-iterative, non-overlapped scheme, where four gray level values reduce to one value in the next highest pyramid level, as in [Burt 84]. Gray-level reduction is performed using the best image-preserving technique described in [Li 89b], where the values are sorted and the average of the second and third sorted values is taken as the gray-level value for the parent of the four elements. Once the gray-level pyramid is established, the edge pyramid is constructed.

Edge detection is performed using the Sobel operator, which is applied to all pyramid levels independently. As in range jump edge detection, edge magnitudes are thresholded to determine edges, and for each edge the gradient and its corresponding section are calculated and stored. Also, the same method of edge thinning is used. Once edge thinning is complete, inter- and intra-level edge linking is performed. This linking is very important to subsequent processing since it provides a unique and powerful method for relating edges at the same and different levels of resolution to each other.

Intra-level edge linking is based on the method for finding supporting neighbours described in [Hong 82]. Their technique was chosen since it provides a simple yet powerful mechanism allowing groups of adjacent edge points to form edge segments, with each edge point keeping track of its predecessor and successor in the segment. For each edge point in the image, six neighbours from the surrounding eight neighbours are chosen based on the edge gradient section as predecessor and successor candidates as in figure 4-7. Of these six neighbours, three are candidates for the predecessor and three for the successor. For both candidate predecessors and successors, neighbours are tested in the order shown. For each candidate, if the difference between its edge gradient and the centre point’s is within a threshold (currently set at 90°) then that neighbour is recorded as the predecessor or successor. The order of the candidate evaluations is important, since if all three neighbours are potential predecessors or successors then a bias will be maintained towards straight edge segments.

Inter-level edge linking is based on the overlapped pyramid concept [Burt 84, Hong 82]. In this
Implementation of Range-Guided Stereo Matching

\[ P_n : \text{Candidate for predecessor } n \]
\[ S_n : \text{Candidate for successor } n \]

**Figure 4-7:** Neighbours used for intra-level edge linking by edge section
approach, each edge point has up to four parents and sixteen children. Inter-level linking begins at the top of the pyramid. Each edge checks its children for compatible edges, where an edge is compatible if the difference between their edge gradients is within a threshold (currently set at 45°). For efficiency, bi-directional edge linking is used, where each edge element has two bitmaps, a four bit map for compatible parents and a sixteen bit map for compatible children. Once the entire pyramid has been linked, pyramid relinking is performed.

The purpose of pyramid relinking is to ensure the best possible hierarchical linking. It is made necessary by the fact that edge gradients are different between consecutive levels of resolution in the pyramid. The edge gradients may be sufficiently different to create a situation where edges which should link to each other do not. Pyramid relinking begins at the top level of the pyramid. All edges in the pyramid recalculate their edge gradient from the average of the edge gradients of the children they are currently linked to. If any calculated edge gradients are different than their previous values, then all child and parent links are cleared and the entire pyramid relinked using the method for calculating edge gradients from children outlined above. If no calculated edge gradients differ from their previous value or three iterations of relinking have already been performed, then the process halts.

4.3. Edge Combination

The goal of edge combination is to find edges in the same relative positions in the registered range and intensity image pair. The method used is based on the local AND method found in [Gil 83]. In the local AND method, a 3x3 neighbourhood in the range image is searched for a roof or jump edge, where the centre of the neighbourhood is in the same row and column as the corresponding point in the intensity map. In this implementation, the depth value from the corresponding edge is used to find a stereo match disparity estimate when matching is attempted at that edge point.

Some problems particular to this implementation were discovered with the local AND method. The most major of these is that a common edge is assumed whenever a range edge occurs in the 3x3 correspondence neighbourhood. This lead to many false correspondences (and thus incorrect
depth estimates) since even an edge of opposite orientation could create a correspondence. Therefore, jump edges are considered to correspond to an intensity edge only if the difference between their orientations lies within a threshold (currently set at 30°). Roof edges are considered to correspond to an intensity edge when they lie within the correspondence neighbourhood of an intensity edge.

The use of orientation for finding common jump and intensity edges lead to a critical observation about the choice of edge detection methods for the range and intensity images. Namely, different methods place edges differently and give varying degrees of accuracy in their estimate of orientation. This observation also applies to edge thinning, since a simple difference in method between that used for range and intensity images can reduce the number of correspondences. For all these reasons, the Sobel operator is used for edge detection and the same method for edge thinning is used on all edge types (roof, jump, and intensity).

Another problem with the local AND method for this implementation is that no order of search is specified for the neighbourhood of a point. This is a problem when more than one possible correspondence lies within the neighbourhood. For this implementation, where the best-possible correspondence (and thus depth estimate) must be found, it is necessary to use a prioritised local AND. The order of search in the neighbourhood is determined by the section of the intensity edge orientation, as shown in figure 4-8. The first non-zero depth encountered in this ordered search is used as the initial depth estimate for that point.

Once all correspondences have been found, the depth values are projected down to the next level of the pyramid using the inter-level linking. It must be noted that all edge correspondence information and data resulting from stereo matching and boundary improvement are stored in a new data structure called the combined image pyramid. Note that while there are two intensity image pyramids, only one combined image pyramid is necessary and it corresponds to the intensity image on which edge combination was performed. This pyramid is constructed and indexed in exactly the same fashion as the intensity image pyramid, but uses a different base structure as depicted in figure 4-9.
Figure 4-8: Order of search for the prioritised local AND
4.4. Stereo Matching

Stereo matching begins at the pyramid level below which edge combination takes place. Stereo matching takes place from a source image onto a target image where features from the source image are matched with corresponding features in the target image. In this implementation, the image with corresponding depth information is the source image.

Four data structures are used in stereo matching. For the source image, there is a corresponding image pyramid and a combined image pyramid. The matching algorithm uses the edge linking information from the image pyramid, but does not modify it. The combined image pyramid records information important to stereo matching, such as match disparities, depth values created from stereo matching, and match status as shown in figure 4-9. For the target image, there is an associated image intensity pyramid. The matching algorithm uses its edge linking information and also sets a flag when an edge point is matched, thus avoiding multiple matches to the same edge. The final data structure is a compatible edge list, which exists to reduce the time spent searching for match candidates. Every edge in the source image records its compatible edges from the target image in this list.
The first processing step is to construct the compatible edge list. For every edge in the source pyramid, a list of compatible edges from the target pyramid is built. Edges are assumed to be compatible if they appear in the same row of the image, if the difference between their edge directions are within a threshold (currently set at 30°), and if the column of the target edge lies within the match search window, which is defined based on the direction of search. If the source is the left image and the target the right image, then the search window is to the left of the source edge column. If the source is the right image and the target the left image, the search window is to the right of the source edge column. The final compatible edge list for a source edge is ordered by increasing match disparity (i.e. by increasing difference between the source and target column).

The first step in stereo matching at a pyramid level is matching edges with associated depth. An algorithmic description of stereo matching using depth as a guide is shown in figure 4-10. At the top of the pyramid, depth values are derived directly from the original range data. On all subsequent levels, depth values are derived from the results of stereo matching and boundary improvement at the previous level.

---

4 Currently, edge magnitude is not considered in this comparison. This is largely due to the fact that the number of correspondences in the images considered is typically low. If more complicated images were considered then edge magnitude would be a useful addition to an edge compatibility test.

5 Currently, the search window extends to the border of the image; this could easily be changed by specifying a window size threshold.
Implementation of Range-Guided Stereo Matching

for each edge in the pyramid
    if the edge has associated depth and is not already matched
        Estimate disparity from the given depth.
        Sort the compatible edge list by "distance" from the estimated disparity.
        for each potential match in the sorted compatible edge list
            if at least one linked-to child agree with the match (where the disparity must lie within the inter-level figural continuity threshold)
                if a sufficient percentage of the edge points and their children on the edge segment agree with this match (where each new disparity along the edge must agree with the previous disparity within the intra-level figural continuity threshold)
                    /* Accept the match. */
                    Set the match disparity of all edges on the segment and mark all matched-to edges as matched.
                    if any edges on the segment are unmatched, set their match disparity as the average of their immediate neighbour’s disparities.
                    endif
            endif
        endif
    endif
endfor
endfor

Figure 4-10: Algorithm for stereo matching with depth as a guide

Matching is performed on edge segments, using intra- and inter-level linking to explicitly enforce the figural continuity constraint. This is a departure from other stereo matching approaches which only have intra-level figural continuity. Matching proceeds from the top left to the bottom right of the source image. Each unmatched edge point encountered, is used as a starting point for the edge segment it lies on. For each such edge, the first step is to convert its associated depth to an estimated disparity. Then, all its match candidates from the target image are ordered by their
distance from this disparity, discounting those not lying within the figural continuity threshold of the estimated disparity. This list is then searched for a match with which at least one linked-to child agrees. The validity of this match is tested by ensuring that a sufficient percentage of the edge points along the edge and their children agree with the match.

A match is evaluated sequentially along an edge using the predecessors and successors of the start point (the first edge point encountered in the image scan). As the edge is traversed, a current disparity measure is kept, recording the disparity of the last matched point on the edge. For each new neighbour, the match with disparity closest to the current disparity is found with which at least one linked-to child agrees. A child is said to agree with a match if it is capable of making a match to the equivalent target edge point within the inter-level figural continuity threshold (currently set at three plus two times the disparity from the candidate match at the previous level). If at least one linked-to child agrees with the match at the current disparity and the difference between the disparity of this match and the current disparity is less than the intra-level figural continuity threshold (currently set at 3), the edge point is said to agree with the match and the current disparity is set to the disparity of the current match. Then, the next neighbour is tested using this new disparity. If the new disparity does not agree, then the next neighbour is tested using the same current disparity. If some percentage (currently 60%) of the edge points (and their children) in an edge segment agree with a match, then it is accepted and all edges in the segment and the edges matched to in the target pyramid are marked as matched. Edge points on the same edge segment which were unable to agree with the match are now marked as matched and a disparity is calculated for them based on the average disparity of their predecessor and successor edge points. If an insufficient percentage of points on the edge agree with a match, then a new match is attempted, resulting in the same process of attempting to reach consensus along the edge at a new disparity. If there are no more potential edges, the edge segment remains unmatched and the image search for unmatched edges continues.

Once all possible matches using depth as a guide have been made, the image is searched again, attempting to match remaining edges using a non-guided matching process. Figure 4-11 shows the high-level algorithm for matching without depth as a guide. This matching is performed in exactly the same fashion as matching by depth except that the candidate matches for the initial edge point on an edge are ordered by decreasing disparity instead of by difference with a disparity estimate.
for each edge in the pyramid
    if the edge is not already matched
        Sort the compatible edge list by decreasing disparity.
        for each potential match in the sorted compatible edge list
            if at least one linked-to child agree with the match (where the disparity must lie within the inter-level figural continuity threshold)
                if a sufficient percentage of the edge points and their children on the edge segment agree with this match (where each new disparity along the edge must agree with the previous disparity within the intra-level figural continuity threshold)
                    /* Accept the match. */
                    Set the match disparity of all edges on the segment and mark all matched-to edges as matched.
                    if any edges on the segment are unmatched, set their match disparity as the average of their immediate neighbour's disparities.
                    endif
                endif
            endif
        endfor
    endif
endfor

Figure 4-11: Algorithm for stereo matching without depth as a guide

Some examples have been derived from actual data to show the effectiveness of the hierarchical approach to stereo matching. As they will show, this approach is capable of dealing with poor depth estimates (which range data sometimes provides), allows matching horizontal line segments (a weakness of traditional stereo matching approaches), and is not restricted to straight edge segments. Also, the method of enforcing intra-level figural continuity along edge segments allows the disparity to gradually increase, decrease, or modulate along the edge. This intra-level figural continuity is enhanced by the inter-level figural continuity constraint, which ensures that matches are compatible with their linked-to children's. Note that all of the following examples are performing matching from the left to the right image.
The first example, shown in figure 4-12, shows how the stereo matcher retries a match until the most consistent match is found, how it is capable of dealing with incorrect depth estimates, and how ambiguities can be dealt with using matches made for immediate neighbours. Matching begins at (18,10), and it has an associated depth value of 33, which translates to an estimated disparity of 1, which means matching to (18,9) in the right image. None of the other edge points on the edge segment in the source image can agree with this match, nor the next match with (18,8) at disparity 2, nor the next with (18,7) at disparity 3 so these are all rejected in turn. The next match, to (18,6) at disparity 4 is accepted, since (19,9), (20,9), (21,9), (22,9), and (23,10) all match at disparity 6, which lies within the figural continuity threshold of a maximum difference of 3 between adjacent disparities. At (24,11) there are two matches, but the match to (24,5) is taken since its disparity (6) is closest to the disparity (6) matched by its successor, (23,10). Likewise, at (25,12) the match at disparity 5 to (25,7) is taken. This example also shows a minor flaw with the matching algorithm; the optimal match for (18,10) at disparity 6 to (18,4) is not made, since the figural continuity threshold is large enough to allow the match at disparity 4. It was felt that the most important consideration in this case is that the depth value for that point has been improved by stereo matching, since the new disparity of 4 (and thus calculated depth) provides a depth value closer to the rest of the points on the edge.

![Figure 4-12: Stereo matching, example 1: An illustration of retrying matching](image)

The second example, shown in figure 4-13, illustrates the same features of the stereo matcher as the first example, but also demonstrates matching of a horizontal segment. In this case, matching
begins at (1,85) and the associated depth of 72 produces an estimated disparity of 5. This provides an immediately consistent match. The predecessors of this point, (1,86), (1,87), (1,88), (1,89), (1,90), and (1,91) all match at disparity 5. All its successors match at disparity 6, except (10,79) to (14,76) at disparity 5.

![Image](image.png)

**Figure 4-13:** Stereo matching, example 2: An illustration of matching a long edge segment containing a horizontal segment

The third example, shown in figure 4-14, illustrates how the algorithm is capable of dealing with ambiguity and also how it can set the disparity of edge points after the edge segment has been matched. Matching begins at (96,72), where it has an associated depth of 62, producing an estimated disparity of 4. Its left neighbour, (95,71) is unable to match since it has no compatible edges. To its right, (97,73) matches at disparity 4, and (98,73) and (99,72) at disparity 5. For each of the next six edges, (100,71) to (105,67), matches are made at disparity 4 or 5 since the current disparity measure and the figural continuity threshold do not allow matches to the other edge segment. (106,67) and (108,66) remain unmatched since they have no compatible edges, and the rest successfully match using the current disparity. Once matching is complete on this edge, three edge points are unmatched: (95,71), (106,67), and (108,66). These are all assigned disparity measures based on the average disparity of their predecessors and successors on the edge segment, which in this case gives them all disparity 4. These assigned disparity measures are important since they increase the density of the disparity map, thus providing more information to later
processing steps which use the disparity map. This example can also be used to illustrate how matching would fail if it were to start at (100,71) with disparity 15. Although all its successors on the edge segment, (101,71) to (107,66) can match at this disparity, the fact that the rest of the edge cannot will cause the match to be rejected since in this case only 7 out 17 edge points on the segment successfully matched. Given that this match would be rejected, the second match, at disparity 4 would be attempted and would succeed. This illustrates the point that the figural continuity constraint and the use of a current disparity measure along an edge segment ensure that the same match should be reached regardless of the starting point for matching.

The final example, shown in figure 4-15 illustrates the role played by inter-level figural continuity in stereo matching. The figure shows the overlapped neighbourhoods edges present at two levels of the pyramid: level 6 and 7. The overlapped neighbourhoods for the edges in the left image at level 6 are shown by arrows leading to 4x4 boxes at level 7. The edges appearing within these boxes at level 7 are all linked to by the edge points at level 6. When matching the edge segment at level 6, the depth estimate for (20,46) is large, indicating that a match should be performed to
(20,40) in the right image even though (20,43) is the correct match. The match to (20,40) is rejected by the children, since none of them can match within the inter-level figural continuity threshold at this disparity. The threshold is violated since the proposed match at disparity 6 means that at least one of the children must be able to match at a disparity between 9 and 15. (As outlined above, the inter-level figural continuity threshold is determined by doubling the disparity from the previous level and ensuring that the difference between this disparity and a child's match disparity is at most 3). In this case, none of (40,91), (41,92) or (42,93) can make such a match so it is rejected. The other match for (20,46), to (20,43) is then attempted, and accepted by all of its children (even though one child need agree). This match at disparity 3 is then tested among all the successors of (20,46) along the edge segment at pyramid level 6. In this case, all of the edges on this segment and their children agree with this match.

4.5. Hierarchical Dense Depth Map Generation with Boundary Improvement

The method of boundary improvement chosen for the implementation is intended to demonstrate the viability of combining a coarse laser range depth map with a hierarchical approach for the generation of a dense depth map. The primary problem is in combining the various sources of depth information. One source of depth is the laser range depth map, which is coarse and cannot be relied upon for a complete surface description, not only due to its resolution, but also due to noise in the image, especially at the edges. The other source of depth information are the disparity measures produced from stereo matching. These values are smoother than along the edges of the range image due to the enforcement of figural continuity, but unfortunately are sparse since they correspond only to the edge locations.

The problem of combining a sparse depth map with a coarse one is addressed using a method based on averaging depth values obtained from both depth sources while maintaining as many common edge boundary locations as possible. This method is also hierarchical in nature; the initial depth map is the coarse laser range map, each level further down in the pyramid refines the depth estimates using the sparse depth information from stereo matching.
Boundary improvement uses depth from the previous level and depth obtained along the edges from stereo matching to construct a new depth map at a pyramid level. The first step is to convert disparity values to depth. Then, depth values from the previous level of the pyramid are assigned to the current pyramid level using a non-overlapped scheme where four image points are assigned the value of their parent. If edge combination took place at the previous level, the original laser range data is copied. Otherwise, the calculated stereo depth is copied. Note that two fields in the combined pyramid are used, one to store the estimated depth, the other the calculated depth. Once the estimated depth values are in place, a k-nearest neighbour averaging on a 3x3 neighbourhood is used for each point in the image to calculate its depth. The depth value chosen for each neighbour is the calculated depth if it has one, or the estimated depth if not. If the average of the k-nearest
neighbours is below a threshold (currently set at 45), the calculated depth is set to zero. This ensures that noisy regions from the original range data are ignored and that the object boundaries are kept intact. Once all the depth values are in place, all edge points project their calculated depth down to the next pyramid level to assist stereo matching at that level.
Chapter 5
System Evaluation

Results presented in this chapter are organised by the order of processing used in the implementation:

- Image formation.
- Gray-level pyramid construction and processing.
- Range image edge detection.
- Edge Combination.
- Stereo Matching.
- Surface Interpolation.

5.1. Image Formation

The problems involved in obtaining data were outlined in Chapter 3; this section shows the data which was actually available for testing was used in this implementation. The "hand" images (shown in figure 5-1) and the "cup" images (shown in figure 5-2) were all obtained from Michigan State University and are all of resolution 128x128 points. In these figures, the range values are shown as gray-levels, with white areas being closer. These images do not have registered intensity images, and therefore required all of the processing steps required in Chapter 3: reducing the resolution of the range image, and creating a synthetic stereo intensity image pair.

Unfortunately, only the "hand1" and "hand2" images turned out to be usable for complete testing. Poor stereo and/or synthetic intensity data was produced from all of the other images. The primary problem is that in order to have accurate stereo matching, the disparity must be reasonably large, which implies fairly large camera separation angles. It was discovered that in order to obtain average disparities of 6 or 7 the angle of separation should be at least 10°. An example of the odd looking data resulting from a camera separation of 10° for the "cup3" image is shown in figure 5-3.
Figure 5-1: Original range images: the "hand" images.

Range values are shown as gray level intensities, with white areas being closer to the viewer. (a) hand1; (b) hand2; (c) hand3. All images are 128x128 points.

As shown in this diagram, both of the images in the resulting stereo pair have surfaces obtainable from the original range data, resulting in streaky patches inserted by the stereo program. These
Figure 5-2: Original range images: the "cup" images.

Range values are shown as gray level intensities, with white areas being closer to the viewer. (a) cup1; (b) cup2; (c) cup3. All images are 128x128 points.

Streaks are a result of the stereo program interpolating between known depth values, and are generally present in images where there is fairly large surface curvature as in the cup images or where there is some surface occlusion as in the "hand3" image.
Figure 5-3: An example of the problem of generating stereo intensity data.

Figure (a) shows the synthetic intensity image generated from a smoothed version of the "cup3" image. Figure (b) and (c) show the stereo pair obtained from the image in (a) with a camera separation of 10°. Note that there are clearly missing surfaces in both images as indicated by the streaky patches inserted by the stereo program. Clearly, neither the left or right image produced from the stereo program for this image could be used as the second image in a stereo pair.
Intensities in the interior of the cup are largely homogeneous due to the inability of the laser range finder to deal with surface reflectance in this area of the image. If a corresponding real intensity images were available, the process outlined in this thesis would result in a correct depth map for the interior regions as well, since the necessary edges at the bottom of the cup would appear in the intensity edge maps.

In some cases, good results would be attainable if real intensity data were available. This is particularly true for the "cup2" image, where the inside of the cup shows one of the main problems of laser range finders; that the surface reflectance in the interior of the cup causes the depth values to be almost homogeneous. This is obvious in a synthetic intensity image, where the homogeneity of the depth values is reflected in the resulting intensity values, as in figure 5-4. If actual intensity data were available for this image, the approach outlined in this thesis would be capable of recovering complete depth information, since the edges on the bottom interior of the cup would show up in the intensity image. These edges would then be matched in stereo matching, and a new, more accurate surface description developed based on these new range values obtained for the bottom of the cup.

From the original range images, reduced resolution range images were built as shown in figure 5-5. For these reduced resolution images, feature and boundary retention is quite good.
The generation of synthetic intensity images from the original range images introduced the problem of smoothing the range images. For the hand images, few visible changes in surface occurred. Therefore, to show the effects that image smoothing can have on a range image, figure 5-6 shows an image of a gasoline bottle after performing k-nearest smoothing for 3, 6, and 9 iterations. This image was obtained from Western Softworks, Ltd., located in North Vancouver, British Columbia. Note the "nulls" in the image caused by surface reflectance and that they gradually disappear with no noticeable affect on the object boundaries.

The smoothed versions of the "hand1" and "hand2" images are shown in figure 5-7. Both of these
Figure 5-6: Effects of $k$-nearest neighbour smoothing on a range image.

(a) Original image; (b) Image after 3 iterations, $k=5$, $n=3$; (c) Image after 6 iterations, $k=5$, $n=3$; (d) Image after 9 iterations, $k=5$, $n=3$.

Images were generated using a number of iterations of $k$-nearest neighbour smoothing, with $k=5$ and $n=3$. The "hand1" image required only 3 iterations since it was fairly good to begin with.
Figure 5-7: Smoothed versions of the "hand1" and "hand2" images. Depicted are range images generated from k-nearest neighbour smoothing, with k=5 and n=3. These images were used to generate synthetic intensity images. (a) hand1, after 3 iterations of smoothing; (b) hand2, after 6 iterations of smoothing.

Figure 5-8: MSU PRIP lab registered range and intensity image of a cone ("cone1"). Image dimensions are 106x238. (a) Range Image shown by gray-level; (b) Intensity Image. However, the "hand2" image contained quite a bit of noise, in particular the blotches near the thumb, pinkie and second finger, all of which disappeared after 6 iterations. Both of these smoothed range images were then used to generate synthetic intensity images using UBC's "synthetic" program (as described in Chapter 3). The synthetic intensity image for "hand1" is shown in the next section on pyramid construction. The synthetic intensity image for the "hand2" image is not shown since it is very similar to the "hand1" image.
As mentioned in Chapter 3, the new images recently obtained from MSU's PRIP lab came with registered range and intensity. There are quite a few images, so for the purposes of this thesis, only the most interesting are shown. The first image, shown in figure 5-8 is of a cone. The most interesting aspect of these images is the intensity image, which clearly shows the streaks from the laser range finder. These streaks are also clearly shown in the two images of a Y pipe, in figures 5-9 and 5-10. The final image of an unidentifiable toy (perhaps a joystick) is shown in figure 5-11. In this image, note the large regions where no range was recorded, especially along the bottom. Also, the "hole" in the middle left of the range image contains quite a bit of noise. These incorrect readings are likely caused by surface reflectance problems, and do not appear in the intensity image.
5.2. Pyramid Construction

When building a gray-level intensity pyramid for an image, the most interesting aspect to observe in the images is the retention of features between levels of resolution in the pyramid. The pyramids generated from the synthetic intensity data are particularly uninteresting. Therefore, instead of showing all of the pyramids, one pyramid constructed from the new MSU PRIP intensity data and one from a stereo pair of synthetic intensity images are shown. The first example, shown in figures 5-12 and 5-13 shows the (non-stereo) pyramid constructed for the "toy" image. Figures 5-14 and 5-15 show the stereo intensity pyramid constructed for the "hand" image. The camera separation angle used for the "hand" image is 10°. Note that the images at the bottom level of the "hand" pyramid are the images constructed from the synthetic and stereo programs. This clearly shows the weaknesses of using synthetic intensity data outlined in Chapter 3. The most obvious thing about these images is the lack of features which would be present in real intensity data.

Edge detection from the "hand" intensity pyramids produces the corresponding unthinned edge maps shown in figures 5-16 and 5-17. These maps were produced using an edge detection threshold of 250. In these images, the feature retention properties of the pyramid are particularly evident. In general, the boundaries of the object are well maintained through each level of resolution. The features which gradually disappear are the fine features, such as the splotches on
Figure 5-12: Top and middle levels of the intensity image pyramid for the "toy" image. (a) is the top level of the pyramid, and is of resolution 32x32; (b) is the next level, and is 64x64 points.

Figure 5-13: Bottom-most levels of the intensity image pyramid for the "toy" image. (a) is the level of the pyramid immediately above its bottom, and is 128x128 points; (b) shows the bottom of the intensity pyramid, and is 256x256 points.

The important thing to notice in these maps is that the edges are narrow and that there are no significant gaps in the edges, especially around the perimeter of the hand.
Figure 5-14: Top and middle levels of the intensity image pyramid for the "hand1" image. Shown are the left and right images respectively. (a) and (b) are the top level of their respective pyramid, and are of 32x32 resolution. (c) and (d) are the middle level of the pyramids, and are 64x64.

One of the most surprisingly difficult issues to tackle during implementation and testing was the problem of edge thinning. In particular, the choice of neighbours in the surrounding neighbourhood of an edge point is crucial. The current configuration produced by far the best results of any of the other approaches tried.

5.3. Range Image Edge Detection

The jump edges before and after thinning for the "hand1" and "hand2" images at resolution 32x32 and 64x64 are shown in figures 5-22 and 5-23 respectively. Unfortunately, roof edges are not relevant in these images, so the roof edge maps are not shown.

Another example of jump edge detection and thinning is shown for the "bottle" image in figure 5-24. All jump edges in this example were detected using a threshold of 32.
Figure 5-15: Bottom level of the intensity image pyramid for the "hand1" image. Shown are the left and right images pyramids, respectively. (e) and (f) are the bottom level of the pyramids, and are 128x128. Image (e) is also the synthetic intensity image obtained from the smoothed range image shown in figure 5-7. Image (f) was obtained from the stereo program, with a camera separation of 10°.

To show the viability of the roof edge detection approach outlined in this thesis, two artificial range images were created. The artificial images were created using a custom program which converted a file of numbers created with a standard editor into an iff format image file. Both are images of two planar surfaces sloping together. In the first image, the two surfaces slope upwards, coming together to form a roof. In the second image, the two surfaces slope downwards, coming together to form a pit. Figure 5-25 shows the jump and roof edge maps after thinning for these two images. As expected, the roof edges are two points wide since on both sides of the peak there is a large surface normal difference with the other side of the peak.
Figure 5-16: Unthinned edge maps from the top and middle levels for the "hand1" intensity image pyramid.

Shown are the left and right images respectively. (a) and (b) are the top level of their respective pyramid, and are of 32x32 resolution. (c) and (d) are the middle level of the pyramids, and are 64x64.

5.4. Edge Combination

One of the major concerns with the implementation of the prioritized local AND was ensuring a minimal loss in efficiency from the original local AND method. In the local AND method, neighbours are searched in the same order for all edges, regardless of their orientation. The prioritized local AND method presented in this thesis maintains efficiency using a neighbour search map. For each edge, the neighbour search map is initialised with an appropriately ordered list of pointers to its neighbours' entries in the range image data structure. Thus, all information about every neighbour is available through a single pointer reference. The prioritized search for the first neighbour with associated depth is therefore simply a search through the neighbour search map. Therefore, the only overhead over the original local AND method is the initialisation of the
neighbour search map, which is performed in one procedure call, a test of the edge direction indicator, and eight assignments to the neighbour search map.

In both the "hand1" and "hand2" images, edge combination takes place between the range image and the left intensity image. The images in figure 5-26 show the depth values recorded at the common edges when edge combination takes place at the 32x32 and 64x64 resolution pyramid levels for both the "hand1" and "hand2" images.

Figure 5-27 shows the common edges obtained for the "toy" image at the 64x64 and 128x128 levels of resolution.
System Evaluation

Figure 5-18: Thinned edge maps from the top and middle levels for the "hand1" intensity image pyramid.

Shown are the left and right images respectively. (a) and (b) are the top level of their respective pyramid, and are of 32x32 resolution. (c) and (d) are the middle level of the pyramids, and are 64x64.

5.5. Stereo Matching

All of the images tested in this thesis match from the left image to the right image. To show the effectiveness of matching by depth, the disparity maps between the first pass (matching by depth) and the second pass (matching remaining edges) were written out. All of the following figures show disparity maps, which were written out as inverse intensity so that the disparity values can be more obvious. Also, since the disparity values are generally quite small (1-20), they have been scaled onto the 1-255 interval for display purposes. Note that the expected result in a disparity map is that depth is proportional to disparity. Therefore, in these images, areas which are "closer" should appear as darker intensities.

Figure 5-28 shows the disparity maps obtained for the "hand1" image when stereo matching starts
Figure 5-19: Thinned edge maps from the bottom level for the "hand1" intensity image pyramid.

(a) and (b) are the bottom level of the pyramid for the left and right images respectively, and are 128x128 points.

at the 64x64 level of resolution. Figure 5-30 shows the corresponding disparity maps obtained for the "hand2" image. In the hand2 image at level of resolution of 64x64 note the incorrect matches near the base of the thumb. These match errors were caused by noisy depth estimates, where the depth from the laser range map was large, causing matching with the perimeter of the hand rather than with an edge on the interior as expected. Now, looking at the 128x128 resolution, these same edges are correct. This points out another interesting aspect of the hierarchical approach. In this case, the incorrect matches at the 64x64 level of resolution create a new, more accurate depth estimate to guide stereo matching at the next level of resolution. In this case, the new depth estimates result in a correct match.

Figure 5-29 shows the disparity maps obtained for the "hand1" image when stereo matching starts at the 128x128 level of resolution. Figure 5-31 shows the corresponding disparity maps obtained for the "hand2" image.

A number of general observations can be made from these images. First of all, matching by depth
Figure 5-20: Thinned edge maps from the top and middle levels for the "hand2" intensity image pyramid.

Shown are the left and right images respectively. (a) and (b) are the top level of their respective pyramid, and are of 32x32 resolution. (c) and (d) are the middle level of the pyramids, and are 64x64.

is very effective; the majority of edges are matched in this fashion. Also, edges matched on the second pass of the stereo matching algorithm are generally interior texture points.

Matching at the 64x64 level of resolution often leaves areas unmatched. The problem is that at this level of resolution due to edge blurring, there are not enough coherent edge segments to ensure that correct edge correspondences are made. Also, the disparity maps for the 128x128 level of resolution are virtually identical regardless of where the matching began. This is important since it shows that the matching algorithm does not require completely accurate information from matching at the previous level of resolution in order to succeed.

One of the major problems encountered in stereo matching was in the translation of depth to disparity and vice versa. This is due to having to use a lookup table to perform these conversions
Figure 5-21: Thinned edge maps from the bottom level for the "hand2" intensity image pyramid.

(a) and (b) are the bottom level of the pyramid for the left and right images respectively, and are 128x128 points.

Instead of using the exact camera parameters, the major problem noticed with this approach is in dealing with different levels of image resolution since depth should remain constant regardless of the level of resolution, while disparity increases with increasing resolution. The method decided upon to deal with this involves assuming that the 128x128 resolution images have a standard conversion and that all other resolutions are scaled with respect to the depth/disparity relationship at that level. For example, a disparity of 4 at the 64x64 level produces the same depth estimate as a disparity of 8 at the 128x128 level. Likewise, a disparity of 16 produces the same depth estimate as a disparity of 8 at the 128x128 level. Note that this reduces the effectiveness of stereo matching at higher resolution, since in this example disparity estimates of 16 and 17 will produce the same depth estimate.

Another problem encountered during testing was the effectiveness of the intra-level linking. In particular, it is very hard to detect edge points linked together to form a closed loop. The worst type of closed loop found were edge segments linked into a "lasso"-like fashion. A frequent
Figure 5-22: Jump edge detection in the "hand1" image.

(a) and (b) show the jump edge maps for the 32x32 resolution version of this image, and (c) and (d) for the 64x64 resolution version. (a) and (c) are the jump edge maps before thinning and (b) (d) are the jump edge maps after thinning. Edges were detected using a threshold of 100.

occurrence when examining these edges was to start searching on the "tail" and then infinitely search around the loop. The change made to guard against these and other types of closed loops was to build a list of the first \( N \) (\( N \) is configurable) edge points on the segment and then for each edge test if it is in this list. Obviously, this adds a fair bit of overhead to the examination of edge segments, especially long ones. However, the actual amount of is quite small since the largest value of \( N \) required so far is four.

Figure 5-32 shows the disparity maps obtained from the "toy" image. The figure shows two levels of resolution of disparity maps, one at 128x128, the other at 256x256, where range and intensity edge combination was performed at the 64x64 level. As in the "hand2" image, there are incorrect matches at the 64x64 level, and again they do not affect the matching at the 256x256 resolution. The 256x256 resolution is a good example of how the stereo matcher is capable of
Figure 5-23: Jump edge detection in the "hand2" image.

(a) and (b) show the jump edge maps for the 32x32 resolution version of this image, and (c) and (d) for the 64x64 resolution version. (a) and (c) are the jump edge maps before thinning and (b) (d) are the jump edge maps after thinning. Edges were detected using a threshold of 120.

dealing with complex scenes. As in the "hand" images, most of the points are matched using estimated depth as a guide, and few edges are added by matching the remaining edges. Another feature to note is the incorrect matches near the bottom left of the image (the large disparities surrounded by smaller values). These matches were added when matching without depth.

As a comparison to matching with depth as a guide, stereo matching was performed for the toy image as if all edges had no associated depth. This test was accomplished by only using the second pass of the stereo matching algorithm, where all depth values are ignored. The results for this test are shown in figure 5-33. This image shows some visibly incorrect stereo matches, particularly in areas where there are a large number of correspondences. The most noticeable area is at the top right of the image, where an entire edge segment was matched incorrectly. Other edges matched incorrectly are scattered throughout the image, appearing as brighter areas than the surrounding
disparity values. Another noticeable difference associated with matching without depth as a guide for this image was the time required to perform the matching. In this case, while stereo matching with depth as a guide required approximately 2 seconds on a SUN 4, matching without depth required 10 seconds. This time difference is due to the matcher repeatedly attempting matching at each edge segment, which in some cases causes all possible matches for an edge to be attempted. Finally, there are a (perhaps) surprising number of correct matches when matching is performed without depth as a guide. In particular, the advantages of the inter and intra-level figural continuity constraints and their interactions make this algorithm capable of performing many correct matches, although currently at a cost in computation time. This is an interesting result, indicating that the benefits of the stereo matching algorithm itself bear some further study as a traditional, non-guided approach.

Another example of matching complex scenes is shown in figure 5-34. The figure shows the disparity maps after stereo matching with depth and at the completion of stereo matching. This example is interesting since it shows the disparity maps produced from edge detection with a low
Figure 5-25: Jump and roof edge maps for two artificial test images designed to highlight roof edge detection.

(a) the first artificial range image: two planar surfaces sloping upwards towards each other and meeting at a peak; (b) jump edge map after thinning. The edge detection threshold is 32; (c) roof edge map after thinning detected with a threshold of 130°. (d) the second artificial range image: two planar surfaces sloping downwards towards each other and meeting at a pit; (b) jump edge map after thinning. The edge detection threshold is 32; (c) roof edge map after thinning detected with a threshold of 130°.

threshold, thus resulting in edges on the interior of the object. As in the previous example, note that matching without depth introduces some incorrect matches, but that in general the matches are good.

5.6. Hierarchical Dense Depth Map Generation with Boundary Improvement

As pointed out in Chapter 2, the method of surface interpolation used is a compromise made necessary by the complexity of this problem. Good results were obtained for the two test images "hand1" and "hand2" when the resolution of the range image is 64x64. However, when the range image is at any resolution lower than this, holes appear in the resulting depth map inside the desired
object boundaries. These holes are a result of the poor boundaries in images at coarse levels of resolution, which make it hard to reliably transfer depth based on this poor boundary information. It was also discovered that by appropriately setting k (the number of neighbours averaged together to form the depth at a given point) and the k-nearest neighbour averaging minimum threshold, the resulting object boundaries can be made to align almost exactly with the intended object boundary indicated by the intensity edges. The depth maps after surface interpolation for the "hand1" and "hand2" images are shown in figure 5-35. Both images were produced by averaging the 6-most similar neighbour calculated depth values. The "hand1" surface required a minimum depth of 30, and "hand2" required a minimum depth of 50.

One of the interesting things about this approach is the interactions between the resolution of the image, the minimum depth threshold, and the number of neighbours used for averaging. It was
System Evaluation

Figure 5-27: Depth values obtained at common edges in the "toy" image.
(a) Common edges in for edge combination at the 64x64 resolution level; (b) Common edges in for edge combination at the 128x128 resolution level.

found that for the required minimum depth threshold and the number of neighbours used for averaging are proportional to the resolution of the image. Therefore, lower resolution versions of the image typically should use fewer neighbours for averaging and have a lower minimum depth threshold, while higher resolution versions of the image should use more neighbours for averaging and have a higher minimum depth threshold. It was also discovered that the number of neighbours can remain fairly constant regardless of the image being tested while the minimum depth threshold requires adjustment on a per-image basis. This adjustment was found to be related to the range of depth values encountered in the image. Images which have a small range of depth values require lower thresholds, while those with higher overall depth require higher thresholds.

Figure 5-36 shows the role of the minimum depth threshold in the generation of new depth maps. Shown are four depth maps produced from edge combination at the 128x128 level of resolution. Each depth map was produced using a different minimum depth threshold. The values used were: 10, 20, 30, and 40. These images show a tradeoff between edge "cleanliness" and feature clarity. In the images produced using a low threshold, note the cleanliness of the edges due to an
antialiasing effect, where the low depth values smooth the edge. However, in these images also note the blurring on the holes in the object. The images produced with a higher threshold show the opposite effect; the holes on the interior of the object are sharper, but the edges are also not as smooth.
Figure 5-29: Disparity maps obtained for the "hand1" image when stereo matching starts at the 128x128 level of resolution.

(a) Disparity map immediately following matching using depth as a guide at 128x128 resolution;
(b) Final disparity map at 128x128 resolution.

Figure 5-37 shows the main weakness of the boundary improvement method. These are the depth maps obtained at the 128x128 and 256x256 point levels of resolution for edge combination at the 64x64 level. The blurring evident in both images illustrates the drawback of averaging the depth information from the range image with that obtained at the edges in stereo matching. As illustrated in this example, the effectiveness of the boundary improvement method is largely restricted by the resolution of the range image, thus making it hard to accurately refine the depth map.
Figure 5-30: Disparity maps obtained for the "hand2" image when stereo matching starts at the 64x64 level of resolution.

(a) Disparity map immediately matching using depth as a guide at 64x64 resolution; (b) Final disparity map at 64x64 resolution; (c) Disparity map immediately following matching using depth as a guide at 128x128 resolution; (d) Final disparity map at 128x128 resolution.
Figure 5-31: Disparity maps obtained for the "hand2" image when stereo matching starts at the 128x128 level of resolution.

(a) Disparity map immediately following matching using depth as a guide at 128x128 resolution;
(b) Final disparity map at 128x128 resolution.
System Evaluation

Figure 5-32: Disparity maps obtained for the "toy" image when stereo matching starts at the 64x64 level of resolution.

(a) Disparity map immediately following matching using depth as a guide at 128x128 resolution; (b) Final disparity map at 128x128 resolution; (c) Disparity map immediately following matching using depth as a guide at 256x256 resolution; (d) Final disparity map at 256x256 resolution.
Figure 5-33: Disparity map obtained for the "toy" image when stereo matching is performed without depth as a guide.

This map was obtained from stereo matching at the 256x256 level of resolution by only performing the second pass of stereo matching.
Figure 5-34: Disparity maps obtained for the "wyel" image when stereo matching starts at the 128x128 level of resolution.

Both images are 256x256. (a) Disparity map immediately following matching using depth as a guide; (b) Final disparity map.
Figure 5-35: Depth maps for the "hand1" and "hand2" images resulting from hierarchical depth map generation.

Both of these depth maps are 128x128, and were produced hierarchically from a 64x64 depth map. (a) hand1 produced by averaging k=6 most-similar neighbours with a minimum required calculated depth of 30; (b) hand2 produced by averaging k=6 most-similar neighbours with a minimum required calculated depth of 50.
Figure 5-36: Depth maps for the "toy" image resulting from boundary improvement and the role of the minimum depth threshold.

All of these depth maps are 256x256 points, and were obtained from edge combination at the 128x128 resolution level by averaging the k=6 surrounding neighbours. (a) Depth map obtained with a minimum threshold of 10; (b) Depth map obtained with a minimum threshold of 20; (c) Depth map obtained with a minimum threshold of 30; (d) Depth map obtained with a minimum threshold of 40.
Figure 5-37: Depth maps for the "toy" image resulting from boundary improvement and the blurring of depth information. Both of these depth maps were obtained from edge combination at the 64x64 level of resolution. (a) Depth map at the 128x128 level of resolution; (b) Depth map at the 256x256 level of resolution.
Chapter 6
Conclusion

This thesis has presented a unique approach to the combined processing of laser range data and stereo intensity data. Drawing on the complementary nature of these disparate data sources, it was shown how stereo matching can be simplified using depth information as a guide for determining edge correspondences. Also, a flexible multiresolution stereo matching algorithm was described which is capable of dealing with noisy range data, providing more accurate depth estimates. In addition, the hierarchical nature of this approach allows the creation of a high-resolution depth map from a low-resolution range image and a stereo pair of high-resolution intensity images.

The order of processing explored in this thesis is in direct contrast to other intensity and range approaches, in particular that of [Chen 89]. Their approach is based on edge combination for the purposes of segmentation and surface parameter extraction. However, their use of a single intensity image forces them to restrict the types of recognizable surfaces and results in an inability to reliably extract surface parameters for non-planar surfaces not in the laser range data. The contention in this thesis has been that segmentation and surface parameter calculation will be easier given a less noisy and more complete depth map along with the various information available from the implementation’s data structures.

This thesis has explored the edge-based correspondence issue further than previous researchers. While separate edge detection methods are used for jump and intensity edges [Gil 83], this research has shown that the most reliable edge correspondences are achieved when the same edge detection operator is used for both types of edges. Also, a more general-purpose method of finding edge correspondences has been presented. Unlike the local AND method it is based on, the prioritised local AND method is capable of detecting the region of correspondence in the local neighbourhood with little speed degradation from the original approach. Knowing the region of correspondence in
the local neighbourhood is important information in this thesis, but would prove useful in other applications requiring knowledge of the exact location of the correspondence.

An efficient and reliable method for range image edge detection has been presented. It was shown that the use of differential geometry is a popular current approach but that these methods are very slow. Therefore, a numerical approximation technique from [Yokoya 89] was used but with a different method for the calculation of first-order derivatives. The Sobel intensity edge detection operator was used to detect edges in the intensity image and jump edges in the range image. Fortunately, based on estimations on the operator’s similarity to masks used by [Yang 86], it also proved useful for calculating the necessary derivatives. This resulted in a range image edge detection method for jump and roof edges requiring fewer passes over the image and a smaller number of calculations than traditional techniques.

Arguably, the most important feature of this thesis is the multiresolution approach to stereo matching. While most other researchers use some form of multiresolution matching (e.g. [Grimson 81, Mayhew 81]) their methods rely on weak, unspecified relationships between different levels of resolution of the edge map. This thesis has presented a method where edge relationships between different levels of resolution of the edge map are explicitly represented using inter-level edge linking in an image pyramid data structure. Furthermore, the image pyramid is used to provide intra-level edge linking to form edge segments. These edge segments form the basis of stereo matching and are not restricted to being a straight line as in [McIntosh 88] or any particular orientation as in [Peng 89]. Edge segments also provide an effective means of enforcing the traditional figural continuity constraint. The inter-level edge linking along edge segments allows the enforcement of the inter-level figural continuity constraint. This constraint, not present in other approaches, provides a powerful method of enforcing surface similarity between levels of resolution of the image.

This thesis has demonstrated some of the complementary aspects of laser range and stereo intensity data. In particular, the noise inherent in laser range data, especially along the edges of an object, has been approached by providing a complementary source of depth information, namely stereo matching. As mentioned previously, this algorithm enforces image smoothness using both
an intra- and inter-level figural continuity constraint. Laser range data has been used to guide a stereo matching process where known depth values from the laser range map are converted to estimated disparity values to reduce the search space for candidate matches, thus reducing the correspondence problem. Furthermore, the stereo matching algorithm is not blindly dependent on estimated disparity values. Instead, it is capable of attempting matches of increasing distance from the original disparity estimate until a match succeeds. If matching fails, then perhaps it can succeed when started from one of its neighbours, in which case a new disparity will be assigned to the edge point consistent with its neighbour’s disparities.

The method presented in this thesis should be capable of dealing with missing parts in the range data. The proposed camera geometry should make this possible, since the location of the stereo camera pair provides a sufficiently different perspective from the laser range finder. Also, the stereo matching algorithm should have enough flexibility, since it works in two passes, matching by known depth first, then matching the remaining edges. Missing parts will not appear in the depth map, and will therefore be dealt with in the second pass. In this second pass there are fewer match candidates due to many candidates being matched in the first pass, resulting in a high likelihood of successful match.

This thesis has shown an alternative method for the derivation of a high-resolution depth map. Instead of building better and more expensive laser range finders, it should be possible to use a technique similar to the one presented in this thesis using existing range finders to acquire lower-than-normal resolution range data.

6.1. Future Research Issues

There are many open issues resulting from this thesis research. In particular, the viability of this approach as an alternative for the rapid acquisition of high-resolution range data is an issue requiring further exploration. Clearly, fast hardware and perhaps even custom hardware is required. Another open issue is the "best" camera geometry for the optimal balance between redundant depth information and hidden surface depth recovery.

The problem of acquiring good registered stereo intensity and range data has also left some
unanswered questions. In particular, this applies to the effectiveness of this method and the proposed camera model in dealing with hidden surfaces. Also, intensity data for complex scenes introduces finer surface details than the resolution of the laser range device allows. It is unclear how the current approach of boundary improvement versus surface interpolation would deal with such new, finer details.

Another issue related to the camera geometry is that more work is required in accurately determining the depth/disparity relationship. This thesis has used a simple lookup table to perform these conversions, but this method is not desirable when more accurate calculations are required.

A major issue open to further research is that of surface interpolation. Clearly, some method not totally dependent on the range data is required to deal with surfaces and finer details on existing surfaces introduced by the stereo intensity data. Also, a true surface interpolation method could provide a technique capable of dealing with the "nulls" from laser range scanning. The current implementation would have problems dealing with these areas if they were large in the original range image. The boundary improvement method described in this thesis is not intended to be a surface interpolation approach, but it should make it clear that the availability of range information is beneficial. The multiresolution approach should make it possible to use a method similar to the multiresolution interpolation method in [Terzopoulos 84] with the added advantage that an initial surface description for the majority of the image is already available. Another approach might be to use a refinement technique, where the stereo disparity measurements are used to progressively refine an initial surface fit from the low-resolution range data.

The research on the stereo matching algorithm can be extended. In particular, as was shown in figure 4-12, some incorrect matches are still made. Fortunately, these only occur at the starting point of an edge segment match where more than one candidate match is possible and where the initial match estimate is low compared to the actual value. A solution for such incorrect matches might be to apply an optimality test to the match made at the starting point to ensure that it has the best match with respect to its neighbours. As mentioned previously, this issue was not addressed in the implementation, since in most cases the resulting depth estimates are still much better than those from the original range data.
Another issue worth exploring in stereo matching is the possible improvements to matching without depth information. For example, for edges which could not be matched in the first pass for matching by depth, the partial results from the failed matches could be used to eliminate redundant match attempts or to possibly guide a search for more compatible edges.

Although the method of edge combination presented is successful, it would be interesting to explore further. For example, information which could assist image segmentation is present in the logical OR of the two edge maps. When such information is combined with the local AND method, the relationships between edge placements might be meaningful. For example, it would be interesting to examine the situations in which a jump edge is present and an intensity edge is absent due to homogeneous intensity. Likewise, the presence of an intensity edge where there is no corresponding range edge may perhaps be an indication of a surface or area of fine detail not present in the laser range map. The correspondences between roof and intensity edges would also be interesting to explore further, since in many cases roof edges will not have corresponding intensity edges.

Some other interesting and currently relevant issues could arise from this research. For example, it would be interesting to collect range data in both cameras in the proposed framework, thus providing doubly-redundant depth information. This depth information could be used to drive a bi-directional stereo matcher (where matching takes place in parallel from the left to the right and right to left images as in [Hoff 87]). Such a process should be able to provide more accurate depth maps due to the redundant depth and stereo information.

To further increase the accuracy of the laser range and stereo triangulations, it might be interesting to consider the implications of increasing the displacement between the devices and capturing range data in both cameras. Of course, such a configuration would result in more occlusions, but it would be interesting to explore a bi-directional stereo matcher which analyzes potential occlusions before matching. Most of this occlusion information is available in the edge maps of the laser range data, since occlusions occur at major surface discontinuities [Hoff 89].

Another possible topic for future research is automated camera calibration. Many researchers
have presented techniques showing how camera calibration can be achieved using known depth (e.g. [Faugeras 86]). In this case, the laser range data could be used to calibrate the cameras before stereo matching, thus reducing the error factors in stereo matching due to inaccurate camera parameters.
Appendix A

Glossary

Compatibility
Compatibility is a stereo matching constraint ensuring that only similar features can match each other.

Figural Continuity
Figural continuity is a constraint commonly used in stereo matching, involving an assumption that disparity (the difference in location of a feature between the images) varies smoothly along an edge.

Hough Transform
The Hough transform is a parametric representation of a mathematically definable object, allowing the points making up that object to be identified. For example, one possible Hough transform for a line is its orientation and distance with respect to the origin. If orientation information were available for each point on the line, then the distance of a line passing through that point to the origin could be used to increment a counter in the 2-D hough space at the location defined for the given orientation and distance. This should cause the Hough space to be filled with peaks where a number of points incremented the same location, indicating that they belong to the same line.
Image Pyramid

The image pyramid is a multiresolution data structure commonly used in computer vision research. The pyramid stores reduced versions of the original image at different resolutions where each is a power of 2. For example, if the original image is 512x512 points, the level above this is 256x256, the next 128x128, and so on, thus forming a pyramid of images with the finest resolution version of the image at the bottom of the pyramid and the coarsest resolution version at the top of the pyramid.

Different methods are used to reduce the resolution of an image. The simplest is the non-overlapped method, where four points reduce to one in the next highest pyramid level as in figure A-1. Another common method is the overlapped pyramid, where an image point at a given pyramid level has sixteen children and four parents as in figure A-2. This is called an overlapped pyramid since both the children and parents are non-unique for all image points in the pyramid, except at the corners.

Figure A-1: Child/parent relationships in a non-overlapped pyramid

Gaussian Sphere

The Gaussian sphere is a three-dimensional gaussian distribution. Unit surface normals for example can be analyzed using a Gaussian sphere by mapping them onto the sphere and examining the clustering or groupings which take place.
Lambertian Lighting model
A Lambertian lighting model is one which assumes a single light source located at an infinite distance away from the viewing plane. All surfaces are assumed to be dull matte surfaces which scatter light equally in all directions.

Lateral camera model
The lateral camera model is the most common camera model used in stereo vision research, where two cameras are placed side-by-side and oriented in a parallel fashion much like the human eyes.

Registered Images
Image registration is necessary to easily identify common features in images. Two images are registered with respect to each other if they are taken from the same camera at an identical location and with the same aspect ratio but not necessarily the same resolution. If the images are of the same resolution, they are called perfectly registered.
Glossary

Uniqueness

Uniqueness is a stereo matching constraint ensuring that a feature in one image has at most one corresponding feature in the other image, thus ensuring that matches are unique.
References

[Alvertos 89] Alvertos, N., Dragana, B., Gonzalez, R.
Camera Geometries for Image Matching in 3-D Machine Vision.
*IEEE Transactions on Pattern Analysis and Machine Intelligence*

[Baker 81] Baker, H., Binford, T.
Depth from Edge and Intensity-Based Stereo.

[Barrow 81] Barrow, H., Tenenbaum, J.
Interpreting Line Drawings as Three-Dimensional Surfaces.

[Besl 85a] Besl, P., Jain, R.
Three-Dimensional Object Recognition.

[Besl 85b] Besl, P., Jain, R.
Range Image Understanding.

[Besl 86] Besl, P., Jain, R.
Invariant Surface Characteristics for 3-D Object Recognition in Range Images.

Range Data Processing: Representation of Surfaces by Edges.

[Boyter 84] Boyter, B., Aggarwal, J.
Recognition with Range and Intensity Data.

[Burt 84] Burt, P.
The Pyramid as a Structure for Efficient Computation.
*Multiresolution Image Processing and Analysis.*
Springer-Verlag, 1984, pages 6-35.

[Chen 89] Chen, W., Lie, W., Chen, Y.
Segmented Description of 3-D Objects with Range and Intensity Data.
Duda, R., Nitzan, D., Barrett, P.
Use of Range and Reflectance Data to Find Planar Surface Regions.

Fan, T., Medioni, G., Nevatia, R.
Description of Surfaces from Range Data using Curvature Properties.

Fan, T., Medioni, G., Nevatia, R.
Segmented Descriptions of 3-D Surfaces.

Faugeras, O., Toscani, G.
The Calibration Problem for Stereo.

Faugeras, O., Hebert, M.
The Representation, Recognition, and Positioning of 3-D Shapes from Range Data.
*Three-Dimensional Machine Vision, Kanade, T., Editor.*

Flynn, P., Jain, A.
On Reliable Curvature Estimation.

Gil, B., Mitchie, A., Aggarwal, J.
Experiments in Combining Intensity and Range Edge Maps.

Grimson, W.

Grimson, W.
Computational Experiments with a Feature Based Stereo Algorithm.

Hoff, W., Ahuja, N.
Extracting Surfaces from Stereo Images: An Integrated Approach.

Hoff, W., Ahuja, N.
[Hoffman 86] Hoffman, R., Jain, A.
Segmentation and Classification of Range Images.

[Hoffman 87] Hoffman, R., Jain, A.
Segmentation and Classification of Range Images.

[Hong 82] Hong, T., Shneier, M., Rosenfeld, A.
Border Extraction Using Linked Edge Pyramids.

[Hurt 86] Hurt, S., Rosenfeld, A.

[IFF library 89] Department of Computing Science.
Image File Format (IFF) Software Library. 1989

[Ittner 85] Ittner, D., Jain, A.
3-D Surface Discrimination from Local Curvature Measures.

[Jarvis 83] Jarvis, R.
A Perspective on Range Finding Techniques for Computer Vision.

[Kim 88] Kim, N., Bovik, A.
A Contour-Based Stereo Matching Algorithm using Disparity Continuity.

[Kreyszig 68] Kreyszig, E.
*Introduction to Differential Geometry and Riemannian Geometry.*

[Laurendeau 87] Laurendeau, Boussart.
Model Building of Three-Dimensional Polyhedral Objects Using 3-D Edge Information.

[Li 89a] Li, Z.N.
Dynamic programming in Hough space for line matching in stereopsis.

[Li 89b] Li, Z.N., Hu, G.
Quantitative Evaluation on Edge Preservation in Multiresolution Images.
[Li 89c] Ze-Nian Li.
Pyramid manipulation software library.
1989

[Lim 88] Lim, H., Binford, T.
Structural Correspondence in Stereo Vision.
In Proceedings DARPA Image Understanding Workshop, pages 794-808.

[Liu 77] Liu, H.
Two and Three-Dimensional Edge Detection.

[Magee 85a] Magee, M., Aggarwal, J.
Using Multisensory Images to Derive the Structure of Three-Dimensional
Objects - A Review.

Experiments in Intensity Guided Range Sensing Recognition of Three-
Dimensional Objects.
IEEE Transactions on Pattern Analysis and Machine Intelligence 7(6):629-636,
1985.

[Magee 85c] Magee, M., Aggarwal, J.
Determining Motion Parameters Using Intensity Guided Range Sensing.
In Proceedings of the Seventh International Joint Conference on Pattern

[Marr 80] Marr, D., Poggio, T.
A Theory of Human Stereo Vision.
Technical Report Memo 558, Massachusetts Institute of Technology,

[Marr 82] Marr, D.
Vision.
Freeman Press, 1982.

[Mayhew 81] Mayhew, J., Frisby, J.
Psychophysical and Computational studies towards a Theory of Human
Stereopsis.

[McIntosh 88] McIntosh, J., Mutch, K.
Matching Straight Lines.

[Medioni 84] Medioni, G., Nevatia, R.
Description of 3-D Surfaces using Curvature Properties.
In Proceedings of the DARPA Image Understanding Workshop, pages 291-299.
1984.

[Medioni 85] Medioni, G., Nevatia, R.
Segment-Based Stereo Matching.
[Milgram 80] Milgram, D., Bjorklund, C.
Range Image Processing: Planar Surface Extraction.

[Mitchie 83] Mitchie, A., Aggarwal, J.
Detection of Edges Using Range Information.

[Monga 86] Monga, O., Deriche, R.
A New Three-Dimensional Boundary Detection.

[Monga 89] Monga, O., Deriche, R.
3-D Edge Detection using Recursive Filtering: Application to Scanner Images.

[Ohta 85] Ohta, Y., Kanade, T.
Stereo by Intra- and Inter-Scanline Search Using Dynamic Programming.

[Orrock 86] Orrock, J., Garfunkel, J., Owen, B.
An Integrated Vision/Range Sensor.
*Robot Sensors* vol. 1: Vision, Pugh, A., Editor.
IFS Publications Ltd., 1986, pages 75-84.

[Parvin 89] Parvin, B., Medioni, G.
Adaptive Multiscale Feature Extraction From Range Data.

[Peng 89] Peng, Z., Xu, G.
Stereo Matching by Using Relational Graph.

[Ponce 87] Ponce, J., Brady, M.
Toward a Surface Primal Sketch.
*Three-Dimensional Machine Vision, Kanade, T., Editor.*

[Stewart 88a] Stewart, C., Dyer, C.
Local Constraint Integration in a Connectionist Model of Stereo Vision.

[Stewart 88b] Stewart, C., Dyer, C.
The Trinocular General Support Algorithm: A Three-Camera Stereo Algorithm for Overcoming Binocular Matching Errors.
[Sugihara 79] Sugihara, K.
Range-data Analysis with Intensity and Range Data.

[Sugihara 87] Sugihara, K.
Use of Vertex-Type Knowledge for Range Data Analysis.
Three-Dimensional Machine Vision, Kanade, T., Editor.

[Terzopoulos 84] Terzopoulos, D.
Multiresolution Image Processing and Analysis.

[Vemuri 84] Vemuri, B., Aggarwal, J.
3-Dimensional Reconstruction of Objects from Range Data.

[Vemuri 87] Vemuri, B., Mitchie, A., Aggarwal, J.
3-D Object Representation from Range Data Using Intrinsic Surface Properties.
Three-Dimensional Machine Vision, Kanade, T., Editor.

[Yang 86] Yang, H., Kak, A.
Determination of the Identity, Position and Orientation of the Topmost Object in a Pile.

[Yokoya 89] Yokoya, N., Levine, M.
Range Image Segmentation Based on Differential Geometry.

[Zucker 81] Zucker, S., Hummel, R.
An Optimal Three-Dimensional Edge Operator.